Bavarian Graduate Program in Economics (BGPE)
A Course on Mechanism Design
August 8-13, 2010

Zvika Neeman

Course Outline

The course provides a short introduction to the field known as mechanism design. Mechanism design theory asks when and how it is possible to design rules that induce behavior that is optimal with respect to some given criterion of social welfare. Mechanism design has many applications in economics, ranging from the design of exchange mechanisms, to auctions, price discrimination, regulation, and law and economics. The importance of the field was recently recognized by the Nobel Committee who awarded the 2007 Nobel Prize in Economics to Leonid Hurwicz, Eric Maskin, and Roger Myerson for their contributions to the theory of mechanism design.

The course begins with the classic results of social choice theory (Arrow's and Gibbard & Satterthwaite's Impossibility Theorems), continues with the essentials of implementation theory, and concludes with mechanism design theory where it presents the classic foundations (VCG and AAGV mechanisms) and applications (Myerson-Satterthwaite Impossibility; Myerson's optimal auctions). The course concludes with a discussion of the subjects of renegotiation, robust mechanism design, and collusion, which are on the frontier of current research in mechanism design.

Prerequisites The course is self contained, but basic knowledge of game theory is assumed.

Tentative Lecture Plan

Monday (9/8) Social Choice

7:00-9:30 Breakfast
9:30-10:45 *Lecture* 1: *Aggregation of Preferences.*
10:45-11:00 Coffee break
11:00-12:15 *Lecture* 2: *Arrow's Impossibility Theorem*
12:15-14:00 Lunch
14:00-15:15 *Lecture* 3: *Strategy-proof Implementation.*
15:15-15:45 Coffee break
15:45-17:00 *Lecture* 4: *Gibbard-Satterthwaite's Impossibility Theorem.*
17:00-19:00 Free time
19:00- Dinner

Tuesday (10/8) Implementation

7:00-9:30 Breakfast
9:30-10:45 *Lecture* 1: *Introduction.*

10:45-11:00 Coffee break
11:00-12:15 *Lecture* 2: *Dominant Strategy Implementation.*
12:15-14:00 Lunch
14:00-15:15 *Lecture* 3: *Nash Implementation.*
15:15-15:45 Coffee break
15:45-17:00 *Lecture* 4: *Subgame Perfect and Virtual Implementation.*
17:00-19:00 Free time
19:00- Dinner

Wednesday (11/8) Mechanism Design: Foundations

7:00-9:30 Breakfast
9:30-10:45 *Lecture* 1: *Introduction.*
10:45-11:00 Coffee break
11:00-12:15 *Lecture* 2: *Groves Mechanisms.*
12:15-14:00 Lunch
14:00-15:15 *Lecture* 3: *AAGV Mechanisms.*
15:15-15:45 Coffee break
15:45-17:00 *Lecture* 4: *Examples & Discussion.*
17:00-19:00 Free time
19:00- Dinner

Thursday (12/8) Mechanism Design: Applications

7:00-9:30 Breakfast
9:30-10:45 *Lecture* 1: *Price Discrimination.*
10:45-11:00 Coffee break
11:00-12:15 *Lecture* 2: *Bilateral Bargaining and Double Auctions.*
12:15-14:00 Lunch
14:00-15:15 *Lecture* 3: *Revenue Equivalence.*
15:15-15:45 Coffee break
15:45-17:00 *Lecture* 4: *Optimal Auctions.*
17:00-19:00 Free time
19:00- Dinner

Friday (13/8) Mechanism Design: Frontiers

7:00-9:30 Breakfast
9:30-10:45 *Lecture* 1: *Renegotiation.*
10:45-11:00 Coffee break
11:00-12:15 *Lecture* 2: *Robust Mechanism Design.*
12:15-14:00 Lunch
14:00-15:15 *Lecture* 3: *Collusion.*
15:15-19:00 Free time.
19:00- Dinner

1. Introduction

The text below is taken from the entry "Mechanism Design" that I wrote for the Encyclopedia for the Social Sciences.

Mechanism Design deals with the following types of problems: How to design a "mechanism" or a *game* that has an equilibrium whose outcome maximizes some objective function, such as the maximization of social welfare, subject to certain constraints that depend on the specific problem.

Mechanism design begins with the assumption that each one of the agents for whom the mechanism is designed has access to a different piece of private information, and that elicitation of this information is important for achieving the desired objective. Mechanism design is thus all about incentives: about how to provide the agents with incentives to reveal their private information, and to act in accordance with the designer's objectives. Accordingly, the most important constraint in mechanism design is called "incentive compatibility," or IC. The IC constraint obliges the designer to take into account the fact that the agents will try to manipulate the mechanism to their advantage.

For example, in a famous mechanism design problem the challenge is how to design an auction that maximizes the expected revenue to the seller under the assumption that the willingness of the potential buyers' to pay for the auctioned object is their private information.

The roots of the question of how to collect decentralized information for the purpose of allocating resources can be found in the early debates by economists regarding the feasibility of a centralized socialist economy. These early discussions emphasized the complexity of the systems involved, but it soon became evident that any system for making decisions over the allocation of resources might be open to manipulation. One of the first to recognize the importance of incentives in this context was Leo Hurwicz who coined the term "incentive compatibility" in 1959.

Mechanism design has established itself as a field of study in the early 70s as a result of Hurwicz's work on the possibility of attaining efficient outcomes in dominant strategy equilibria in "economic environments," of Mirrlees's investigation into optimal income taxation schemes, and of the studies of Clarke and Groves of efficient dominant strategy mechanisms for the provision of public goods, which are known today as Vickrey-Clarke-Groves, or VCG, mechanisms (Vickrey has studied such mechanisms in the 60s in the context of his work on auctions). In the late 70s, Arrow and d'Aspremont and Gerard-Varet showed that it was possible to obtain incentive compatible, efficient, and budget-balanced mechanisms. However, in 1983, in their research into optimal mechanisms for bilateral trade, Myerson and Satterthwaite showed that these earlier possibility results might break down if the agents were permitted to refrain from participation in the mechanism if it does not give them an expected utility that is larger than their reservation utility. In 1982, Myerson published a paper on optimal auctions, which to this day acts as the model for implementing mechanism design.

The literature on mechanism design subsequently continued to expand and presently encompasses price discrimination, regulation, public good provision, taxation, auction design, procurement, the organization of markets and trade, and more.

Mechanism Design has not had the effect on policy anticipated by its early practitioners. This is probably because many of its main results are not robust against changes in the details of the underlying environment (as argued by Robert Wilson in the so called "Wilson Critique"). It still remains to be seen whether the current work on "robust mechanism design" would make the theory more practicable.

References

Arrow, K. (1979) "The Property Rights Doctrine and Demand Revelation under Incomplete Information," in *Economics and Human Welfare*, ed. M. Boskin. Academic Press.

d'Aspremont, C. and L.-A. Gerard-Varet (1979) "Incentives and Incomplete Information," *Journal of Public Economics* 11, 25-45.

Clarke, E. (1971) "Multipart Pricing of Public Goods," *Public Choice* 8, 19-33.

Groves, T. (1973) "Incentives in Teams," *Econometrica* 41, 617-631.

Hurwicz, L. (1959) "Optimality and Informational Efficiency in Resource Allocation Processes," in *Mathematical Methods in the Social Sciences*, ed. K. Arrow et al. Stanford University Press.

Hurwicz, L. (1972) "On informationally Decentralized Systems," in *Decision and Organization*, ed. M. McGuire and R. Radner. North-Holland.

Mirrlees, J. (1971) "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies* 38, 175-208.

Myerson, R. (1981) "Optimal Auction Design," *Mathematics of Operations Research* 6, 58-73.

Myerson, R. and M. Satterthwaite (1983) "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory* 28, 265-281.

Vickrey, W. (1961) "Counterspeculation, Auctions, and Competitive Sealed Tenders," *Journal of Finance* 16, 8-37.

Wilson, R. (1987) "Game-Theoretic Analyses of Trading Processes," in *Advances in Economic Theory: Fifth World Congress*, Ed. T. Bewley. Cambridge University Press.

2. Social Choice

"Utilitarianism judges collective action on the basis of the utility levels enjoyed by the individual agents and those levels only. This is literally justice by the ends rather than by the means." (Moulin, 1988) Sen calls the theoretical formulation of utilitarianism as defined above, welfarism. Notice that in such a framework, basic ideals such as freedom, rights, dignity, justice, etc. have no value in and of themselves, but only to the extent that they enhance and are reflected in individuals' utilities. This is a serious shortcoming of the theory (and economics more generally?), and I believe the main reason that social choice theory is so difficult to apply in "real world" situations.[1]

2.1. Aggregation of Preferences and Arrow's Impossibility Theorem

- A – a finite set of alternatives, $\#A \geq 3$.

- $\mathcal{L}$ – set of linear orderings over A. A linear ordering is an ordering of the alternatives by an order of preference, from the most to the least preferred alternative, with no ties. A linear order is a *binary relation* (a set of ordered pairs of elements from a given set) that is characterized by the following properties:

Completeness: $\forall x, y \in A$, either $x \succeq y$ or $y \succeq x$.
Transitivity: $\forall x, y, z \in A$, $x \succeq y$ and $y \succeq z \Longrightarrow x \succeq z$.
Asymmetry: $\forall x, y \in A$, $x \succeq y$ and $y \succeq x \Longrightarrow x = y$. (No indifferences are allowed)

- N – a finite set of agents (that includes n agents).

- For every $i \in N$, preferences are described by $u_i \in \mathcal{L}$. For convenience, we sometimes write $u_i(x) \geq u_i(y)$ instead of $x \succeq_{u_i} y$.

- A social welfare function (SWF) ("social aggregator of preferences") is a mapping $R : \mathcal{L}^N \to \widehat{\mathcal{L}}$ where $\widehat{\mathcal{L}}$ is the set of complete and transitive orderings. Indifferences are allowed.

We are interested in SWF that satisfy "nice" or "attractive" properties (sometimes, the notion of "nice" is a little ad-hoc – for example, a property that helps prove a nice theorem is considered nice).

Definition. A SWF is said to satisfy *unanimity* if it ranks alternative a strictly above b whenever every agent ranks a strictly above b.

[1] For a defense of welfarism, and an argument that shows that any non welfarist rule necessarily gives rise to Pareto inefficient outcomes, see Kaplow and Shavel's recent book, *Fairness vs. Welfare,* Harvard University Press (2002).

Observe that Unanimity captures the idea behind Pareto efficiency in this context.

Another property that is considered by many to be a nice property is *independence of irrelevant alternatives* (IIA).

Definition. A social welfare function is said to satisfy *independence of irrelevant alternatives* (IIA) if the relative social ranking of any two alternatives depends only on their relative ranking by every individual. Formally, a SWF $\succsim$ is said to satisfy IIA if for all $a, b \in A$, and profiles $u, v \in \mathcal{L}^N$:

$$\{i \in N : u_i(a) > u_i(b)\} = \{i \in N : v_i(a) > v_i(b)\} \Longrightarrow \{a \succsim_u b \Longleftrightarrow a \succsim_v b\}.$$

Is IIA attractive? If IIA is violated, then the "decision making body" may have an incentive to manipulate by restricting the set of alternatives to some $B \subseteq A$ and individuals may have an incentive to misrepresent their preferences over irrelevant alternatives.

Example. Borda Rule[2]
Borda rule is an example of a scoring rule. The family of scoring rules is a family of Pareto efficient rules that violate IIA. Consider the following preference profile:

points	agent 1	agent 2	agent 3	Borda Score
3	a	d	b	$b:6$
2	b	a	c	$a:5$
1	c	b	d	$d:4$
0	d	c	a	$c:3$

Alternative b is the winner among $\{a, b, c, d\}$. If the set shrinks to $\{a, b, d\}$ (or if alternative c is pushed to the bottom of individuals' preferences) then any anonymous (symmetric w.r.t. to agents) and neutral (symmetric w.r.t. to alternatives) rule makes $\{a, b, d\}$ tie, and if the set shrinks further to $\{a, b\}$ (or if alternatives c and d are pushed to the bottom of individuals' preferences), then a is favored by majority rule.

The next property is definitely very "unattractive."

Definition. A SWF is dictatorial if there exists an agent (the dictator) such that the SWF coincides with the preferences of this agent (for any profile of preferences!).

Theorem (Arrow's impossibility, 1951). *Suppose that $\#A \geq 3$. Any SWF that satisfies IIA and unanimity is dictatorial. (And conversely, a dictatorial SWF satisfies IIA, and unanimity)*

[2] This rule was devised by the French Academician the Chevalier de Borda for the purpose of the election of members for the French Academy of Sciences. Borda's rule avoids the Condorcet Paradox (presnted in the main text below), but was recognized to be open to manipulation by unscrupulous politicians. For additional historical backgroud, see *The Best of All Possible Worlds: Mathematics and Destiny* by Ivar Ekeland (Chicago University Press).

Example. The Condorcet Paradox[3]

Majority rule, which satisfies IIA and unanimity, may fail to be transitive. This is shown for the following profile of preferences:

agent 1	agent 2	agent 3
a	b	c
b	c	a
c	a	b

By majority rule $a \succ b \succ c \succ a$. A contradiction to transitivity.

Proof. Proof 3 from Geanakoplos (2005).

Strict Neutrality Lemma. If IIA is satisfied, then all binary social rankings are made the same way. Consider two profiles of preferences u and v and two pairs of alternatives a, b and α, β. Suppose each individual has the same relative ranking of α, β in v as he does of a, b in u. Then the social preference between a, b under u is identical to the social preference between α, β under v and both social preferences are strict.

Proof. Assume the pair α, β is not identical to the pair a, b (if they are equal, then the proof follows immediately from IIA). Fix a profile u in which, WLOG, $a \succeq b$ socially. Create a new profile w which is the same as u except that α is just above a (if $\alpha \neq a$) and β is just below b (if $\beta \neq b$).[4] By unanimity, in w $\alpha \succ a$ socially and $b \succ \beta$ socially. By IIA, $a \succeq b$ socially in w. By transitivity $\alpha \succ \beta$ socially in w,[5] and by IIA also $\alpha \succ \beta$ socially in any profile v where individuals hold the same preferences over α, β as over a and b. By reversing the roles of a, b and α, β, we conclude that $a \succ b$ socially also in u. ■

Next, take two distinct alternatives a and b and start with a profile in which every individual strictly prefers b to a. Beginning with individual 1, let each individual successively

[3] The French philosopher Jean-Jacques Rousseau helped pave the way for the French revolution of 1789 by arguing that human beings were naturally virtuous and wise and needed only to be set free from tyrannical governments to order their affairs harmoniously. However, before the French revolution could put these ideas to a practical test, the Marquis de Condorcet, who for the first time (!) used mathematics to model human behavior, showed that majority rule (supposedly representing what a democratic government that is responsive to the will of a free people would do) is logically inconsistent. For additional historical backgroud, see *The Best of All Possible Worlds: Mathematics and Destiny* by Ivar Ekeland (Chicago University Press) and the review article by Freeman Dyson "Writing Nature's Greatest Book" that was published in the *New York Review of Books* in October 19, 2006.

[4] Observe that this can be arranged such that individuals have the same relative preferences over α, β as over a, b. If $a = \beta$ or $b = \alpha$ then for the argument to work it is necessary to repeat it several times for different pairs. For example, if there are three alternatives $\{a, b, c\}$ and the two pairs are (a, b) and (b, a) then the result can be first established for the pairs (a, b) and (c, b) , then for (c, b) and (c, a), and finally for (c, a) and (b, a) .

[5] See Exercise 7.

move a above b. By unanimity and the Strict Neutrality Lemma there will be an individual i^* that moves the social preference from $b \succ a$ to $a \succ b$ when a moves up.[6]

We show that i^* is a dictator. Take an arbitrary pair of alternatives α and β and suppose that $\alpha \succ_{i^*} \beta$. Consider a profile u where the α, β ranking for $i \neq i^*$ is arbitrary. Take an alternative $c \notin \{\alpha, \beta\}$ and consider a new profile v in which c is above everything for individuals $1 \leq i < i^*$, c is below everything for $i^* < i \leq n$, and $\alpha \succ_{i^*} c \succ_{i^*} \beta$. By IIA, the Neutrality Lemma, and by comparison with the profile introduced in the previous paragraph, socially $\alpha \succ c$ and $c \succ \beta$ in profile v. It follows that by transitivity, $\alpha \succ \beta$ in v. Finally, by IIA, $\alpha \succ \beta$ also in the original profile u. ■

Remark 1. Does Arrow's impossibility theorem imply "the impossibility of democracy" as sometimes claimed? I don't think so. For many preference profiles there is no problem to aggregate individuals' preferences. Rather, the Theorem shows that it is impossible to aggregate preferences in a certain *consistent* way (IIA imposes certain consistency requirements among social rankings of alternatives on different preference profiles), and that IIA is "too strong" a consistency requirement in the presence of unanimity.[7] The Theorem tells us that we cannot ignore information about "strength" of preferences (as implied by IIA) if we want non-dictatorial SWFs.

