Lecture Plan

1. Gibbard-Satterthwaite Impossibility Theorem
2. Dominant Strategy Implementation
3. Nash Implementation
4. Subgame perfect and Virtual Implementation
5. Mechanism Design
6. Groves Mechanisms
7. AAGV Mechanisms
8. Buyer-Seller Problem
9. Optimal Auctions
10. Bayesian Persuasion
11. Robustness
12. “Applied Mechanism Design”

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.

small space!

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). and unanimity w/a

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. up to strict part

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 individ~~uals' 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. bold? bold?

can possibly be re-established under different assumptions?

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. show! or give as exercise!

[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.

Alternative methods of aggregation?

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

present model from p. 5(3)

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 $f(u) = a$. why would we want that? perhaps we want to consider also SCF that don't permit choice of some alt.?

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 $f(L_1, ..., L_N) = a$. implies onto!

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~~ outside 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 $f(L_1', ..., L_N') = a$.

Ex. Is Borda Rule monotonic? No, see counter-example below!!

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~~ state 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). ∎

Example: Borda Rule is <u>not</u> monotonic

3:	a	a	b	b	c
2:	c	c	a	a	b
1:	b	b	c	c	a

$\longmapsto$

a : 11
b : 10
c : 9

3:	a	a	b	b	b
2:	b	b	a	a	c
1:	c	c	c	c	a

$\longmapsto$

b : 13
a : 11
c : 6

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). to reporting pref.

strategy-proof vs. dominant strat. (specific vs. general)

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'. But 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.

   Note: $D \subseteq \mathcal{L}$ and $u \in D^N$ because we don't want a model where i's preferences may depend on j's preferences. This could happen if $D \subseteq \mathcal{L}^N$ and $u \in D$.

   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 $(v_T, u_{N\setminus T})$.

   ↳ because only those who prefer b to a manipulate and some of them may manipulate ↳ but it's for a over b!

   Example. Single-peaked preferences. See below.

(*) generally, there are two types of rules: CW Condorcet consistent and scoring. see back page

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.

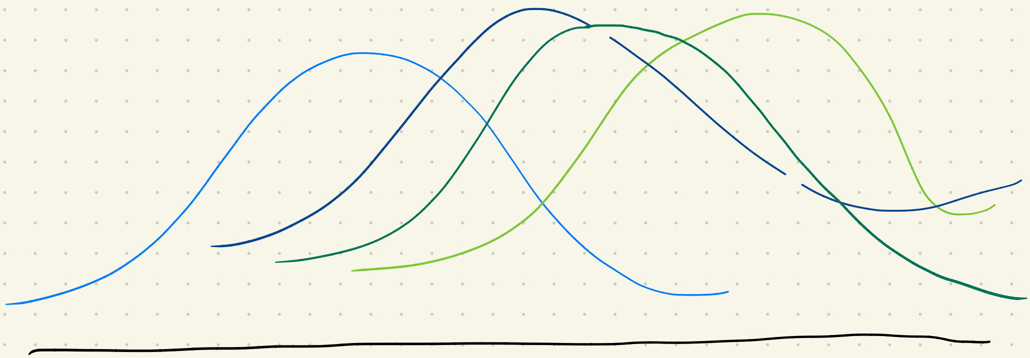
see below ⇓

Exercises

! Pareto - Liberal (Sen)

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.

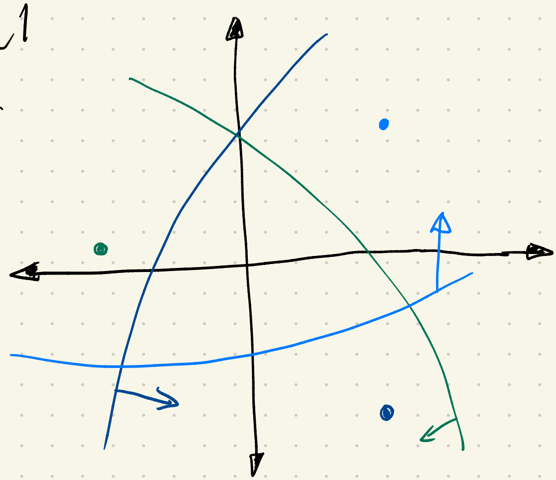
Single peaked preferences on a uni-dimensional space (real line)

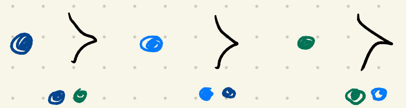


(*) The "median voter's ideal point" is a condorcet winner

(*) This is not true if the space is 2-dimensional

ideal points of 3 different voters and indiff curves are given by different colors





Condorcet Consistent vs. Scoring Rules

Example

5: a a c

4: b b b

1: c c a

$a \succ b$, $a \succ c$ $\Rightarrow$ Condorcet consistent picks a

But scoring rule w. weights 5, 4, 1 picks

b : 12

a : 11

c : 7

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 that satisfies assumption of ... cannot be indifferent between any two alternatives (Hint: use the Strict Neutrality Lemma).

(*) 5. (2) Define single-peaked preferences. (1) Show that majority rule satisfies unanimity and IIA if individuals' preferences are all unidimensional single-peaked. ~~and~~ and

(3) show that Majority rule is transitive

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 strategy-proof:

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

11. 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?

האם ב בורדא מונוטוני ?

לא!

a a ·· a
b b b

שים לב ~~ש~~ שעלה —

b למעלה הגם מיקום.

b ← !b וכן ירד במקום

a

a a d c | Borda
2 3 b b b b | a: 5
1 2 c c d a d | b: 4
0 1

a d c b b a: 5
b b a c c b: 6
c c b a a

[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 $\langle N, (A_i)_{i \in N}, g \rangle$ 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 $\langle N, (A_i)_{i \in N}, (\succsim')_{i \in N} \rangle$ 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.

This definition implicitly assumes game is finite. Can accomodate infinite games by defining $g : H \to C$.

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 form 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

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

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 $f : \mathcal{P} \rightarrow C$ 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 $\langle N, (A_i)_{i\in N}, (\succsim)_{i\in N}\rangle$ 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~~ (and for social choice rules instead of social choice functions). 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

θ (1 is real mother)	$a \succ_1 b \succ_1 d$	$b \succ_2 d \succ_2 a$
θ' (2 is real mother)	$a \succ'_1 d \succ'_1 b$	$b \succ'_2 a \succ'_2 d$

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

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

Truthful implementation induces the following game form:
show that cannot have a, b, or d instead of ?
⟹ truthful implementation is impossible

	θ' mine	θ hers
θ mine	?	a
θ' hers	b	d(?)

What about implementation of the social choice function? next result shows impossible too.

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. G^*. A report of $\succsim'$ by player i in the truthful mechanism G^* produces the same outcome that would be obtained by the original mechanism when all other players play their equilibrium strategies under $\succsim$ and player i plays the equilibrium strategy it plays when players' preferences are give by $(\succsim'_i, \succsim_{-i})$. Since we have a Nash equilibrium under the original mechanism at $\succsim$, it follows that player i cannot benefit from not reporting its preferences truthfully ■

don't check! don't think it's true...