Remark 2. Arrow's impossibility result generated a large literature that tried to figure out how possibility can be re-established. We mention two such attempts.

- if instead of transitivity (i.e., $a \succeq b$, $b \succeq c \Longrightarrow a \succeq c$ which implies $a \succ b$, $b \succ c \Longrightarrow a \succ c$ and $a \sim b$, $b \sim c \Longrightarrow a \sim c$), we only required that the strict part of the SWF be transitive (i.e., only that $a \succ b$, $b \succ c \Longrightarrow a \succ c$, indifference need not be transitive as in example of amount of sugar in coffee) then it can be shown that instead of a dictator, there would be an oligarchy – a set of agents each of which can at least force a tie. An oligarchy is not a big improvement over a dictatorship. If it is small, then it is very similar to a dictatorship, and if it is large, then it means that society is seldom capable of breaking indifference among different alternatives.

- Arrow's impossibility Theorem demonstrates that it is impossible to establish consistent social preferences over the entire domain of individuals' preferences. The power of consistency to rule out social preferences is weakened if the domain of individuals' preferences becomes smaller. This suggests that it may be possible to re-establish possibility by restricting attention to an "interesting" subset of individuals' preferences. Indeed, if the domain of preferences is restricted to single peaked preferences (such preferences arise naturally in a political context), then majority rule (which satisfies IIA and unanimity) can be shown to satisfy transitivity.

[6] The strict neutrality lemma implies that (1) social preferences are always strict; and (2) it's the same individual i^* who "moves" preferences for each pair of alternatives.

[7] Besides, democracy is more complicated than mere "majority rule," even broadly definted. It also requires protection of the rights of the minority, due process, equality before the law, etc.

Remark 3. It seems natural to aggregate preferences by "proximity." Define a metric over linear orders (for example, the minimal number of "flips" of pairs of alternatives that is required to change one linear order to the other). Let social preferences be given by whatever linear order that minimizes the sum of distances from individuals' preferences. It is not clear to me why the literature has not investigated this approach. Indeed, there is no characterization of the social choice rule that corresponds to the metric proposed above (but see Nitzan and Lehrer, JET, 1985).

2.2. Strategy-proof Implementation & Gibbard-Satterthwaite's Impossibility Theorem

Arrow's theorem shows that assuming we know agents' preferences, it is impossible to aggregate them in a "satisfactory way." But in practice we cannot observe agents' preferences. Rather, we must rely on the agents to truthfully reveal them. The Gibbard-Satterthwaite Theorem shows this is impossible to do in a way that is "strategy-proof."

A decision function (voting rule) is a function $f : \mathcal{L}^N \to A$. We focus on voting rules that are single valued (vs. correspondences $f : \mathcal{L}^N \to 2^A$) deterministic (vs. stochastic $f : \mathcal{L}^N \to \Delta(A)$) and that are onto: $\forall a \in A$, $\exists u \in \mathcal{L}^N$ such that $f(u) = a$.

Definition. A SCF f is Pareto efficient if whenever some alternative a is at the top of every individual i's ranking L_i, then $f(L_1, ..., L_N) = a$.

Remark. Observe that this is a weak definition of Pareto efficiency. A stronger definition would require that if all the individuals rank the alternatives in a set F above all the other alternatives, then the decision function does not select an alternative that is not in F.

Definition. A SCF f is monotonic if whenever $f(L_1, ..., L_N) = a$ and for every individual i and every alternative b the ranking L_i' ranks a above b if L_i does (i.e., a "moves up" weakly in i's ranking in L_i' relative to L_i), then $f(L_1', ..., L_N') = a$.

Remark. Notice that because it allows the relative ranking of other alternatives to also change, monotonicity implies a type of independence of irrelevant alternative.

Definition. A SCF f is dictatorial if there is an individual i such that $f(L_1, ..., L_N) = a$ if and only if a is at the top of i's ranking L_i.

We first prove a theorem which is a version of a theorem of Muller and Satterthwaite (JET, 1977).

Theorem. *If $\#A \geq 3$ and $f : \mathcal{L}^N \to A$ is Pareto efficient and monotonic, then f is a dictatorial social choice function.*

Proof. Proof of Theorem A in Reny (2001). ■

Definition. A SCF f is strategy-proof if for every individual i, every $L \in \mathcal{L}^N$, and every $L'_i \in \mathcal{L}$, $f(L_i, L_{-i})$ is ranked weakly above $f(L'_i, L_{-i})$ according to L_i (i.e., it is a dominant strategy for individual i to reveal its preferences truthfully).

Theorem (Gibbard-Satterthwaite's Impossibility, 1973, 1975). *If $\#A \geq 3$ and $f : L^N \to A$ is strategy-proof and onto, then f is dictatorial. (In plain words, any rule that is not dictatorial is sometimes subject to manipulation.)*

Proof. We show that a strategy-proof and onto social choice function is Pareto efficient and monotonic. The proof is taken from Reny's (2001). First, we establish monotonicity. Suppose that $f(L) = a$ and that for every alternative b, the ordering L'_i ranks a above b whenever L_i does. We want to show that $f(L'_i, L_{-i}) = a$. Suppose to the contrary that $f(L'_i, L_{-i}) = b \neq a$. Strategy-proofness implies that $a = f(L)$ is ranked above $f(L'_i, L_{-i}) = b$ according to L_i (if not, then L_i can manipulate). The fact that the ranking of a does not fall in the move to L'_i implies that $a = f(L)$ must also be ranked above $b = f(L'_i, L_{-i})$ according to L'_i. This is a contradiction to strategy-proofness because in this case L'_i can manipulate by reporting L_i. Hence, $f(L'_i, L_{-i}) = f(L) = a$.

Suppose that $f(L) = a$ and that for every individual i and every alternative b, the ordering L'_i ranks a above b whenever L_i does. Because we can move from $L = (L_1, ..., L_n)$ to $L' = (L'_1, ..., L'_n)$ by changing the ranking of each individual i from L_i to L'_i one at a time, and because we have shown that the social choice must remain unchanged for every such change, we must have $f(L') = f(L)$. Hence, f is monotonic.

Next, we establish Pareto efficiency. Choose $a \in A$. Because f is onto, $f(L) = a$ for some $L \in \mathcal{L}^N$. By monotonicity the social choice remains equal to a when a is raised to the top of every individual's ranking. Again by monotonicity, the social choice must remain a regardless of how the alternatives below a are ranked by each individual. Consequently, whenever a is at the top of every individual's ranking the social choice is a. Because a was arbitrary f is Pareto efficient. ■

Remark 1. It is possible to generalize the Gibbard-Satterthwaite Theorem to permit an arbitrary game form where agents are endowed with general message spaces. The Revelation Principle (to be defined and discussed later in the course) implies that strategy-proofness can be replaced by the requirement that equilibria be in dominant strategies.

Remark 2. A random dictator rule is strategy-proof (also anonymous and neutral), but is likely to be inefficient in a quasi-linear world (where individuals' utilities are all measured on the same scale). Suppose for example that preferences are quasi-linear and are given by

Utility	1	2
10	a	c
8	b	b
0	c	a

Random dictator rule: Ex-ante welfare of $\frac{1}{2}(10+0)+\frac{1}{2}(10+0)=10$
Choosing b : Ex-ante welfare of 16.

This shows that a random dictator rule can be very inefficient (although still, of course, Pareto efficient).

Random dictator rules are the only rules that are strategy-proof with probabilistic decision functions, but, some additional mild "attainability" conditions have to be satisfied for that. Without attainability conditions, probabilistic versions of scoring rules, Copeland's and Simpson's rules are also strategy-proof.

Remark 3. The Gibbard-Satterthwaite Theorem also fails to hold when some reasonable restrictions are imposed on the domain of individuals' preferences.

1. **Condorcet Winner.** An alternative is called a "Condorcet winner" if it beats any other alternative in majority comparison.

   Lemma. *Fix an odd N and a restricted domain $D \subseteq \mathcal{L}$ such that for all $u \in D^N$ a Condorcet winner exists [i.e., restrict attention to the set of environments where majority rule produces a well defined winning set]. Then, the decision function that associates with every profile in D^N its Condorcet winner is coalitionally strategy-proof.*

   Proof. Let $CW(u)$ denote the Condorcet winner at $u \in D^N$. Suppose there exists a profile $u \in D^N$, a coalition T and a joint lie $v_T \in D^{\#T}$ such that $CW(u)=a$ but $CW(v_T, u_{N\setminus T})=b$ and $u_i(a)<u_i(b)$ for all $i \in T$ $(*)$. By definition of CW, the set of individuals who prefer a to b under u, denoted $N(u,a,b)$, is a strict majority and by $(*)$ $N((v_T,u_{N\setminus T}),a,b)$ contains $N(u,a,b)$. Hence b cannot be a Condorcet winner at $(v_T,u_{N\setminus T})$.

   Example. Single-peaked preferences.

2. **"Economic Environments."** As we will show later in the course, Groves mechanisms permit dominant strategy implementation in environments with quasi-linear preferences. That is, the space of alternatives is given by $D \times \mathbb{R}^n$ where D is an arbitrary set with no particular structure and $u_i(d,p)=v_i(d)+p_i$. A choice of $d \in D$ is interpreted as the selection of a social alternative, and a choice of $p \in \mathbb{R}^n$ is interpreted as a vector of payments made to the agents. Notice that an individual always prefers a higher p_i to a lower one.

Exercises

1. (Mas-Colell, Whinston, and Green, 21.D.1) Suppose that X is a finite set of alternatives. Construct a reflexive and complete preference relation $\succsim$ on X with the property that $\succsim$ has a maximal element on every strict subset $X' \subseteq X$, and yet $\succsim$ is not acyclic.

2. Define Pareto efficiency of a social welfare function. Does your notion of Pareto efficiency imply or is implied by unanimity?

3. In Step 2 of the proof of Arrow's impossibility Theorem, we argued that there is an individual who "moves the social preference from $b \succ a$ to $a \succ b$." How do we know that the individual does not move the social preference from $b \succ a$ to $b \sim a$?

4. Show that a social welfare function cannot be indifferent between any two alternatives (Hint: use the Strict Neutrality Lemma).

5. Define single-peaked preferences. Show that majority rule satisfies unanimity and IIA if individuals' preferences are all single-peaked.

6. (Mas-Colell, Whinston, and Green, 21.D.7) Construct an example with three alternatives in $\mathbb{R}^2$ and three agents. Each agent should have single peaked preferences on $\mathbb{R}^2$, and yet majority rule should cycle on the three alternatives.

7. Show that transitivity ($a \succeq b$, $b \succeq c \Longrightarrow a \succeq c$) implies that $a \succ b$, $b \succ c \Longrightarrow a \succ c$ and $a \sim b$, $b \sim c \Longrightarrow a \sim c$.

8. Suppose that there are only two alternatives. Give three different examples of a social welfare function that satisfies unanimity and IIA. Give three different examples of a social choice function that is strategyproof.

9. Define a notion of "distance" between two linear orders as the minimal number of "flips of two alternatives" that is needed in order to transform one linear order to another (e.g., the distance between the order *abc* and *bca* is 2). Consider a method for the aggregation of preferences that maps every profile of individuals' preferences into the preference ordering that is closest to this profile (i.e., minimizes the sum of distances from the linear orderings in the profile). What does this method of aggregating preferences produces for an environment with 3 individuals and 3 alternatives? Does your answer generalize to more individuals and alternatives? (Hint: the answer is the Borda rule, and, yes, it generalizes; this method of aggregation seems as good to me as Arrow's, and I don't understand why it didn't receive much attention in the literature).

10. (Osborne and Rubinstein, exercise 183.1) Explain, without making reference to the Gibbard-Satterthwaite Theorem, why the following social choice function is not strategyproof:

$$f(\succsim) = \begin{cases} a & \text{if for all } i \in N,\ a \succ_i b \text{ for every } b \neq a \\ a^* & \text{otherwise} \end{cases}$$

1. Given a social welfare function, can you give a social choice function that would be consistent with it?

2. Given a social choice function, can you give a social welfare function that would be consistent with it?

3.* Show that a social welfare function that is consistent with a strategyproof social choice rule is monotonic and satisfies IIA (Hint: see Moulin, p. 299-300). Arrow's impossibility Theorem then implies that the social welfare function must be dictatorial, which implies that the social choice rule must be dictatorial. This allows us to use Arrow's Impossibility Theorem to provide a short proof for the Gibbard-Satterthwaite Impossibility Theorem.

12.* Can a strategic manipulation of Borda rule ever result in the choice of a Pareto inefficient alternative? Prove or find a counter-example.[8]

13.** 1. Show that Borda rule is "asymptotically strategyproof." That is, show that the proportion of profiles on which an individual can successfully manipulate the social decision in its favor decreases to zero with the number of individuals.[9]

2. Can you find an example of a "Condorcet consistent" rule (a rule that always selects the Condorcet winner when it exists) that is not asymptotically strategyproof?

[8] Hint: see the paper by Baharad and Neeman that is forthcoming in *Social Choice & Welfare*; the paper can be downloaded from my homepage at http://www.tau.ac.il/~zvika/.

[9] This question is based on the paper by Baharad and Neeman that was published in the *Review of Economic Design* in 2002 and that can be downloaded from my homepage at http://www.tau.ac.il/~zvika/. For (1), see the numerical example in p. 337, and for (2) see the example in p. 339.

3. Implementation

This lecture is based on Chapter 10 of Osborne and Rubinstein's text "A Course in Game Theory."

3.1. Introduction

Consider the following set-up.
$N = \{1, ..., n\}$ is a set of individuals.
C is a set of outcomes.
$\mathcal{P}$ is a set of preference profiles over C, $\succsim = (\succsim)_{i \in N} \in \mathcal{P}$.
A social choice rule is a mapping $f : \mathcal{P} \to 2^C$.
A social choice function is a mapping $f : \mathcal{P} \to C$.
The objective is to design a game that would implement the social choice function f in the following way:

$$\begin{tikzcd}
\mathcal{P} \arrow[rr, "f"] \arrow[dr, "s"'] & & C \\
& \underset{\text{game form}}{G} \arrow[ur, "g"'] &
\end{tikzcd}$$

The idea is of "design behind a veil of ignorance." Before players know what their preferences would be, they design a "constitution" that would determine how they would decide what to do later, after they would obtain their preferences. One of the differences between implementation and mechanism design literatures, which we'll study later, is that in implementation literature it is usually assumed that the players' preferences become commonly known among them after the players obtain them.

Example. A "market" implements a Pareto efficient allocation through a "competitive equilibrium" (Hurwicz, 1970s). This result illustrates the triangle above although it is not strictly speaking an example of it because consumers in a "market" are not strategic and so a market is not a game form.

Definition. A strategic game form with consequences in C is a triplet $\left\langle N, (A_i)_{i \in N}, g \right\rangle$ where A_i is a set of actions for player i, and $g : A \to C$ is an outcome function.

A strategic game form and a preference profile $(\succsim)_{i \in N}$ induces a strategic game $\left\langle N, (A_i)_{i \in N}, (\succsim')_{i \in N} \right\rangle$ where each $\succsim'_i$ is defined by $a \succsim'_i b$ if and only if $g(a) \succsim_i g(b)$. (Observe that $\succsim'$ is defined over actions while $\succsim$ is defined over outcomes.)

Definition. An extensive game form with perfect information with consequences in C is a four-tuple $\langle N, H, P, g \rangle$ where
H is a set of histories;
$P : H \backslash Z \to N$ is a player function ($Z \subseteq H$ is the set of terminal nodes)
$g : Z \to C$ is an outcome function.

An extensive game form and a preference profile $(\succsim)_{i\in N}$ induces an extensive form game.

An environment for the planner consists of:
N – a set of players.
C – a set of outcomes.
$\mathcal{P}$ – a set of preference profiles over C
$\mathcal{G}$ – a set of game forms with consequences in C.

The planner must have some idea about how the game it designs will be played. The planner's idea is captured by the solution concept that is used for the game.

Definition. A solution concept for an environment $\langle N, C, \mathcal{P}, \mathcal{G}\rangle$ is a set valued function

$$S : \mathcal{G} \times \mathcal{P} \to \begin{array}{c} \Delta\left(2^A\right) \text{ (for strategic form games, a lottery over a set of action profiles)} \\ \Delta\left(2^Z\right) \text{ (for extensive form games, a lottery over a set of terminal nodes)} \end{array}$$

Definition. Let $\langle N, C, \mathcal{P}, \mathcal{G}\rangle$ be an environment and let S be a solution concept. The game form $G \in \mathcal{G}$ with outcome function g is said to S-implement the social choice rule f if for every profile $\succsim \in \mathcal{P}$, $g\left(S\left(G, \succsim\right)\right) = f\left(\succsim\right)$. In this case, we say that f is S-implementable in $\langle N, C, \mathcal{P}, \mathcal{G}\rangle$.

Remark. Often, $g\left(S\left(G, \succsim\right)\right) = f\left(\succsim\right)$ denotes equality between two sets rather than single outcomes.

It is often the case that the set of actions is equal to the set of preference profiles, and where each player is required to report the entire profile of players' preferences, including the preferences of other players.

Definition. Let $\langle N, C, \mathcal{P}, \mathcal{G}\rangle$ be an environment in which $\mathcal{G}$ is a set of game forms in which the set of actions for each player is the set $\mathcal{P}$ of preferences profiles. Let S be a solution concept. The strategic game form $G \in \mathcal{G}$ with outcome function g is said to truthfully S-implement $f : \mathcal{P} \to C$ if for every profile $\succsim \in \mathcal{P}$,
– $a^* \in S\left(G, \succsim\right)$ where $a_i^* = \succsim$ for every $i \in N$ (truth-telling is a solution), and
– $g\left(a^*\right) \in f\left(\succsim\right)$.
In this case, we say that f is truthfully S-implementable in $\langle N, C, \mathcal{P}, \mathcal{G}\rangle$.

Remark. There are three important differences between truthful implementation and implementation:

1. in truthful implementation, $A_i = \mathcal{P}$ and truth-telling is a solution;

2. a non truthful solution may lie outside $f\left(\succsim\right)$; and

3. in the case of truthful implementation, not every outcome in $f\left(\succsim\right)$ necessarily corresponds to a solution of the induced game.

3.2. Implementation in Dominant Strategies

Suppose that $\mathcal{G}$ is the set of strategic game forms, and S is dominant strategy equilibrium.

Definition. A dominant strategy equilibrium of a strategic game $\left\langle N, (A_i)_{i \in N}, (\succsim)_{i \in N} \right\rangle$ is a profile of actions $a^* \in A$ such that for every player $i \in N$,

$$(a_i^*, a_{-i}) \succsim_i (a_i, a_{-i})$$

for every $a \in A$.

Theorem (Gibbard & Satterthwaite). Let $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$ be an environment in which C contain at least three members each, $\mathcal{P}$ is the set of all possible preferences profiles, and $\mathcal{G}$ the set of strategic game forms. Let $f : \mathcal{P} \to C$ be a choice rule that is dominant strategy implementable and that satisfies the following condition: for every $c \in C$, there exists a profile $\succsim \in \mathcal{P}$ such that $f(\succsim) = \{c\}$. Then f is dictatorial (there exists a player $j \in N$ such that for every preference profile $\succsim \in \mathcal{P}$ and $c \in f(\succsim)$, $c \succsim_j b$ for every $b \in C$).

This theorem is more general than the one we proved in the previous lecture because it is formulated for a general message space and for dominant strategy instead of strategy-proof implementation. However, by using the Revelation Principle (explained below) it is straightforward to generalize the previous argument to this more general case.

Remark. It is possible to implement efficient decision rules in dominant strategies in quasi-linear environments using 'Groves mechanisms.' We will discuss this result in detail in the next chapter of the course, when we talk about mechanism design.

3.3. Nash Implementation

Example (Solomon's trial as a problem of truthful implementation). The example is based on the biblical story in which two women came to King Solomon, each arguing that a certain baby is hers. Solomon, who is considered in Jewish tradition to have been "the wisest of all men" ordered that the baby be cut in half, and each half be given to one woman. One of the women said, fine, neither I nor the other woman will have the baby. The other woman said, no, let her have the baby but just don't cut the baby in two, upon which Solomon declared her the true mother (for showing true motherly love towards the child).

Let's consider this as an implementation problem. The set of consequences is given by:

$$C = \begin{cases} a & \text{give baby to 1} \\ b & \text{give baby to 2} \\ d & \text{cut baby in two} \end{cases}$$

Preferences are given by

$$\begin{array}{lll} \theta \text{ (1 is real mother)} & a \ \succ \ {}_1 b \succ_1 d & b \succ_2 d \succ_2 a \\ \theta' \text{ (2 is real mother)} & a \ \succ \ {}'_1 d \succ'_1 b & b \succ'_2 a \succ'_2 d \end{array}$$

According to the story, the difference between the real and pretend mother is that the real mother cares about the baby itself, not just about herself and about her fight with the baby's real mother.

We want to implement the following social choice function

$$f(\theta) = \{a\}, f(\theta') = \{b\}.$$

The following game form truthfully implements this social choice function

	mine	hers
mine	d	a
hers	b	d

Notice, however, that in addition to the "good" equilibria that implement f there are also "bad" equilibria in which the baby is given to the wrong mother. The notion of truthful implementation is too weak to rule out these "bad" equilibria. Implementation is sufficiently strong, but as we shall see below, it is too strong to be of help in this problem.

The next observation about the "revelation principle" is straightforward and yet powerful. It applies in many different contexts.

Lemma (The Revelation Principle for Nash Implementation). Let $\langle N, C, \mathcal{P}, \mathcal{G}\rangle$ be an environment in which $\mathcal{G}$ is the set of strategic game forms. If a choice rule is Nash-implementable then it is truthfully Nash-implementable.

Proof. Osborne and Rubinstein, p. 185-6. Suppose that players' preferences are given by $\succsim$ and that all players except for i report their preferences truthfully in the truthful mechanism. A report of $\succsim'$ by player i in the truthful mechanism produces the same outcome that would be obtained by the original mechanism when all other players play their equilibrium strategies and player i plays the equilibrium strategy it plays when players' preferences are give by $\succsim$. Since we have a Nash equilibrium under the original mechanism, it follows that player i cannot benefit from not reporting its preferences truthfully (moreover, note that the range of outcomes that player i can achieve by deviating in the truthful mechanism is contained in the range of outcomes it can achieve in the original mechanism). ■

Remark. The Revelation Principle *does not* imply that we may restrict attention to games in which each player announces a preference profile because the game that truthfully Nash-implements a choice rule may have other non truthful Nash equilibria that generate outcomes outside $f(\succsim)$. However, it does imply that we may restrict attention to such games if we want to prove that a certain choice rule is not Nash-implementable.

Definition. A choice rule $f : \mathcal{P} \to 2^C$ is monotonic if whenever $c \in f(\succsim)$ and $c \notin f(\succsim')$, then there exists some player $i \in N$ and some consequence $b \in C$ such that $c \succsim_i b$ and $b \succ_i' c$.

Intuitively, this implies that it is impossible that c has weakly improved its ranking from $\succsim$ to $\succsim'$. Rather, c must have gone down in at least one player's ranking. This definition of monotonicity generalizes the one given in the previous lecture to social choice rules. It coincides with the definition given in the previous lecture for social choice functions (in previous lecture, weak improvement $\Longrightarrow$ choice is preserved; here, choice is not preserved $\Longrightarrow$ not a weak improvement).

Example 1. $f(\succsim)$ is the set of weakly Pareto efficient outcomes,

$$f(\succsim) = \{c \in C : \nexists b \in C \text{ such that } b \succ_i c \text{ for every } i \in N\}$$

Example 2. $f(\succsim)$ consists of every outcome that is the favorite of at least one player,

$$f(\succsim) = \{c \in C : \exists i \in N \text{ such that } c \succsim_i b \text{ for every } b \in C\}.$$

(Observe that this may be a strictly smaller set than the set of weakly Pareto efficient outcomes.)

Proposition (Maskin, 1985). Let $\langle N, C, \mathcal{P}, \mathcal{G}\rangle$ be an environment in which $\mathcal{G}$ is the set of strategic game forms. If a choice rule is Nash-implementable, then it is monotonic.

Proof. Osborne and Rubinstein, p. 186. ■

Example (Solomon's trial as an implementation problem). Recall that the set of consequences is given by:

$$C = \begin{cases} a & \text{give baby to 1} \\ b & \text{give baby to 2} \\ d & \text{cut baby in two} \end{cases}$$

and preferences are given by

$$\begin{array}{lcc} \theta \text{ (1 is real mother)} & a \ \succ \ {}_1 b \succ_1 d & b \succ_2 d \succ_2 a \\ \theta' \text{ (2 is real mother)} & a \ \succ \ {}'_1 d \succ'_1 b & b \succ'_2 a \succ'_2 d \end{array}$$

The social choice function

$$f(\theta) = \{a\}, f(\theta') = \{b\}$$

is not Nash-implementable because f is not monotonic. To see this, note that $a \in f(\theta)$, and $a \notin f(\theta')$, but there does not exist a $y \in C$ and a player $i \in N$ such that $a \succsim_i y$ and $y \succ'_i a$.

So, how come Solomon solved the problem successfully? Osborne and Rubinstein write (tongue in cheek?) that the women probably didn't perceive the situation as a strategic form game. In my opinion, Solomon was bluffing, but the women either believed him, or even if not did not dare call his bluff. In any case, there is no reason that the pretend mother couldn't have also said "give the baby to the other women but don't cut it."

Definition. A choice rule $f : \mathcal{P} \to C$ has no veto power if $c \in f(\succsim)$ whenever for at least $|N| - 1$ players $c \succsim_i y$ for every $y \in C$.

Proposition (Maskin, 1985). Let $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$ be an environment in which $\mathcal{G}$ is the set of strategic game forms. If $|N| \geq 3$, then any choice rule that is monotonic and has no veto power is Nash-implementable.

Proof. Osborne and Rubinstein, pp. 187-8. ■

Remark. The proof relies on a "natural" or "plausible" component. A complaint against the consensus is accepted only if the suggested alternative is no better for the player who complains under the preference profile that is reported by everyone else. I.e., we listen to you (and agree to do what you say is best) only if it does not appear to benefit you. Since this is not supposed to happen if everyone was truthful, the fact that you contest the decision suggests that someone lied. (Compare to the way "whistle-blowers" tend to be treated by the media vs. the organization they criticize.)

A less "plausible" component is the "shouting match," especially since shouting is costless. Jackson (RES, 1992) investigates whether the same result can be obtained with bounded mechanisms. For the case of implementation in undominated strategies, he shows that the answer is negative and that only strategy-proof social choice functions can be implemented.[10]

Remark. Muller and Satterthwaite (1977) have shown that if $|C| \geq 3$ and $\mathcal{P}$ contains all preference profiles then no monotone choice function has no veto power. Since we showed that monotonicity + Pareto efficiency imply dictatorship, this implies that the sufficiency part of Maskin's result applies only to choice rules and on limited domains.

Example (Solomon's trial with money). Osborne and Rubinstein, pp. 190-1. Observation: the 2×2 version of example 190.1 truthfully implements f.

3.4. Subgame Perfect Implementation (with money)

Example (Solomon's trial redux). Once money is introduced, it is possible to implement the choice function

$$\begin{aligned} f(\succsim) &= (1, 0, 0) \\ f(\succsim') &= (2, 0, 0) \end{aligned}$$

where the first coordinate denotes the woman who gets the baby and the next two denote the payments made by the two women, respectively, in a subgame perfect equilibrium as follows. Suppose that the value of the baby to the true mother is strictly larger than M

[10] What about "modulo games"? Check!

and the value of the baby to the pretend mother is strictly smaller than M. Consider the following extensive form game

$$\begin{array}{ccccccc}
 & & \text{mine} & & \text{mine} & \\
 & 1 & — & 2 & — & (2,\varepsilon,M) \\
\text{hers} & | & \text{hers} & | & & \\
 & (2,0,0) & & (1,0,0) & &
\end{array}$$

Remark. Notice that this game gives rise to Nash equilibria that are outside f. When 1 is the real mother she can choose "hers" because she expects 2 to "irrationally" choose "mine."

We now show that every social choice function can "almost" be implemented in a SPE.

Suppose that,

$N = \{1, ..., n\}$ is a set of individuals.

C^* is a set of deterministic consequences;

$C = \{(L, m) : L$ is a lottery over C^* and $m \in \mathbb{R}^n\}$

m_i is interpreted as the fine paid by player i. m_i is not transferred to another player;

Each player has a utility function $u_i : C^* \to \mathbb{R}$; it evaluates the consequence or lottery (L, m) according to $E_L [u_i (c^*)] - m_i$;

A profile of preference profiles is given by $(u_i)_{i \in N}$.

$\mathcal{P} = U^n$ is a finite set that excludes constant functions;

$\mathcal{G}$ is the set of extensive game forms with prefect information and consequences in C.

Definition. A choice function $f : \mathcal{P} \to C^*$ is virtually SPE-implementable if for every $\varepsilon > 0$ there exists an extensive game form $\Gamma \in \mathcal{G}$ such that for every profile $u \in \mathcal{P}$ the extensive game $\langle \Gamma, u \rangle$ has a unique SPE in which the outcome is $f(u)$ with probability greater than or equal to $1 - \varepsilon$.

Proposition (Osborne and Rubinstein, 193.1).

Remark. Abreu and Matsushima (*Econometrica*, 1992) proved a similar result for implementation via iterated elimination of strictly dominated strategies in strategic game forms (which is a little stronger than SPE because it rules out the existence of other equilibria). The variant that is presented here is due to Glazer and Perry.

3.5. Conclusion

Observe that relaxation of the notion of implementation implies an expansion of the set of implementable social choice rules as follows:

notion of implementation		**classes of social choice rule**
dominant strategy implementation	$\Leftrightarrow$	dictatorial social choice rule
Nash implementation	$\Longrightarrow$	monotone social choice rule
	$\Longleftarrow$	monotone social choice rules + no veto power
virtual subgame perfect implementation	$\Leftrightarrow$	every social choice rule

Exercises

1. Osborne and Rubinstein, exercise 191.1, p. 191
2. Mas-Colell, Whinston, and Green, exercise 23.BB.1, p. 925
3. Mas-Colell, Whinston, and Green, exercise 23.BB.2, p. 925
4. Mas-Colell, Whinston, and Green, exercise 23.BB.3, p. 925
5. Is the closure of majority rule monotonic? Prove or give a counter-example.
6. Consider an auction environment with complete information about buyers valuations for the good. Suppose that the objective is to assign the object to the player who has the highest valuation for it. Does a second price auction Nash-implement this objective? Does a second price auction truthfully Nash-implement this objective?
7. Is the monotonicity condition alone sufficient for truthful Nash implementation? Prove or provide a counter example.

4. Mechanism Design

4.1. Introduction

The first part of this lecture is based on chapter 7 of Fudenberg and Tirole's text "Game Theory," and on chapter 23 in Mass-Colell, Whinston, and Green's (MWG) "Microeconomic Theory."

Mechanism design is a subfield of the general theory of implementation. It is distinguished by the fact that (1) it typically assumes that agents have quasi-linear utility functions; (2) it focuses on the case in which the agents are asymmetrically informed, and (3) it focuses on truthful implementation. That is, it typically abstracts away from the fact that a mechanism that implements a certain desired outcome function may also have other, undesired, equilibria. (4) Also, in mechanism design he objective is usually to maximize some objective function such as social welfare rather than implement a given social choice function.

This focus allows mechanism design to consider decidedly more applied problems. The subjects that have received attention in the mechanism design literature include (for each problem, only the first or "classic" reference is given):

– monopolistic price discrimination (Mussa and Rosen, JET 1979),
– optimal taxation (Mirlees, 1971)
– auctions (Myerson, 1981)
– public good provision (Mailath and Postlewaite, RES 1990)
– "market design," or the organization of trade, (Myerson and Satterthwaite, 1983)
– regulation of a monopolist (Baron and Myerson, 1982; Laffont and Tirole, 1986, 1987),
and more.

Mechanism design usually employs the following set-up.

$N = \{1, ..., n\}$ is a set of individuals. The space of alternatives is given by $X \times \mathbb{R}^n$ where X is an arbitrary set of consequences with no particular structure, and $\mathbb{R}^n$ represent monetary transfers to the individuals. Individuals' payoffs depend on their types, which they draw from a *common prior* distribution P on the set of types $\Theta = \Theta_1 \times \cdots \times \Theta_n$. The distribution P is assumed to be commonly known among the agents.[11] In many applications, it is further assumed that agents' types are independent and one-dimensional, or that $P = \prod\limits_{i \in N} P_i$ and $\Theta_i = \left[\underline{v_i}, \overline{v_i}\right]$. A player's type describes the private information of the player. A player may either have private information about its preferences, or about its beliefs, or about both its preferences and beliefs.

The environment is called a *private values* environment if each player i's payoff function is given by $u_i(x, t_i, \theta_i)$ where $x \in X$ denotes a consequence and $t_i \in \mathbb{R}$ denotes the payment made to i. The general case, where u_i may depend on other players' types is called *interdependent values*. For simplicity, we focus our attention in this section on the case of private values. Interdependent values are interesting and give rise to the famous "winner's curse" but are harder to work with.

The environment is said to have *quasi-linear* payoff functions if $u_i(x, t_i, \theta) = V_i(x, \theta) + t_i$ for every $i \in N$, and either $u_0(x, t, \theta) = V_0(x, \theta) - \sum_{i \in N} t_i$ (self interested principal) or $u_0(x, t, \theta) = \sum_{i=0}^n V_i(x, \theta)$ (benevolent principal). In the latter case, it is also usually required that the sum of players' payments sum up to the cost of whatever decision is implemented (ex-post budget balance).

A mechanism is a game form $\left\langle N, (A_i)_{i \in N}, g\right\rangle$ where $g : (A_i)_{i \in N} \rightarrow X \times \mathbb{R}^n$ is a mapping from the set of actions available to each player to a set of consequences X and to a payment to each player. In particular, the game form $\left\langle N, (\Theta_i)_{i \in N}, (x, t)\right\rangle$, which we denote more simply by $\{x(\theta), t(\theta)\}$ is referred to as a *direct revelation mechanism*.

The *revelation principle* implies that for any Bayesian Nash equilibrium under any mechanism, there exists an *incentive compatible* direct revelation mechanism $\{x(\theta), t(\theta)\}$ that implements it in the sense that its truth-telling equilibrium induces the same outcome as the original equilibrium.[12] The proof of the revelation principle is similar to the proof of the revelation principle for Nash implementation. Given a mechanism $M = \langle N, A, g\rangle$ define a direct revelation mechanism $\langle x(\theta), t(\theta)\rangle$ so that it implements the same outcome as M

[11] More generally, it can be assumed that there is a prior P_i for each player. We say that players' beliefs are consistent if $P_i = P \; \forall i$ where P is the common prior. For an interesting discussion about the generality of the common prior assumption (CPA) and the assumption that the model itself is commonly known among the players, see the discussion between Aumann and Gul in *Econometrica* 1998. Aumann is a strong proponent of the common prior assumption. In this discussion, Gul argues, convincingly in my opinion, that the assumption that the model itself is commonly known among the players, which in itself need not involve any loss of generality, does not imply the common prion assumption.

[12] This direct revelation mechanism may also have other equilibria. But these are seldom investigated. This is not a problem if the goal is to establish an impossibility result, but it could undermine the practicability of a possibility result.

if players report truthfully. Suppose that the players' true types are given by θ and that all players except for i report their types truthfully under $\langle x, t\rangle$. By not reporting his type truthfully under $\langle x, t\rangle$ player i can induce a different outcome (one that would have obtained in equilibrium when players' types are (θ_i', θ_{-i})). This cannot possibly benefit player i because if it did, then player i would also have had a profitable deviation opportunity from the equilibrium that is played under M. A contradiction.

Definition. A direct revelation mechanism $\{x(\theta), t(\theta)\}$ is (Bayesian) incentive compatible if every type of every agent prefers to report its type truthfully provided all other types do, or

$$E_{\theta_{-i}}\left[u_i\left(x\left(\theta\right), t_i\left(\theta\right), \theta_i\right)\right] \geq E_{\theta_{-i}}\left[u_i\left(x\left(\widehat{\theta}_i, \theta_{-i}\right), t_i\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_i\right)\right] \qquad \text{for every } i \in N, \text{ and } \theta_i, \widehat{\theta}_i \in \Theta_i.$$

Another important definition is the following.

Definition. A direct revelation mechanism $\{x(\theta), t(\theta)\}$ is ex-post efficient if

$$x\left(\theta\right) \in \underset{x \in X}{\operatorname{argmax}} \sum_{i=0}^{n} V_i\left(x, \theta\right) \qquad \text{for every } \theta.$$

Ex-post efficiency thus requires that an efficient decision be made for any profile of players' types. The notions of *ex-ante* and *interim efficiency* are defined in a similar way. A mechanism is said to be *interim efficient* if, at the interim stage, when each player knows its own type but not the other players' types, there is no other mechanism that gives each type of each player a weakly higher expected payoff and a strictly higher expected payoff to at least one type of one player. A mechanism is said to be *ex-ante efficient* if, at the ex-ante stage, before the players even learn their own types, there is no other mechanism that gives each player a weakly higher expected payoff and a strictly higher expected payoff to at least one player.

Ex-ante efficiency implies interim efficiency which, in turn, implies ex-post efficiency. It follows that a mechanism that is incentive compatible and ex-ante efficient is incentive compatible and interim efficient, and a mechanism that is incentive compatible and interim efficient is incentive compatible and ex-post efficient. A mechanism that is incentive compatible and ex-ante, interim, or ex-post efficient, is sometimes called ex-ante, interim, or ex-post incentive efficient, respectively.[13]

Example. Suppose that $N = \{1, 2\}$, $\Theta_1 = \Theta_2 = \{a, b\}$, players' types are independent and equally likely, and $u_1(x, \theta) = u_2(x, \theta) = \sqrt{x}$. Suppose that total income is equal to 1 in

[13] See Holmström and Myerson (*Econometrica*, 1983) for an interesting discussion about these notions and the relationship among them.

every state of the world. The mechanism

$$x(\theta) = \begin{array}{c|c|c|} & a & b \\ \hline a & 1,0 & 0,1 \\ \hline b & 0,1 & 1,0 \\ \hline \end{array}$$

is ex-post efficient, but not interim efficient. The mechanism

$$x(\theta) = \begin{array}{c|c|c|} & a & b \\ \hline a & 1,0 & 1,0 \\ \hline b & 0,1 & 0,1 \\ \hline \end{array}$$

is ex-post efficient and interim efficient, but not ex-ante efficient. And the mechanism

$$x(\theta) = \begin{array}{c|c|c|} & a & b \\ \hline a & 1/2,1/2 & 1/2,1/2 \\ \hline b & 1/2,1/2 & 1/2,1/2 \\ \hline \end{array}$$

is ex-post efficient, interim efficient, and ex-ante efficient.

Suppose that the cost of implementing decision $x \in X$ is given by $C_0(x)$. To be "practicable," a mechanism should often also be budget balanced.

Definition. A direct revelation mechanism $\{x(\theta), t(\theta)\}$ is ex-post budget balanced if

$$\sum_{i=1}^{n} t_i(\theta) = -C_0(x(\theta)) \qquad \text{for every } \theta.$$

Remark. The notation is a little awkward because t_i is the payment *to* player i. Observe that unless $\sum_{i=1}^{n} t_i(\theta) \leq -C_0(x(\theta))$ the mechanism would not collect enough payment to cover the cost of implementing $x(\theta)$; and unless $\sum_{i=1}^{n} t_i(\theta) \geq -C_0(x(\theta))$ whatever extra payment has been collected has to be taken out of the system so as not to distort incentives.

Remark. In some cases, if the principal has access to a well functioning credit market, it may be possible to replace ex-post budget balance with ex-ante budget balance, or the weaker requirement that:

$$E_\theta\left[\sum_{i=1}^{n} t_i(\theta)\right] = -E_\theta\left[C_0(x(\theta))\right].$$

Another constraint that is often relevant is voluntary participation, or individual rationality.

Definition. A direct revelation mechanism $\{x(\theta), t(\theta)\}$ is (interim) individually rational if

$$E_{\theta_{-i}}\left[u_i(x(\theta), t_i(\theta), \theta_i)\right] \geq 0 \qquad \text{for every } i \in N, \text{ and } \theta_i \in \Theta_i.$$

Observe that the right-hand-side of the IR constraint may depend on the particular problem that is studied, and may therefore be different from zero.

Ex-post and ex-ante IR are defined analogously.

4.2. Groves Mechanisms

Groves mechanisms have been discovered by Vickrey (1961), Clarke (1971) and Groves (1973). Vickrey and Clarke have each described an example of a particular Groves mechanism, and Groves identified the entire class of such mechanisms. Groves mechanisms implement the ex-post efficient outcome in dominant strategies in quasi-linear private values environments. (Contrast with Gibbard and Satterthwaite's Impossibility Theorem).

Definition. A direct revelation mechanism $\{x(\theta), t(\theta)\}$ is incentive compatible in dominant strategies if truth-telling is a dominant strategy, or

$$u_i(x(\theta), t_i(\theta), \theta_i) \geq u_i\left(x\left(\widehat{\theta}_i, \theta_{-i}\right), t_i\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_i\right) \qquad \text{for every } \theta \in \Theta, \text{ and } \widehat{\theta}_i \in \Theta_i.$$

Denote the ex-post efficient decision by $x^*(\theta) \in \underset{x \in X}{\operatorname{argmax}} \sum_{i=0}^{n} V_i(x, \theta_i)$. Define

$$t_i^*\left(\widehat{\theta}\right) = \sum_{j \neq i} V_j\left(x^*\left(\widehat{\theta}\right), \widehat{\theta}_j\right) + \tau_i\left(\widehat{\theta}_{-i}\right)$$

where $\tau_i(\cdot)$ is an arbitrary function of $\widehat{\theta}_{-i}$.

Definition. A direct revelation mechanism $\{x^*(\theta), t^*(\theta)\}$ is called a Groves mechanism. By changing the function τ_i it is possible to span the entire collection of Groves mechanisms.

Definition. A Vickrey-Clarke-Groves (VCG) mechanism is a Groves mechanism where

$$t_i^*\left(\widehat{\theta}\right) = \sum_{j \neq i} V_j\left(x^*\left(\widehat{\theta}\right), \widehat{\theta}_j\right) - \sum_{j \neq i} V_j\left(x^*\left(\overline{\theta}_i, \widehat{\theta}_{-j}\right), \widehat{\theta}_j\right)$$

for some $\overline{\theta}_i \in \Theta_i$. Usually $\overline{\theta}_i$ is i's lowest possible type, or the type that contributes least to social welfare. In auctions for example $\overline{\theta}_i = 0$. So in VCG mechanisms the payment to each player is equal to the player's contribution to social surplus, not taking the player's own payoff into account.

Proposition. A Groves mechanism $\{x^*(\theta), t^*(\theta)\}$ is incentive compatible in dominant strategies and ex-post efficient.

Proof. Because x^* is ex-post efficient by definition, it is enough to show that $\{x^*(\theta), t^*(\theta)\}$ is incentive compatible in dominant strategies. Suppose to the contrary that some type θ_i of player i strictly prefers to announce $\widehat{\theta}_i$ instead of θ_i for some types θ_{-i} of the other players. It follows that

$$\sum_k V_k\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_k\right) + \tau_i(\theta_{-i}) = V_i\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_i\right) + \sum_{j \neq i} V_j\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_j\right) + \tau_i(\theta_{-i})$$

$$> V_i(x^*(\theta_i, \theta_{-i}), \theta_i) + \sum_{j \neq i} V_j(x^*(\theta_i, \theta_{-i}), \theta_j) + \tau_i(\theta_{-i}) = \sum_k V_k(x^*(\theta), \theta_k) + \tau_i(\theta_{-i}),$$

which contradicts the assumption that $x^*(\theta_i, \theta_{-i}) \in \underset{x \in X}{\operatorname{argmax}} \sum_{i=0}^{n} V_i(x, \theta_i)$. ■

Intuitively, the idea of a Groves mechanism is to choose each agent's transfer t_i in such a way that agent i's payoff is the same as the total surplus to all the agents (given their reports) up to a constant. Because agent i already internalizes its own surplus, it suffices to set the transfer equal to the total surplus minus its own surplus, up to a constant. Note that i's report affects its payment only through its effect on the decision $x^*(\theta)$. Hence i's payment is equal to the externality it imposes on the other players, up to a constant. For this reason, the payments under Groves mechanisms are sometimes referred to as "externality payments."

Example. A sealed bid second price auction is a VCG mechanism. Suppose that there are n bidders for an object and a seller, and that it is commonly known that the value of the object for the seller is zero.

The set of consequences X is given by:

$$X = \left\{ (q_0, q_1, ..., q_n) : q_i \geq 0 \quad \forall i \in N, \text{ and } \sum_{i=0}^{n} q_i = 1 \right\}.$$

The bidders' types are their willingness to pay for the object.

The ex-post efficient decision rule is given by $x^*(\theta) = \left(0, ..., 0, \underset{i\text{th place}}{1}, 0, ..., 0\right)$ if i's willingness for the object is positive and is the highest.

In a VCG mechanism, the bidder with the highest valuation, suppose it is i, should win the object, so

$$\begin{aligned} t_i^*\left(\widehat{\theta}\right) &= \sum_{j \neq i} V_j\left(x^*\left(\widehat{\theta}\right), \widehat{\theta}_j\right) + \tau_i\left(\widehat{\theta}_{-i}\right) \\ &= \tau_i\left(\widehat{\theta}_{-i}\right), \end{aligned}$$

because $V_j = 0$ for every $j \neq i$. For all other bidders j,

$$\begin{aligned} t_j^*\left(\widehat{\theta}\right) &= \sum_{k \neq j} V_k\left(x^*\left(\widehat{\theta}\right), \widehat{\theta}_k\right) + \tau_j\left(\widehat{\theta}_{-j}\right) \\ &= \max_{k \neq j}\left\{\widehat{\theta}_k\right\} + \tau_j\left(\widehat{\theta}_{-j}\right) \end{aligned}$$

In a VCG mechanism,

$$\tau_j\left(\widehat{\theta}_{-j}\right) = -\max_{k \neq j}\left\{\widehat{\theta}_k\right\},$$

which means that the winner pays $\max_{k \neq j}\left\{\widehat{\theta}_k\right\}$, which is equal to the value and bid of the bidder with the second highest valuation, and losers pay nothing. This is Vickrey's famous second price auction. The highest bidder wins and pays the second highest bid. Losers pay nothing, and it is unimportant how ties are resolved.

Remark. Green and Laffont (1977) show that the converse of the proposition above is also true in the following sense: if the type space is "rich enough" in the sense that no restrictions are imposed on the set of agents' types Θ, or that $u_i(x, t_i, \theta_i) = V_i(x, \theta_i) + t_i$ ranges over all the possible functions V_i as θ_i ranges over Θ_i, then if a mechanism implements the ex-post efficient outcome in dominant strategies, then it is a Groves mechanism. See MWG (Proposition 23.C.5) for a straightforward proof.

Remark. Another important result of Green and Laffont (1977) is that if the set of agents' types is sufficiently rich (so that agents may hold any payoff function V_i), then no Groves mechanism is ex-post budget balanced. For example, consider a public good problem with two agents and two types each (high and low), and denote the cost of the public good by c. Suppose that the public good should be efficiently provided unless both agents' have low types or low willingness to pay.

Denote agent i's payment under a Groves mechanism by $t_i\left(\widehat{\theta}_1, \widehat{\theta}_2\right)$. Agent i's payment under a Groves mechanism is

$$t_i\left(\widehat{\theta}\right) = \begin{cases} \widehat{\theta}_j + \tau_i\left(\widehat{\theta}_j\right) & \text{if } \widehat{\theta}_1 + \widehat{\theta}_2 \geq c \\ \tau_i\left(\widehat{\theta}_j\right) & \text{if } \widehat{\theta}_1 + \widehat{\theta}_2 < c \end{cases}$$

and the sum of the agents' payments is given by

$$t_i\left(\widehat{\theta}\right) + t_j\left(\widehat{\theta}\right) = \begin{cases} \widehat{\theta}_i + \widehat{\theta}_j + \tau_j\left(\widehat{\theta}_i\right) + \tau_i\left(\widehat{\theta}_j\right) & \text{if } \widehat{\theta}_1 + \widehat{\theta}_2 \geq c \\ \tau_j\left(\widehat{\theta}_i\right) + \tau_i\left(\widehat{\theta}_j\right) & \text{if } \widehat{\theta}_1 + \widehat{\theta}_2 < c \end{cases}$$

The definition of i's payment under a Groves mechanism implies:

$$\begin{array}{rrcl} (1) & t_1(H, L) - t_1(L, L) & = & L \\ (2) & t_1(H, H) - t_1(L, H) & = & 0 \\ (3) & t_2(L, H) - t_2(L, L) & = & L \\ (4) & t_2(H, H) - t_2(H, L) & = & 0 \end{array}$$

and Ex-post budget balance requires that:

$$\begin{array}{rrcl} (5) & t_1(L, L) + t_2(L, L) & = & 0 \\ (6) & t_1(L, H) + t_2(L, H) & = & c \\ (7) & t_1(H, L) + t_2(H, L) & = & c \\ (8) & t_1(H, H) + t_2(H, H) & = & c \end{array}$$

Algebraic manipulation reveals that these eight equations are inconsistent. This can be seen by manipulating the matrix that describes these equations as follows:

$$\left(\begin{array}{cccccccccc} & 1LL & 1LH & 1HL & 1HH & 2LL & 2LH & 2HL & 2HH & \\ (1) & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & L \\ (2) & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ (3) & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & L \\ (4) & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ (5) & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ (6) & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & c \\ (7) & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & c \\ (8) & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & c \end{array}\right)$$

(1) + (5) implies the vector

$$(9) \qquad (0,0,1,0,1,0,0,0,L)$$

(2) + (6) implies the vector

$$(10) \qquad (0,0,0,1,0,1,0,0,c)$$

(9) − (7) implies the vector

$$(11) \qquad (0,0,0,0,1,0,-1,0,L-c)$$

(10) − (8) implies the vector

$$(12) \qquad (0,0,0,0,0,1,0,-1,0)$$

(11) + (3) implies the vector

$$(13) \qquad (0,0,0,0,0,1,-1,0,2L-c)$$

(12) − (13) implies the vector

$$(14) \qquad (0,0,0,0,0,0,1,-1,c-2L)$$

Finally, (14) + (4) implies the vector

$$(14) \qquad (0,0,0,0,0,0,0,0,c-2L)\,,$$

which implies that $c - 2L = 0$. A contradiction to the assumption that $2L < c$.

This failure of budget balance is probably the main reason that Groves mechanisms are so seldom used in practice. In fact, when they are used, it is usually in contexts where budget balance is unimportant because there is an agent who acts as a "budget breaker."

It is straightforward to make a Groves mechanism ex-ante budget balanced by adjusting the functions $\tau_i(\cdot)$, but at the possible price of violating the agents' individual rationality constraints.[14]

[14] Observe that the second price auction is dominant strategy implementable and budget balanced. However, the seller has no private information so the "richness" condition is violated.

4.3. AAGV mechanisms

AAGV mechanisms are named after Arrow (1979), and d'Aspremont and Gérard-Varet (1979), who discovered these mechanisms independently. AAGV mechanisms implement the ex-post efficient outcome and are ex-post budget balanced. However, they are only Bayesian incentive compatible, not incentive compatible in dominant strategies.

The idea is that instead of being paid the surplus of the other agents based on their reports as in a Groves mechanism, each agent is paid the *expected* value of the other agents' surpluses based on its own report. Then each agent again internalizes the social surplus and has no incentive to distort the decision by manipulating its announcement, and the functions $\{\tau_i(\cdot)\}_{i\in N}$ can be chosen to ensure budget balance.

Let

$$t_i^*\left(\widehat{\theta}\right) = E_{\theta_{-i}}\left[\sum_{j\neq i} V_j\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_j\right)\right] + \tau_i\left(\widehat{\theta}_{-i}\right).$$

Observe that the first term is independent of the other players' reports, and the second term is independent of i's own report, which implies that it does not affect i's incentives.

Lemma. A direct revelation mechanism $\{x^*(\theta), t^*(\theta)\}$ is Bayesian incentive compatible.

Proof. We have to show that $\widehat{\theta}_i = \theta_i$ maximizes i's expected payoff, or

$$E_{\theta_{-i}}\left[V_i\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_i\right) + t_i^*\left(\widehat{\theta}_i, \theta_{-i}\right)\right],$$

which is equal to

$$E_{\theta_{-i}}\left[V_i\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_i\right) + \sum_{j\neq i} V_j\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_j\right)\right]$$

up to a constant. Observe that the function $\tau_i\left(\widehat{\theta}_{-i}\right)$ is independent of i's report and does not affect i's incentives. The same reasoning that applied in the case of Groves mechanisms implies that $\widehat{\theta}_i = \theta_i$ maximizes the function

$$V_i\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_i\right) + \sum_{j\neq i} V_j\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_j\right)$$

for each possible realization of θ_{-i}. It therefore follows that $\widehat{\theta}_i = \theta_i$ maximizes the expectation

$$E_{\theta_{-i}}\left[V_i\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_i\right) + \sum_{j\neq i} V_j\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_j\right)\right]$$

as well. ■

. . . .

Example. Consider the "AAGV version" of the 2nd price auction. In a 2nd price auction, for the winner i:

$$\sum_{j\neq i} V_j\left(x^*\left(\widehat{\theta}_i,\theta_{-i}\right),\theta_j\right)=0$$

for loser j

$$\sum_{k\neq j} V_k\left(x^*\left(\widehat{\theta}_k,\theta_{-k}\right),\theta_k\right)=\max\left\{\theta_k\right\}$$

and for all bidders

$$\tau_k\left(\theta_{-k}\right)=-\underset{j\neq k}{\max}\theta_j.$$

In the AAGV version of the 2nd price auction

$$\begin{aligned} t_i^*\left(\widehat{\theta}_i\right) &= E_{\theta_{-i}}\left[\sum_{j\neq i} V_j\left(x^*\left(\widehat{\theta}_i,\theta_{-i}\right),\theta_j\right)-\underset{j\neq i}{\max}\widehat{\theta}_j\right] \\ &= \Pr\left(\underset{j\neq i}{\max}\theta_j\leq\widehat{\theta}_i\right)\cdot 0+\Pr\left(\underset{j\neq i}{\max}\theta_j>\widehat{\theta}_i\right)\cdot E_{\theta_{-i}}\left[\underset{j\neq i}{\max}\widehat{\theta}_j\middle|\underset{j\neq i}{\max}\theta_j>\widehat{\theta}_i\right]-E_{\theta_{-i}}\left[\underset{j\neq i}{\max}\widehat{\theta}_j\right]\end{aligned}$$

For example, if $n=2$ and the bidder's valuations are uniformly distributed on the unit interval, this is equal to:

$$\left(1-\widehat{\theta}_i\right)\cdot\frac{1+\widehat{\theta}_i}{2}-\frac{1}{2}.$$

The expected payoff to a bidder who reports $\widehat{\theta}_i$ is therefore given by

$$\widehat{\theta}_i\cdot\theta_i+\left(1-\widehat{\theta}_i\right)\cdot\frac{1+\widehat{\theta}_i}{2}-\frac{1}{2}$$

The first-order condition is:

$$\theta_i+\frac{1-\widehat{\theta}_i}{2}-\frac{1+\widehat{\theta}_i}{2}=0$$

if and only if $\widehat{\theta}_i=\theta_i$.

But if it is known that the other player reported $\widehat{\theta}_j=.5$, and say that my valuation is .25, then reporting truthfully yields

$$(1-.25)\cdot\frac{1+.25}{2}-\frac{1}{2}=-.031$$

while reporting a little over .5 yields

$$.25+(1-.5)\cdot\frac{1+.5}{2}-\frac{1}{2}=.125$$

....

We now show how the functions $\tau_i\left(\widehat{\theta}_{-i}\right)$ can be chosen in such a way that $\left\{x^*\left(\theta\right),t^*\left(\theta\right)\right\}$ satisfies budget balance, or such that $\sum_{i=1}^n t_i^*\left(\theta\right)=C_0\left(x^*\left(\theta\right)\right)$ for every θ.

Suppose first that $C_0(x) \equiv 0$. Define

$$T_i\left(\widehat{\theta}_i\right) = E_{\theta_{-i}}\left[\sum_{j\neq i} V_j\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_j\right)\right]$$

and let

$$\tau_i\left(\widehat{\theta}_{-i}\right) = -\frac{1}{n-1}\sum_{j\neq i} T_j\left(\widehat{\theta}_j\right).$$

Observe that this implies that $\sum_{i=1}^n t_i^*(\theta) = \sum_{i=1}^n \left(T_i\left(\widehat{\theta}_i\right) - \frac{1}{n-1}\sum_{j\neq i} T_j\left(\widehat{\theta}_j\right)\right) = 0$ for every θ as required.

Suppose now that $C_0(x)$ is an arbitrary function. Consider the "fictional problem" where the agents' utility functions are given by

$$\widetilde{V}_i(x, \theta_i) = V_i(x, \theta_i) - \frac{C_0(x)}{n}$$

and the principal's cost is given by $\widetilde{C}_0(x) \equiv 0$. Compute the IC and BB transfers for this fictional problem, $\widetilde{t}_i(\cdot)$, and let

$$t_i^*(\cdot) = \widetilde{t}_i(\cdot) - \frac{C_0(x^*(\cdot))}{n}.$$

We show that the transfers $\{t_i^*(\cdot)\}$ implement the efficient decision with budget balance in the original problem. Budget balance follows from the fact that

$$\begin{aligned}\sum_{i=1}^n t_i^*\left(\widehat{\theta}\right) &= \sum_{i=1}^n\left(\widetilde{t}_i(\cdot) - \frac{C_0(x^*(\cdot))}{n}\right) \\ &= -C_0\left(x^*\left(\widehat{\theta}\right)\right)\end{aligned}$$

because $\sum_{i=1}^n \widetilde{t}_i\left(\widehat{\theta}\right) = 0$ for all $\widehat{\theta}$. The fact that

$$x^* \in \arg\max\sum_{i\in N}\widetilde{V}_i(x, \theta_i)$$

implies that

$$x^* \in \arg\max\sum_{i\in N} V_i(x, \theta_i) - C_0(x)$$

so that $\{x^*(\theta), t^*(\theta)\}$ is ex-post efficient. Finally, incentive compatibility follows from the fact that

$$\begin{aligned}\widetilde{V}_i\left(x^*\left(\widehat{\theta}\right), \theta_i\right) + \widetilde{t}_i\left(\widehat{\theta}\right) &= \left[V_i(x, \theta_i) - \frac{C_0\left(x^*\left(\widehat{\theta}\right)\right)}{n}\right] + \left[t_i^*\left(\widehat{\theta}\right) + \frac{C_0\left(x^*\left(\widehat{\theta}\right)\right)}{n}\right] \\ &= V_i\left(x^*\left(\widehat{\theta}\right), \theta_i\right) + t_i^*\left(\widehat{\theta}\right)\end{aligned}$$

for every $\widehat{\theta}$ and θ_i. So, the fact that truthful reporting is an equilibrium in the fictional problem under transfers $\widetilde{t}$, implies that it is also an equilibrium in the original problem under the transfers t^*.

Remark 1. Notice that the same argument would work also if players' types are not independent. The function T has to be defined relative to the reported type as before, and everything works in the same way.

Remark 2. Although AAGV mechanisms can be made to satisfy ex-ante IR, they might violate interim IR in "rich enough" environments. In the next section, we show that interim IR may be incompatible with ex-post efficiency. For some problems, there are no ex-post efficient budget balanced mechanisms that satisfy interim individual rationality.

Remark 3. Although they are budget balanced, AAGV mechanisms are a lot less practicable than Groves mechanisms because they depend on the prior and preferences of the agents, which are difficult if not impossible to verify in practice.

4.4. Optimal Monopolistic Price Discrimination

Consider first the two type case and then the continuum case.

4.5. Bilateral Bargaining under Asymmetric Information

Some mechanism design problems (such as those encountered in theory of auctions) admit the existence of ex-post efficient, ex-post budget balanced, and individually rational mechanisms, but others do not. The most famous result that establishes the impossibility of ex-post efficient, budget balanced, and individually rational mechanisms is due to Myerson and Satterthwaite (1983). Here, we consider a simpler 2×2 version of their model that is due to Matsuo (1989).

Consider the following mechanism design problem. There is a buyer and a seller. The buyer is interested in buying an object which is owned by the seller. The buyer's value (type) for the object is either v_1 or v_2. The seller's reservation value (cost, type) is either c_1 or c_2. Suppose that each profile of types is equally likely (i.e., the buyer's and seller's types are independent, and both the buyer and the seller are equally likely to be of either type). Suppose that the buyer's and seller's types are "symmetric" in the following sense:

$$\overset{c_1}{\cdot} \longleftarrow A \longrightarrow \overset{v_1}{\cdot} \longleftarrow -D- \longrightarrow \overset{c_2}{\cdot} \longleftarrow A \longrightarrow \overset{v_2}{\cdot}$$

[explain the sense in which this covers all the "interesting" cases]

A bargaining game is any game form that specifies a message set for each player and a mapping from message profiles to outcomes, or the probability with which the buyer obtains the object and the price it pays. The revelation principle implies that if we are interested in studying the range of equilibrium outcomes, then no loss of generality is entailed by restricting attention to incentive compatible and individually rational direct revelation mechanisms. Individual rationality follows from the fact that trade is voluntary. Each trader may refuse to participate in the mechanism if it does not give it a nonnegative expected payoff.

A direct revelation mechanism is composed of two functions: $t(c,v)$, which described the expected payment from the buyer to the seller when their types are given by (c,v), and $q(c,v)$, which described the probability with which the buyer obtains the object when types are given by (c,v). A direct revelation mechanism is thus characterized by the following eight-tuple $\langle q_1,q_2,q_3,q_4,t_1,t_2,t_3,t_4\rangle$:

$$\begin{array}{c|c|c|} q(c,v) & c_1 & c_2 \\ \hline v_1 & q_1 & q_2 \\ \hline v_2 & q_3 & q_4 \\ \hline \end{array} \qquad \begin{array}{c|c|c|} t(c,v) & c_1 & c_2 \\ \hline v_1 & t_1 & t_2 \\ \hline v_2 & t_3 & t_4 \\ \hline \end{array}$$

Let

$$\begin{aligned} u(v',v) &\equiv E_c\left[vq(c,v') - t(c,v')\right] \\ h(c',c) &\equiv E_v\left[t(c',v) - cq(c',v)\right] \end{aligned}$$

Definition. A direct revelation mechanism $\langle t,q\rangle$ is incentive compatible and individually rational if and only if

$$\begin{aligned} U(v) &\equiv u(v,v) \geq u(v',v) \qquad \forall v,v' \in \{v_1,v_2\} && (\mathrm{IC}-\mathrm{B}) \\ H(c) &\equiv h(c,c) \geq h(c',c) \qquad \forall c,c' \in \{c_1,c_2\} && (\mathrm{IC}-\mathrm{S}) \\ U(v) &\geq 0 \qquad \forall v \in \{v_1,v_2\} && (\mathrm{IR}-\mathrm{B}) \\ H(c) &\geq 0 \qquad \forall c \in \{c_1,c_2\} && (\mathrm{IR}-\mathrm{S}) \end{aligned}$$

Remark. Alternatively, instead of focusing our attention on incentive compatible direct revelation mechanisms we could consider the probability of trade and expected payment in a BNE and denote those by $q(c,v)$ and $t(c,v)$. In this case, IC and IR would follow from the fact that what we consider is a Bayesian Nash equilibrium. The fact that these two approaches are identical illustrates, or rather is a proof of, the revelation principle in this context.

Definition. A direct revelation mechanism $\langle t,q\rangle$ is ex-post efficient if $q(c,v)=1$ whenever $v>c$, or in matrix form:

$$\begin{array}{c|c|c|} q(c,v) & c_1 & c_2 \\ \hline v_1 & 1 & 0 \\ \hline v_2 & 1 & 1 \\ \hline \end{array}$$

Proposition. There exists an ex-post efficient incentive compatible and individually rational direct revelation mechanism $\langle t,q\rangle$ if and only if

$$c_2 - v_1 \leq (v_2 - c_2) + (v_1 - c_1) \qquad \text{(i.e., iff } D \leq 2A\text{)}$$

Proof. We first show that if $D > 2A$, then no ex-post efficient mechanism exists. Suppose to the contrary that such a mechanism exists. The IC constraint for v_2 implies

$$U(v_2) \geq v_2 E_c[q(c, v_1)] - E_c[t(c, v_1)] + v_1 E_c[q(c, v_1)] - v_1 E_c[q(c, v_1)]$$

or

$$U(v_2) \geq U(v_1) + (v_2 - v_1) E_c[q(c, v_1)]. \tag{1}$$

Similarly, the IC constraint for c_1 implies

$$H(c_1) \geq H(c_2) + (c_2 - c_1) E_v[q(c_2, v)]. \tag{2}$$

Ex-post efficiency implies that $E_c[q(c, v_1)] = E_v[q(c_2, v)] = \frac{1}{2}$. Plug this into (1) and use IR – B to get

$$U(v_2) \geq \frac{v_2 - v_1}{2}.$$

Similarly, (2) and IR – S imply

$$H(c_1) \geq \frac{c_2 - c_1}{2}.$$

Since v_2 and c_1 each occurs with probability $\frac{1}{2}$, the sum of the ex-ante expected payoffs is at least

$$\frac{1}{2} \cdot (U(v_2) + H(c_1)) \geq \frac{v_2 - v_1}{4} + \frac{c_2 - c_1}{4} = \frac{A + D}{2}.$$

But the maximum ex-ante surplus is

$$\frac{1}{4} \cdot ((v_2 - c_2) + (v_2 - c_1) + (v_1 - c_1) + 0) = \frac{4A + D}{4},$$

which is smaller than $\frac{A+D}{2}$ if $D > 2A$. A contradiction.

Next, we show that if $D \leq 2A$ then the following direct revelation mechanism is incentive compatible, individually rational, and ex-post efficient: $q(c, v) = 1$ if and only if $c < v$, and $t(c, v)$ is given by the following matrix:

$t(c,v)$	c_1	c_2
v_1	v_1	0
v_2	$\frac{v_2+c_1}{2}$	c_2

Intuition: The problem is to get the "high value" types v_2 and c_1 to reveal their types. To accomplish this goal, these types are given the most favorable prices possible if they reveal their identity.

Ex-post efficiency, IR, and IC for v_1 and for c_2 are immediate. IC for v_2 follows from

$$\frac{1}{2} \cdot \underbrace{\left(v_2 - \frac{v_2 + c_1}{2}\right)}_{A + \frac{D}{2}} + \frac{1}{2} \cdot \underbrace{(v_2 - c_2)}_{A} \geq \frac{1}{2} \cdot \underbrace{(v_2 - v_1)}_{A+D}.$$

$$\begin{aligned} 2A+\frac{D}{2} &\geq A+D \\ &\iff 2A\geq D. \end{aligned}$$

IC for c_1 follows similarly. ■

Remark. The ex-post efficient mechanism for the case where $D \leq 2A$ is not incentive compatible when $D > 2A$ because in this case v_2 and c_1 would rather report v_1 and c_2, respectively. If $D \leq 2A$, the benefit that v_2 and c_1 would obtain from reporting v_1 and c_2, respectively, is too small relative to the lower probability they would get to trade, and so ex-post efficiency is possible.

The remark above suggests that by replacing the ex-post efficient allocation rule $q(\cdot,\cdot)$ with the following allocation function

$q^*(c,v)$	c_1	c_2
v_1	p	0
v_2	1	p

where $p = \min\left\{\frac{2A+D}{2D},1\right\}$, or as high as possible such that IC still holds, it is possible to retain IC. Indeed, it can be shown that the mechanism $\langle q^*, t\rangle$ with this p and the following t

$t(c,v)$	c_1	c_2
v_1	pv_1	0
v_2	$\frac{v_2+c_1}{2}$	pc_2

maximizes the ex-ante surplus of the buyer and seller. The argument is as follows. The mechanism design problem is to find a mechanism $\langle q_1,q_2,q_3,q_4,t_1,t_2,t_3,t_4\rangle$ where $q_i \in [0,1]$ and $t_i \in \mathbb{R}$ for $i \in \{1,...,4\}$ that maximizes the objective function:

$$\frac{1}{4}\left(q_1(v_1-c_1)+q_2(v_1-c_2)+q_3(v_2-c_1)+q_4(v_2-c_2)\right)$$

subject to IC for the buyer

$$\begin{aligned} v_1E_c\underbrace{[q(c,v_1)]}_{.5(q_1+q_2)} - E_c\underbrace{[t(c,v_1)]}_{.5(t_1+t_2)} &\geq v_1\underbrace{E_c[q(c,v_2)]}_{.5(q_3+q_4)} - E_c\underbrace{[t(c,v_2)]}_{.5(t_3+t_4)} \\ v_2E_c\underbrace{[q(c,v_2)]}_{.5(q_3+q_4)} - E_c\underbrace{[t(c,v_2)]}_{.5(t_3+t_4)} &\geq v_2E_c\underbrace{[q(c,v_1)]}_{.5(q_1+q_2)} - E_c\underbrace{[t(c,v_1)]}_{.5(t_1+t_2)} \end{aligned}$$

and seller

$$\begin{aligned} E_c\underbrace{[t(c_1,v)]}_{.5(t_1+t_3)} - c_1E_c\underbrace{[q(c_1,v)]}_{.5(q_1+q_3)} &\geq E_c\underbrace{[t(c_2,v)]}_{.5(t_2+t_4)} - c_1E_c\underbrace{[q(c_2,v)]}_{.5(q_2+q_4)} \\ E_c\underbrace{[t(c_2,v)]}_{.5(t_2+t_4)} - c_2E_c\underbrace{[q(c_2,v)]}_{.5(q_2+q_4)} &\geq E_c\underbrace{[t(c_1,v)]}_{.5(t_1+t_3)} - c_2E_c\underbrace{[q(c_1,v)]}_{.5(q_1+q_3)} \end{aligned}$$

and IR for the buyer and seller. Observe that this is a linear programming problem, and as such, can be solved using known methods (and software). We proceed to present a direct solution below.

Rewrite the IC and IR constraints as:

$$\begin{aligned} v_1(q_1+q_2)-(t_1+t_2) &\geq v_1(q_3+q_4)-(t_3+t_4) \\ v_2(q_3+q_4)-(t_3+t_4) &\geq v_2(q_1+q_2)-(t_1+t_2) \\ (t_1+t_3)-c_1(q_1+q_3) &\geq (t_2+t_4)-c_1(q_2+q_4) \\ (t_2+t_4)-c_2(q_2+q_4) &\geq (t_1+t_3)-c_2(q_1+q_3) \end{aligned}$$

and

$$\begin{aligned} v_1(q_1+q_2)-(t_1+t_2) &\geq 0 \\ v_2(q_3+q_4)-(t_3+t_4) &\geq 0 \\ (t_1+t_3)-c_1(q_1+q_3) &\geq 0 \\ (t_2+t_4)-c_2(q_2+q_4) &\geq 0 \end{aligned}$$

Step 1. We solve a relaxed problem in which we ignore the IC constraints of v_1 and c_2 and the IR constraints of v_2 and c_1. We will later show that the solution satisfies these constraints. The remaining IC and IR constraints are:

$$\begin{aligned} v_2(q_3+q_4)-(t_3+t_4) &\geq v_2(q_1+q_2)-(t_1+t_2) \\ (t_1+t_3)-c_1(q_1+q_3) &\geq (t_2+t_4)-c_1(q_2+q_4) \end{aligned}$$

and

$$\begin{aligned} v_1(q_1+q_2)-(t_1+t_2) &\geq 0 \\ (t_2+t_4)-c_2(q_2+q_4) &\geq 0 \end{aligned}$$

Step 2. In the optimal solution $q_3 = 1$. If not, then increase q_3 by d and t_3 by md, $v_1 < m < c_2$. Observe that this increases the value of the objective function and that the choice of m implies that the IC and IR constraints are not violated.

Step 3. In the optimal solution $q_2 = 0$. If not, then decrease q_2 by d and t_2 by md, $v_1 < m < c_2$. Observe that this increases the value of the objective function and that the choice of m implies that the IC and IR constraints are not violated. This implies that the constraints can be further simplified as follows:

$$\begin{aligned} v_2(1+q_4)-(t_3+t_4) &\geq v_2q_1-(t_1+t_2) \\ (t_1+t_3)-c_1(1+q_1) &\geq (t_2+t_4)-c_1q_4 \end{aligned}$$

and

$$\begin{aligned} v_1q_1-(t_1+t_2) &\geq 0 \\ (t_2+t_4)-c_2q_4 &\geq 0 \end{aligned}$$

Step 4. The remaining IC and IR constraints are binding in the optimal solution. If not, then in the first IC constraint increase q_1 by d and t_1 by $v_1 d$. In the second IC constraint increase q_4 by d and t_4 by $v_1 d$. In the first IR constraint increase q_1 by d and t_1 by $v_2 d$. In the second IR constraint increase q_4 by d and t_4 by $v_1 d$.

Step 5. The problem now becomes:

Maximize the objective function:

$$q_1 (v_1 - c_1) + q_4 (v_2 - c_2)$$

subject to

$$\begin{aligned} v_2 (1 + q_4) - (t_3 + t_4) &= v_2 q_1 - (t_1 + t_2) \\ (t_1 + t_3) - c_1 (1 + q_1) &= (t_2 + t_4) - c_1 q_4 \end{aligned}$$

and

$$\begin{aligned} v_1 q_1 - (t_1 + t_2) &= 0 \\ (t_2 + t_4) - c_2 q_4 &= 0 \end{aligned}$$

Observe that if $q_1, q_4, t_1 = x, t_2 = y, t_3 = z, t_4 = w$ is a solution to the problem, then so is $q_1, q_4, t_1 = x + y, t_2 = 0, t_3 = z - y, t_4 = w + y$. This means that we may restrict our attention to solutions in which $t_2 = 0$, which simplifies the IR constraints to:

$$\begin{aligned} t_1 &= v_1 q_1 \\ t_4 &= c_2 q_4 \end{aligned}$$

Upon plugging $t_2 = 0$ and these two equations into the IC constraints, we obtain

$$\begin{aligned} v_2 (1 + q_4) - (t_3 + c_2 q_4) &= (v_2 - v_1) q_1 \\ (v_1 q_1 + t_3) - c_1 (1 + q_1) &= (c_2 - c_1) q_4 \end{aligned}$$

Summing these two equations, we get

$$2A + D = D (q_1 + q_4)$$

Thus, the objective function is maximized at a point where $q_1 + q_4 = \dfrac{2A + D}{D}$, or in particular where $q_1 = q_4 = \dfrac{2A + D}{2D}$ which is smaller than 1 if $D > 2A$.

Remark. Myerson and Satterthwaite (1983) considered a similar model to the 2×2 model presented here with a continuum of buyer's and seller's types. They showed that if the supports of the buyer's and seller's distributions overlap, then ex-post efficiency is impossible. Notice that in the case analyzed here, if $D \leq 2A$ then ex-post efficiency is possible in spite of the "overlapping" supports.

4.6. Double Auctions

A double auction is a trade mechanism in which buyers and sellers are each required to post bid and ask prices. These bids and asks are used to construct demand and supply functions, respectively, and trade takes places at a market clearing price[15] among the buyers who bid at or above the price and sellers who bid below or at the price, with rationing on the long side of the market, among low paying buyers or high asking sellers, if needed.

The double auction mechanism is attractive because it is *simple*. It does not depend on the players' payoff functions and beliefs, and it does not employ integer games, etc.

Myerson and Satterthwaite (1983) showed that in a setting with just one buyer and one seller the 1/2-double auction has an equilibrium that optimal in the sense of maximizing ex-ante efficiency subject to IC and IR.

In a series of subsequent papers, Satterthwaite and Williams (together with Rustichini and others) showed that as the number of traders increases the bids and asks in any non trivial equilibrium converge to the traders' true willingness to pay and reservation values. Therefore, equilibria are asymptotically ex-post efficient. They have also showed that the double-auction is asymptotically worst-case optimal. That is, every other mechanism has an equilibrium that is not more efficient than an equilibrium of the double-auction for some distribution of traders' types.

This work and subsequent generalizations provide a "micro" or "strategic" foundation for "competitive" behavior and for the first welfare theorem.

4.7. Private Values Auctions

This lecture is based on Krishna's "Auction Theory," and on Milgrom's "Putting Auction Theory to Work."

4.7.1. The Symmetric Model

A Single object is offered for sale.
There are N potential buyers or bidders who are interested in buying the object.[16]
The valuation, or willingness to pay, of bidder i for the object is X_i. Since we analyze an auction as a Bayesian game, X_i also describes bidder i's type.

The X_i are i.i.d. and distributed according to an increasing distribution function F with a continuous density f on $[0, \omega]$. The support of F may be unbounded $(\omega = \infty)$.

Bidder i knows the realization x_i of the random variable X_i. It believes that other bidder's valuations are independently distributed according to F. [explain how this is different from a common values setting]

[15] Typically, there would be an interval $[a, b]$ of market-clearing prices. A k-double auction where $k \in [0, 1]$ refers to a double auction in which the price is equal to $ka + (1 - k)\, b$.

[16] The number of bidders is exogenous. Bulow and Klemperer (AER, 1996) show that for a seller who employs a "standard" auction, attracting one more bidder is more valuable to the seller than employing the optimal auction.

Each bidder i is a risk neutral expected utility maximizer who seeks to maximize its expected payoff, which is given by

$$q_i \cdot x_i - p_i$$

where q_i is the probability that bidder i wins the object and p_i is bidder i's expected payment.

Bidders are not subject to liquidity or budget constraints.

All the above, except for the realization of the bidders' types is commonly known among the bidders.

4.7.2. (Sealed Bid) First Price Auction

Description. Bidders submit their bids simultaneously.[17] The highest bidder wins the object and pays its bid. Other bidders pay nothing. In case of a tie, the winning bidder is chosen randomly from among those who submitted the highest bid.

We compute a Bayesian-Nash equilibrium of the first price auction. Suppose that for each bidder to bid $\beta(x_i)$ where $\beta : [0, \omega] \to \mathbb{R}$ is increasing and differentiable is a Bayesian-Nash equilibrium of the first price auction.

A heuristic computation of β. The expected payoff to bidder 1 with valuation x who bids b when other bidders bid according to β is given by

$$\Pr(1 \text{ wins with } b) \cdot (x - b) = G\left(\beta^{-1}(b)\right) \cdot (x - b)$$

where $G = F^{N-1}$ denotes the distribution function of the random variable Y_1, which is the maximum of $N-1$ independently drawn valuations that are drawn according to F.[18]

Maximizing this expression with respect to b yields the first-order condition:

$$\frac{g\left(\beta^{-1}(b)\right)}{\beta'\left(\beta^{-1}(b)\right)}(x - b) - G\left(\beta^{-1}(b)\right) = 0$$

[17] Bidding need not be literally simultanous. What's important is that bidders don't know other bidders' bids at the time they make their own bids. So, for example, writing bids into envelopes qualifies as simultanous bidding even if it's not all done at the exact same time.

[18] Observe that

$$\begin{aligned}
\Pr(1 \text{ wins with } b) &= \Pr(b > \beta(x_2), ..., \beta(x_n)) \\
&= \Pr(b > \beta(x_2)) \cdots \Pr(b > \beta(x_n)) \\
&= \Pr\left(x_2 < \beta^{-1}(b)\right) \cdots \Pr\left(x_n < \beta^{-1}(b)\right) \\
&= \Pr\left(x < \beta^{-1}(b)\right)^{N-1} \\
&= F\left(\beta^{-1}(b)\right)^{N-1} \\
&= G\left(\beta^{-1}(b)\right).
\end{aligned}$$

where $g = G'$ denotes the density of Y_1 (recall that for any function $f : X \to Y$, $\frac{df^{-1}(y)}{dy} = \frac{1}{f'(x)}$). Because bidding according to β is an equilibrium, $b = \beta(x)$ (or $\beta^{-1}(b) = x$) the previous equation yields the following differential equation:

$$\begin{aligned} \frac{g(x)}{\beta'(x)}(x - \beta(x)) - G(x) &= 0 \\ G(x)\beta'(x) + g(x)\beta(x) &= xg(x) \end{aligned}$$

or

$$\frac{d}{dx}(G(x)\beta(x)) = xg(x).$$

Integrating both sides according to x yields:

$$G(x)\beta(x) = \int_0^x yg(y)\,dy + C$$

(observe that the fact that $G(0) = 0$ implies that $C = 0$) or[19]

$$\begin{aligned} \beta(x) &= \frac{1}{G(x)}\int_0^x yg(y)\,dy \\ &= E[Y_1 | Y_1 \leq x]. \end{aligned}$$

Remark. This derivation is heuristic because the differential equation is only a necessary condition for equilibrium.

Remark. Observe that the fact that F is continuous and increasing implies that $E[Y_1 | Y_1 \leq x] < x$ for $x > 0$. This formula also shows that the bid is increasing with N because the first order statistic Y_1 increases with N.

We now show that $\beta^{\mathrm{I}}(x) = E[Y_1 | Y_1 \leq x]$ is indeed a Bayesian-Nash equilibrium of the first-price auction. Suppose that all $N - 1$ bidders bid according to β^{I}. We show that to bid according to β^{I} is a best response. It is not optimal to bid $b > \beta^{\mathrm{I}}(\omega)$. The expected payoff to a bidder who has valuation x if she bids b is calculated as follows. Denote $\beta^{\mathrm{I}}(z) = b$ or $z = \left(\beta^{\mathrm{I}}\right)^{-1}(b)$.

$$\begin{aligned} \Pi(b, x) &= G(z)\left(x - \beta^{\mathrm{I}}(z)\right) \\ &= G(z)x - G(z)E[Y_1 | Y_1 \leq z] \\ &= G(z)x - \int_0^z yg(y)\,dy \\ &= G(z)x - G(z)z + \int_0^z G(y)\,dy \\ &= G(z)(x - z) + \int_0^z G(y)\,dy \end{aligned}$$

[19] Note that by L'Hopital's rule:

$$\lim_{x \searrow 0} \beta(x) \overset{\mathcal{L}}{=} \frac{0 \cdot g(0)}{g(0)} = 0.$$

where the 4th equality follows from integration by parts.

Sufficiency follows from the fact that

$$\begin{aligned}\Pi\left(\beta^{\mathrm{I}}(x),x\right)-\Pi\left(\beta^{\mathrm{I}}(z),x\right) &= \int_0^x G(y)\,dy-\left(G(z)(x-z)+\int_0^z G(y)\,dy\right)\\ &= G(z)(z-x)-\int_x^z G(y)\,dy\\ &\geq 0\end{aligned}$$

regardless of whether $z > x$ or $z < x$ (demonstrate this on a figure with a plot of G).

Reserve Price. If the seller sets a reserve price $r > 0$, then bidders with valuations below r cannot possibly win. A bidder with valuation r bids $\beta^{\mathrm{I}}(r) = r$ in equilibrium (because by bidding r it wins if every other bidder has a valuation below r). The analysis above can be repeated to show that in this case $\beta^{\mathrm{I}}(x) = E[\max\{Y_1, r\}|Y_1 \leq x]$ for $x \geq r$ and zero otherwise is a Bayesian-Nash equilibrium of the first-price auction. The fact that $E[\max\{Y_1, r\}|Y_1 \leq x] > E[Y_1|Y_1 \leq x]$ for $x \geq r$ suggests that the seller may be able to increase its expected revenue by setting a positive reserve price.

Uniqueness of Equilibrium. See Lebrun (IER, 1999) for a proof that an equilibrium exists for the first price auction in private value environments and for sufficient conditions it is unique. The method of proof used by Lebrun is based on the mathematical theory of existence and uniqueness of solutions to systems of partial differential equations.

4.7.3. Second Price Auction

Description. Bidders submit their bids simultaneously. The highest bidder wins the object and pays the second highest bid. Other bidders pay nothing. In case of a tie, the winning bidder is chosen randomly from among those submitted the highest bid.

Bidding the true valuation is a dominant strategy in the second-price auction. To see this, let b_1 denote the highest bid made by the other bidders and distinguish among the cases in which $x > b_1$, $x < b_1$, and $x = b_1$.

Reserve Price. The setting of a positive reserve price by the seller has no effect on the bidders' incentives to bid truthfully. Bidding the true willingness to pay is still a dominant strategy for the bidders. A positive reserve price may nevertheless increase the expected revenue to the seller in the event that the second highest bid falls below the reserve price.

4.8. Revenue Equivalence

The expected revenue to the seller in the second-price auction is equal to the expected value of the second highest valuation, or the second-order statistic from among $X_1, X_2, ..., X_N$, denoted Y_2 (recall that Y_1 denotes the highest value from among the $N-1$ values $X_2, ..., X_N$).

Because bidders in the first-price auction bid $E\left[Y_1 \left|Y_1 \leq x\right.\right]$, the expected revenue to the seller in a first price auction is given by

$$\begin{aligned}\int E\left[Y_1 \left|Y_1 \leq x\right.\right] dF^N(x) &= E\left[E\left[Y_2 \left|Y_2 < x\right.\right]\right] \\ &= E\left[Y_2\right]\end{aligned}$$

by the law of iterated expectation.[20]

Since the first price auction is equivalent to the Dutch auction (where the price is lowered until one of the bidders stops it and claims the object), and in private values environments, the second price auction is strategically equivalent to the English auction (or the oral, or open outcry auction), then the expected revenue to the seller under each one of these four auctions is identical.

We show that this equivalence holds more generally.

We relax the symmetry assumption. We denote the distribution of bidder i's valuation and its support by F_i and $\mathcal{X}_i$, respectively, and let $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$.

The revelation principle implies that for every equilibrium of every auction there exists an incentive compatible and individually rational direct revelation mechanism $\langle Q, M\rangle$ that generates the same outcome, where $Q : \mathcal{X} \to \Delta_N$ ($\Delta_N = \left\{(\theta_0, \theta_1, ..., \theta_N) : \theta \geq 0 \text{ and } \sum_{i=0}^N \theta_i = 1\right\}$ is the N dimensional simplex) and $M : \mathcal{X} \to \mathbb{R}^N$. The functions Q and M denote the probability that each bidder wins and its expected payment, respectively, as a function of the bidders' types.

Given a direct revelation mechanism $\langle Q, M\rangle$, let

$$q_i(z_i) \equiv \int_{\mathcal{X}_{-i}} Q_i(z_i, x_{-i}) f_{-i}(x_{-i}) dx_{-i}$$

and

$$m_i(z_i) \equiv \int_{\mathcal{X}_{-i}} M_i(z_i, x_{-i}) f_{-i}(x_{-i}) dx_{-i}$$

denote the expected probability of winning and the expected payment of bidder i with report z_i.

[20] Perhaps an easier way of seeing this is the following: the expected payment of a bidder with valuation x in the first price auction is

$$G(x) \times E\left[Y_1 \left|Y_1 < x\right.\right].$$

The expected payment of a bidder with valuation x in a second price auction is

$$\begin{aligned}& \Pr[Win] \times E\left[\text{2nd highest bid} \left|x \text{ is the highest bid}\right.\right] \\ =\ & \Pr[Win] \times E\left[\text{2nd highest value} \left|x \text{ is the highest value}\right.\right] \\ =\ & G(x) \times E\left[Y_1 \left|Y_1 < x\right.\right].\end{aligned}$$

A direct revelation mechanism $\langle Q, M\rangle$ is incentive compatible if

$$q_i(x_i)\,x_i - m_i(x_i) \geq q_i(z_i)\,x_i - m_i(z_i)$$

for every $x_i, z_i \in \mathcal{X}_i$.

A direct revelation mechanism $\langle Q, M\rangle$ is individually rational if

$$q_i(x_i)\,x_i - m_i(x_i) \geq 0$$

for every $x_i \in \mathcal{X}_i$.

Proposition 1. A direct revelation mechanism $\langle Q, M\rangle$ is incentive compatible if and only if q_i is nondecreasing for every i and

$$\begin{aligned} U_i(x_i) &\equiv q_i(x_i)\,x_i - m_i(x_i) \\ &= U_i(0) + \int_0^{x_i} q_i(t_i)\,dt_i. \end{aligned}$$

Proof. If $\langle Q, M\rangle$ is incentive compatible then for every $x_i, z_i \in \mathcal{X}_i$

$$U_i(x_i) \equiv q_i(x_i)\,x_i - m_i(x_i) \geq q_i(z_i)\,x_i - m_i(z_i)$$

and

$$U_i(z_i) \equiv q_i(z_i)\,z_i - m_i(z_i) \geq q_i(x_i)\,z_i - m_i(x_i).$$

It follows that

$$q_i(z_i)(x_i - z_i) \leq U_i(x_i) - U_i(z_i) \leq q_i(x_i)(x_i - z_i)$$

from which it follows that q_i is nondecreasing. Dividing by $x_i - z_i$ and taking the limit as z_i tends to x_i implies that

$$U_i'(x_i) = q_i(x_i)$$

whenever q_i is continuous, which because of monotonicity of q_i is for almost every $x_i \in \mathcal{X}_i$ (that is, it except possibly for a set of measure zero). The function $U_i(x_i)$ is absolutely continuous[21] and as such it is the integral of its derivative, or such that $U_i(x_i) = U_i(0) + \int_0^{x_i} q_i(t_i)\,dt_i$.

[21] Recall that a function f is continuous at a point x if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x') - f(x)| < \varepsilon$$

if

$$|x' - x| < \delta.$$

A function f is absolutely continuous if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\sum_{i=1}^{n} |f(x_i') - f(x_i)| < \varepsilon$$

Conversely, suppose that q_i is nondecreasing and $U_i(x_i) = U_i(0) + \int_0^{x_i} q_i(t_i)\,dt_i$. Incentive compatibility is satisfied if and only if

$$\begin{aligned} U_i(x_i) &\geq q_i(z_i)\,x_i - m_i(z_i) \\ &= q_i(z_i)\,z_i - m_i(z_i) + q_i(z_i)\,x_i - q_i(z_i)\,z_i \end{aligned}$$

if and only if

$$U_i(x_i) \geq U_i(z_i) + q_i(z_i)(x_i - z_i)$$

for every $x_i, z_i \in \mathcal{X}_i$, if and only if

$$\int_{z_i}^{x_i} q_i(t_i)\,dt_i \geq q_i(z_i)(x_i - z_i).$$

The fact that q_i is nondecreasing implies that this last inequality holds for every $x_i, z_i \in \mathcal{X}_i$. ■

We thus have,

Proposition 2 (Revenue Equivalence Theorem; Vickrey, 1961; Myerson, 1981). If the direct revelation mechanism $\langle Q, M\rangle$ is incentive compatible, then for every i and type x_i, x_i's expected payment is

$$\begin{aligned} m_i(x_i) &= q_i(x_i)\,x_i - U_i(x_i) \\ &= q_i(x_i)\,x_i - \int_0^{x_i} q_i(t_i)\,dt_i - U_i(0). \end{aligned}$$

for every finite collection $\{(x_i, x_i')\}$ of nonoverlapping intervals satisfying

$$\sum_{i=1}^{n} |x_i' - x_i| < \delta.$$

It can be shown (see, e.g., Royden's "Real Analysis") that a function is absolutely continuous if and only if it is the indefinite integral of its derivative (i.e., $F(b) = F(a) + \int_a^b f(x)\,dx$ where $f = \frac{dF}{dx}$).

An example of a continuous function that is not absolutely continuous is (show first $\sin x$ and $\sin 1/x$)

$$x \sin \frac{1}{x}.$$

It should be noted that a monotone nondecreasing function is not necessarily absolutely continuous. For example, the Cantor ternary function is continuous monotone nondecreasing on the interval $[0,1]$, is equal to zero at zero and to 1 at 1, and has a derivative that is equal to zero a.e. on $[0,1]$.

The function U_i is absolutely continuous because

$$|U_i(x_i) - U_i(z_i)| \leq |x_i - z_i|$$

which tends to zero as z_i approaches x_i. More generally, it can be shown that any Lipschitz function is absolutely continuous (see Royden).

Finally, Krishna avoids dealing with absolutely continuous functions by showing that U is convex ($U'' = q' \geq 0$), which implies it is absolutely continuous.

Thus, the expected payment in any two incentive compatible mechanisms with the same allocation rule, Q, that provide the lowest type of each bidder with the same expected payoff $U_i(0)$ is equal.

Corollary. The revenue equivalence result implies that in the symmetric model, the first and the second-price auctions, and the Dutch and English auctions generate the same expected revenue to the seller, as would the all-pay auction, the third-price auction, and many other auction forms. Observe however that the first and second price auctions do not induce the same allocation rule in asymmetric environments.

4.9. Optimal Auctions

The optimal auction, or the auction that maximizes the expected revenue to the seller is the solution to the following problem:

$$\max_{\langle Q,M\rangle} \sum_{i=1}^{N} E\left[m_i\left(X_i\right)\right]$$

subject to incentive compatibility and individual rationality.

Proposition 1 implies

$$\begin{aligned} E\left[m_i\left(X_i\right)\right] &= \int_0^{\omega_i} m_i\left(x_i\right) f_i\left(x_i\right) dx_i \\ &= \int_0^{\omega_i} \left(q_i\left(x_i\right) x_i - U_i\left(x_i\right)\right) f_i\left(x_i\right) dx_i \\ &= \int_0^{\omega_i} q_i\left(x_i\right) x_i f_i\left(x_i\right) dx_i - \int_0^{\omega_i}\int_0^{x_i} q_i\left(t_i\right) f_i\left(x_i\right) dt_i dx_i - U_i(0) \end{aligned}$$

By changing the order of integration,

$$\begin{aligned} \int_0^{\omega_i}\int_0^{x_i} q_i\left(t_i\right) f_i\left(x_i\right) dt_i dx_i &= \int_0^{\omega_i}\int_{t_i}^{\omega_i} q_i\left(t_i\right) f_i\left(x_i\right) dx_i dt_i \\ &= \int_0^{\omega_i} \left(1 - F_i\left(t_i\right)\right) q_i\left(t_i\right) dt_i \end{aligned}$$

which implies that

$$\begin{aligned} E\left[m_i\left(X_i\right)\right] &= \int_0^{\omega_i} q_i\left(x_i\right) x_i f_i\left(x_i\right) dx_i - \int_0^{\omega_i} \left(1 - F_i\left(x_i\right)\right) q_i\left(x_i\right) dx_i - U_i(0) \\ &= \int_0^{\omega_i} \left(x_i - \frac{1 - F_i\left(x_i\right)}{f_i\left(x_i\right)}\right) q_i\left(x_i\right) f_i\left(x_i\right) dx_i - U_i(0) \\ &= \int_{\mathcal{X}} \left(x_i - \frac{1 - F_i\left(x_i\right)}{f_i\left(x_i\right)}\right) Q_i(x) f(x)\, dx - U_i(0) \end{aligned}$$

The objective is thus to choose a mechanism $\langle Q, M\rangle$ maximize

$$\sum_{i=1}^{N} \int_{\mathcal{X}} \left(x_i - \frac{1 - F_i\left(x_i\right)}{f_i\left(x_i\right)}\right) Q_i(x) f(x)\, dx - \sum_{i=1}^{N} U_i(0)$$

subject to incentive compatibility and individual rationality. By Proposition 1, incentive compatibility is equivalent to the requirement that each q_i be nondecreasing. Individual rationality requires that $U_i(0) \geq 0$, and since the objective is to maximize the expected revenue to the seller, this implies that each $U_i(0)$ should be optimally set equal to 0.

Define

$$\psi_i(x_i) = x_i - \frac{1 - F_i(x_i)}{f_i(x_i)}$$

to be the virtual valuation of bidder i with value x_i.[22] The seller's problem is to maximize

$$\sum_{i=1}^{N} \int_{\mathcal{X}} \psi_i(x_i) Q_i(x) f(x)\, dx.$$

This expression is maximized if for each x, $Q_i(x)$ is set equal to 1 if $i = \arg \max_{i \in \{1,\ldots,N\}} \{\psi_i(x_i)\}$ provided this maximum is nonnegative and zero otherwise (the tie-breaking rule is unimportant). If the virtual valuations are nondecreasing, the resulting q_i's are nondecreasing too because if $z_i < x_i$ then $\psi_i(z_i) \leq \psi_i(x_i)$, and thus for every x_{-i}, $Q_i(z_i, x_{-i}) \leq Q_i(x_i, x_{-i})$, which implies that

$$q_i(z_i) = \int_{\mathcal{X}_{-i}} Q_i(z_i, x_{-i}) f_{-i}(x_{-i})\, dx_{-i} \leq \int_{\mathcal{X}_{-i}} Q_i(x_i, x_{-i}) f_{-i}(x_{-i})\, dx_{-i} = q_i(x_i).$$

We have thus solved for the optimal auction for the "regular" case, in which the virtual valuations are nondecreasing[23]: Q is defined as above, and M_i is defined such that

$$m_i(x_i) = q_i(x_i) x_i - \int_0^{x_i} q_i(t_i)\, dt_i$$

or

$$M_i(x) = Q_i(x) x_i - \int_0^{x_i} Q_i(t_i, x_{-i})\, dt_i.$$

More intuitively, observe that because

$$Q_i(x) = \begin{cases} 1 & \text{if } \psi_i(x_i) > \max\{\psi_j(x_j), 0\} \text{ for every } j \neq i \\ 0 & \text{otherwise} \end{cases}$$

[22] This virtual valuation may be interpreted as bidder i's marginal revenue. Demand is given by $q(p) = 1 - F(p)$ where q is quantity = probability of purchase. Inverse demand is $p(q) = F^{-1}(1-q)$. The revenue for the seller is $p(q)q = qF^{-1}(1-q)$. Marginal revenue is

$$\begin{aligned} \frac{d}{dq}\left[qF^{-1}(1-q)\right] &= F^{-1}(1-q) - \frac{q}{f(F^{-1}(1-q))} \\ &= p - \frac{1-F(p)}{f(p)} \\ &= \psi(p). \end{aligned}$$

See Bulow and Roberts (JPE, 1989) or Krishna's textbook.

[23] Many distributions, such as the uniform, normal, etc., are indeed "regular" in this sense.

the winner is the bidder with the highest virtual valuation, and that the bidder pays

$$x_i - \int_{y_i(x_{-i})}^{x_i} Q_i(t_i, x_{-i})\, dt_i = y_i(x_{-i})$$

where $y_i(x_{-i})$ is equal to the lowest valuation with which i could still win the auction.[24] Other bidders pay nothing.

This implies that in a symmetric environment, where all the virtual valuation functions are identical, the second-price auction with a reserve price r that is such that $\psi_i(r) = r - \frac{1-F_i(r)}{f_i(r)} = 0$ or $r = \psi_i^{-1}(0)$ is an optimal auction (notice that the optimal reserve price is *independent* of the number of bidders, N). The revenue equivalence result implies that the first-price auction with the same reserve price, as well as many other auction forms are optimal as well.

Remark. This derivation, including the revelation principle, the revenue equivalence result, and the derivation of the optimal auction appeared in Myerson (1981). Myerson (1981) also contains a general solution for the case in which the virtual valuations are not necessarily nondecreasing[25] and some ideas about how to generalize the solution for the case in which the bidders' valuations are correlated, which was later solved by Crémer and McLean.

Remark. How important is the reserve price? Bulow and Klemperer (AER, 1996) show that a standard auction with no reserve price and $n+1$ bidders generates a higher expected revenue to the seller than an optimal auction with n bidders, which they interpret as "optimal negotiations." They interpret their result as establishing the superiority of "more competition" over "optimal negotiations."

4.10. Risk Averse Bidders

Risk aversion, namely the assumption that the payoff function of a bidder in an auction in which it pays only when it wins is given by $\Pr[Win] \times u(x-p) + \Pr[Lose] \times u(0)$ where u is

[24] Does this similarity to the second price auction imply that under the optimal auction bidders have a weakly dominant strategy to bid truthfully?

[25] In this case the q_i need to be "ironed" to ensure their monotonicity. Ironing is similar to what a discriminating monopolist who engages in 3rd degree price discrimination does. Suppose that

$$\begin{aligned} p &= 100 - 2q \qquad 0 \le q \le 20 \\ p &= 70 - .5q \qquad 20 \le q \le 100 \end{aligned}$$

Then marginal revenue is given by

$$\begin{aligned} MR &= 100 - 4q \qquad 0 \le q \le 20 \\ MR &= 70 - q \qquad 0 \le q \le 20 \end{aligned}$$

Suppose that the marginal cost is equal to 40. In this case the monopolist operates as if it has an "ironed out" MR curve that has MR = 40 for $15 \le q \le 30$. The way to implement this ironing is by selling it to buyers with probability 1/3.See Bulow and Roberts (pp. 1078-80).

concave, leads to higher bidding in the first price auction. To see this consider bidder 1 with valuation x in a first price auction Fix the strategies of all the other bidders and suppose bidder 1 bids b. Now suppose that this bidder considers decreasing his bid slightly to $b - \Delta$. If he wins the auction with this lower bid, this leads to a gain of Δ. A lowering of his bid could, however, cause the bidder to lose the auction. For a risk averse bidder, the effect of a slightly lower winning bid on his wealth level has a smaller utility consequence than does the possible loss if this lower bid, were, in fact, to result in his losing the auction. Compared to a risk neutral bidder, a risk averse bidder will thus bid higher. Put another way, by bidding higher, a risk averse bidder "buys insurance" against the possibility of losing.

Risk aversion does no affect bidders' behavior in a second price auction, where bidding the true valuation is still a weakly dominant strategy. It follows that risk aversion does not affect the expected revenue to the seller in a second price auction, but increases the expected revenue to the seller in a first price auction.

4.11. Renegotiation

One of the practical concerns of mechanism design theory is that players might have incentives to change the rules of the game they are playing. Although in some cases the mechanism designer can prevent such changes, in many situations it is impossible or nearly impossible to do so, especially when a change in the rules of the game, contract, or mechanism is mutually beneficial for the players. Such mutually consensual changes, which are known as *renegotiation*, can occur at different stages of the contractual process. *Interim renegotiation* takes place before the mechanism is played and involves a change of the mechanism and the equilibrium the players intend to play. *Ex post renegotiation* takes place after the mechanism is played and involves a change of the outcome or recommendation proposed by the mechanism. The consequences of both interim and ex post renegotiation crucially depend on the details of the renegotiation process: what alternative outcomes or mechanisms are considered? How do the players communicate with each other, and how do they select among the alternative proposals? How is the surplus that is generated by renegotiation shared among the players?

4.11.1. Interim Renegotiation

The following two examples are taken from Holmström and Myerson (1983).

Suppose that there are two individuals in the economy, and each individual may be one of two possible types. Individual 1 may be type $1a$ or $1b$, individual 2 may be type $2a$ or $2b$, and all four possible combinations of types are equally likely. There are three possible decisions called A, B, and C. The payoff of each individual from each decision depends only on his own type (private values), as shown in the following table.

	$U1a$	$U1b$	$U2a$	$U2b$
$d = A$	2	0	2	2
$d = B$	1	4	1	1
$d = C$	0	9	0	-8

In this example, individual 2 in either type and individual 1 in type $1a$ both prefer A over B and B over C. However if individual 1 is type $1b$ then his preference ordering is reversed and he strongly prefers C. Type $2b$ differs from $2a$ in that $2b$ has a greater aversion to decision C. (These are von Neumann-Morgenstern utility numbers.) Among all incentive-compatible decision rules, the following decision rule δ uniquely maximizes the sum of the two individuals' ex ante expected utilities:

$$\begin{aligned} \delta(1a, 2a) &= A, \qquad \delta(1a, 2b) = B \\ \delta(1b, 2a) &= C, \qquad \delta(1b, 2b) = B \end{aligned}$$

Notice that this decision rule selects decision C, type lb's most preferred decision, if the types are $1b$ and $2a$; but if 2's type is $2b$ (so that 2 is more strongly averse to C) then the decision rule selects B instead. To check that δ is incentive compatible, notice that type $2a$ can get decisions A or C with equal probability if he is honest, or he can get B for sure if he lies and reports his type as $2b$. Since both of these prospects give the same expected utility to $2a$, he is willing to report his type honestly when δ is implemented.

The decision rule δ is incentive efficient (in both the interim and ex ante senses), so no outsider could suggest any other incentive-compatible decision rule that makes some types better off without making any other types worse off than in δ.

But if individual 1 knows that his type actually is $1a$, then he knows that he and individual 2 both prefer decision A over the decision rule δ. Thus, rather than let δ be implemented, individual 1 in type $1a$ would suggest that decision A be implemented instead, and individual 2 would accept this suggestion.

Thus, although δ is an incentive-efficient decision rule, it is possible for the individuals to unanimously approve a change to some other decision rule (namely A-for-sure). Of course, this unanimity in favor of A over δ depends on l's type being $1a$, but consider what would happen if 1 were to insist on using δ rather than A. Individual 2 would infer that l's type must be $1b$. Then decision rule δ would no longer be incentive compatible, because both types of individual 2 would report "$2b$", to get decision B rather than C.[26]

[26] Note that the rule that would end up being implemented is the mechanism

	$2a$	$2b$
$1a$	A	A
$1b$	B	B

which is durable.

Thus, if the individuals can redesign their decision rule when they already know their own types, then the decision rule δ could not be implemented in this example, even though it is incentive compatible and incentive efficient. In the terminology of Holmström and Myerson, δ is incentive efficient but not *durable*.

Holmström and Myerson proceed to provide a definition of durable mechanisms and to establish their existence.

Informal Definition. An incentive compatible mechanism is *durable* if any alternative mechanism that is proposed to the players is blocked with probability one in a *non trivial equilibrium* of the voting game in which the players vote simultaneously for either the alternative or original mechanism, and the alternative mechanism is implemented if and only if everyone votes in its favor.

The definition is informal because the definition of "non trivial equilibrium" is unspecified. It refers to an equilibrium that is the limit of a sequence of strictly mixed profiles of strategies. This rules out the trivial equilibrium where everyone votes against the alternative mechanism.

However, their definition only requires that for every alternative mechanisms that is suggested to the players, there is a nontrivial equilibrium where this alternative is rejected. A stronger definition would have required that every alternative mechanism is rejected in every plausible mechanism. To see that this can make a difference, consider the next example.

Suppose that there are two individuals with two independent and equally likely types ($1a$, $1b$; $2a$, $2b$), and there are two possible decisions, A and B. The two individuals get the same payoffs, as follows:

$$\begin{aligned}
u_1(A,t) &= u_2(A,t) = 2, \qquad \forall t \\
u_1(B,t) &= u_2(B,t) = \begin{cases} 3 & \text{if } t = (1a,2a) \text{ or } t = (1b,2b) \\ 0 & \text{if } t = (1a,2b) \text{ or } t = (1b,2a) \end{cases}
\end{aligned}$$

In this example, let $\delta(t) = A$ for all t. Then δ is not interim incentive efficient (it is dominated by the mechanism

δ^*	$2a$	$2b$
$1a$	B	A
$1b$	A	B

) but it is durable. The two individuals would both gain from changing to B when their types match; but in any voting game with any alternative mechanism, there is always an equilibrium rejection in which both individuals always use uninformative voting and reporting strategies. The notion of durability merely assumes that the individuals would play noncooperatively in the voting game. Individuals cannot be forced to communicate effectively in a noncooperative game with incomplete information.

The question of what environments admit such interim renegotiation-proof mechanisms and what environments do not admit interim renegotiation-proof mechanisms is open. It is not even known if there exists an example where an interim renegotiation-proof mechanism fails to exist.

4.11.2. Ex-Post Renegotiation

Neeman and Pavlov (2010) propose the following definition for ex-post renegotiation proofness under complete information. See Neeman and Pavlov (2010) for how this definition can be extended to cover incomplete information as well.

Definition 1. *An equilibrium σ of a mechanism $\langle S, m\rangle$ is* ex post renegotiation-proof *if both of the following conditions hold:*

(i) An outcome that is obtained under the equilibrium play of the mechanism cannot be renegotiated in a way that benefits both players. And,

(ii) No agent can improve upon his equilibrium payoff in any state by a unilateral deviation from σ followed by renegotiation of the resulting outcome to another outcome that benefits both players.

The first part of the definition is straightforward. If there is another outcome that Pareto dominates the outcome that was produced by the mechanism, then the latter will be renegotiated. The following example illustrates the second part of the definition.

Example. A buyer and a seller can trade a single good. The buyer values the good at V that can be either 0 or 2, the seller values the good at 1. The realization of V and the seller's valuation are commonly known between the agents. Consider a mechanism where the buyer is asked to report his value: after a report "$V = 2$" the good is transferred from the seller to the buyer at a price p_2, and after a report "$V = 0$" there is no trade and the buyer pays p_0 to the seller. It is easy to see that the buyer has a dominant strategy to report his true valuation if $p_2 - p_0 \in (0, 2)$, and the resulting outcome is ex post efficient. However, as we show below, this equilibrium is not ex post renegotiation-proof unless $p_2 - p_0 = 1$.

Suppose $p_2 - p_0 \in (1, 2)$. If the buyer with $V = 2$ reports "$V = 0$" then the payoffs of the buyer and the seller (without renegotiation) would be $-p_0$ and p_0, respectively. This outcome is Pareto dominated by a decision to trade at a new price $\widehat{p}$ that satisfies $\widehat{p} - p_0 \in (1, 2)$. Hence, for any such $\widehat{p} < p_2$, the buyer would prefer to misreport and then renegotiate the outcome to trade at the price $\widehat{p}$ rather than report his true valuation. Thus, the original equilibrium is not ex post renegotiation-proof.[27]

Neeman and Pavlov proceed to show that under complete information, any budget balanced and ex-post efficient rule can be implemented if the number of agents is larger than or equal to three, but only Groves mechanisms are ex-post renegotiation proof with two agents. The fact that budget balanced Groves mechanisms often fail to exist implies that in many problems, there is no ex post renegotiation proof mechanism.

[27] The argument for the case $p_2 - p_0 \in (0, 1)$ is similar. The buyer with $V = 0$ will find it profitable to report "$V = 2$" and then renegotiate to "no trade" as long as long a new payment $\widehat{p}$ is smaller than p_0.

4.12. Robust Mechanism Design

4.13. Collusion

Exercises

1. Construct a type space that describes the following information structure. Two firms compete in a market. The cost of firm B is zero. The cost of firm A is equally likely to be either zero or one. Firm B sends a spy to check whether firm A has the machine that enables costless production. If firm A has the machine, then the spy discovers it with probability $\frac{1}{2}$. If firm A does not have the machine, then the spy obviously cannot discover it.

2. Describe an example of a mechanism that is ex-post incentive efficient but not interim incentive efficient. Describe an example of a mechanism that is interim incentive efficient but not ex-ante incentive efficient. Describe an example of a mechanism that is ex-ante incentive efficient.

3. Show that in a public good problem with 2 agents who have two types each no Groves mechanism is budget balanced.

4. Find a budget balanced AAGV mechanism for a public good problem with 2 agents whose valuations are uniformly distributed on the unit interval. Show that the mechanism you found is not dominant strategy incentive compatible.

5. A government agency writes a procurement contract with a firm to deliver q units of a good. The firm has constant marginal cost c, so that its profit is $P - cq$, where P denotes the payment for the transaction. The firm's cost is either high (c_H) or low $(c_L$, with $0 < c_H < c_H)$. The agency makes a take-it-or-leave-it offer to the firm (whose default profit is zero). The benefit to the agency of obtaining q units is given by a concave function $B(q)$.

    1. What is the optimal contract for the agency if it knows the firm's cost?
    2. What is the optimal contract for the agency if the firm's cost is private information, and the agency's prior belief about the firm's cost is $\Pr(c = c_L) = \beta$? Formulate the agency's problem, but do not solve it.
    3. Solve the agency's problem for the case where $B(q) = 4x - x^2$, $c_H = 2$, $c_L = 1$, and $\beta = \frac{1}{4}$.

6. Example 23.F.3, p. 906 from MWG (who took it from Myerson, 1991) and the exercises therein.

7. Redo the 2×2 version of the Myerson and Satterthwaite model under the assumption that c represents the value of the object to the seller and that the object is jointly owned by the buyer and seller (Hint: in this case, if there is disagreement, then the buyer and seller each win the object with probability $\frac{1}{2}$; observe that this formulation affects the buyer's and seller's IR constraints but not their IC constraints). Show that in this case there always exist an incentive compatible and individually rational mechanism. See Cramton, Gibbons, and Klemperer (*Econometrica*, 1987) for a general treatment of this case.

8. An object is worth v to a buyer and costs either $\underline{c}$ or $\overline{c}$ to produce, where $v > \overline{c} > \underline{c} > 0$. The cost of production is the private information of the seller. The buyer believes that the cost is high/low with probability p, $1-p$, respectively. What is the optimal buying mechanism for the buyer? Is this mechanism ex-post efficient? Suppose now that the buyer obtains a signal s about the cost of production that is correct with probability $q > .5$ (that is, $\Pr(s = \overline{c} \,|c = \overline{c}) = \Pr(s = \underline{c}\,|c = \underline{c}) = q$). Identify the mechanism that maximizes the expected payoff for the buyer. Hint: in this mechanism the object is traded with probability 1 at an expected price that is equal to its cost of production.

9. Consider a private values auction environment with 2 bidders. Suppose that the common prior is given by the following matrix:

	$v = 1$	$v = 2$
$v = 1$	$\frac{1}{3}$	$\frac{1}{6}$
$v = 2$	$\frac{1}{6}$	$\frac{1}{3}$

Show that the seller can design a dominant strategy auction that extracts the full surplus of the bidders. Hint: consider a Vickrey or a sealed bid second price auction. Show that there exists a participation fee (that depends on the other bidder's bid in the auction) that, for each type of each of the bidders, is equal to the expected surplus of this type from participating in the auction. See Crémer and McLean (ECM, 1985, 1988) for the original construction of such full surplus extraction auctions. See Neeman (JET, 2004) and Heifetz and Neeman (ECM, 2006) about the generality of this method of extracting the full surplus of the bidders.

10. Consider a first-price auction with independent and private values. Show that in equilibrium the bidders' bid function are nondecreasing.

11. Consider a first-price auction with independent private values. Suppose that bidder i's valuation is distributed according to a distribution F_i with support $[a, b]$ where $a > 0$. Show that if $b_i(v_i)$ is bidder i's equilibrium bid function, then $\lim_{v_i \searrow 0} b_i(v_i) = a$.[Be explicit about what you need to assume in order to prove your answer.]]

12. A seller of an object faces a single buyer. The seller believes that the buyer's willingness to pay for the object is uniformly distributed over the interval $[0, 1]$. The value of the

object for the seller is 0. What auction maximizes the expected revenue to the seller? Prove your result.

13. Suppose there are two bidders and that each bidder observes an independent signal $x_1, x_2 \sim U[0,1]$ about the value of the object. The value to both bidders is given by $v_1 = v_2 = x_1 + x_2$.

    1. Find an asymmetric (linear) Bayesian-Nash equilibrium of the second-price auction. Characterize the set of asymmetric linear Bayesian-Nash equilibria.
    2. Find a symmetric Bayesian-Nash equilibrium for the first-price auction. Can you characterize the set of asymmetric linear Bayesian-Nash equilibria in this case?

14. Give an example of n identically distributed random variables that satisfy the MLRP but are not conditionally i.i.d. Give an example of n identically distributed random variables that are conditionally i.i.d. but fail the MLRP.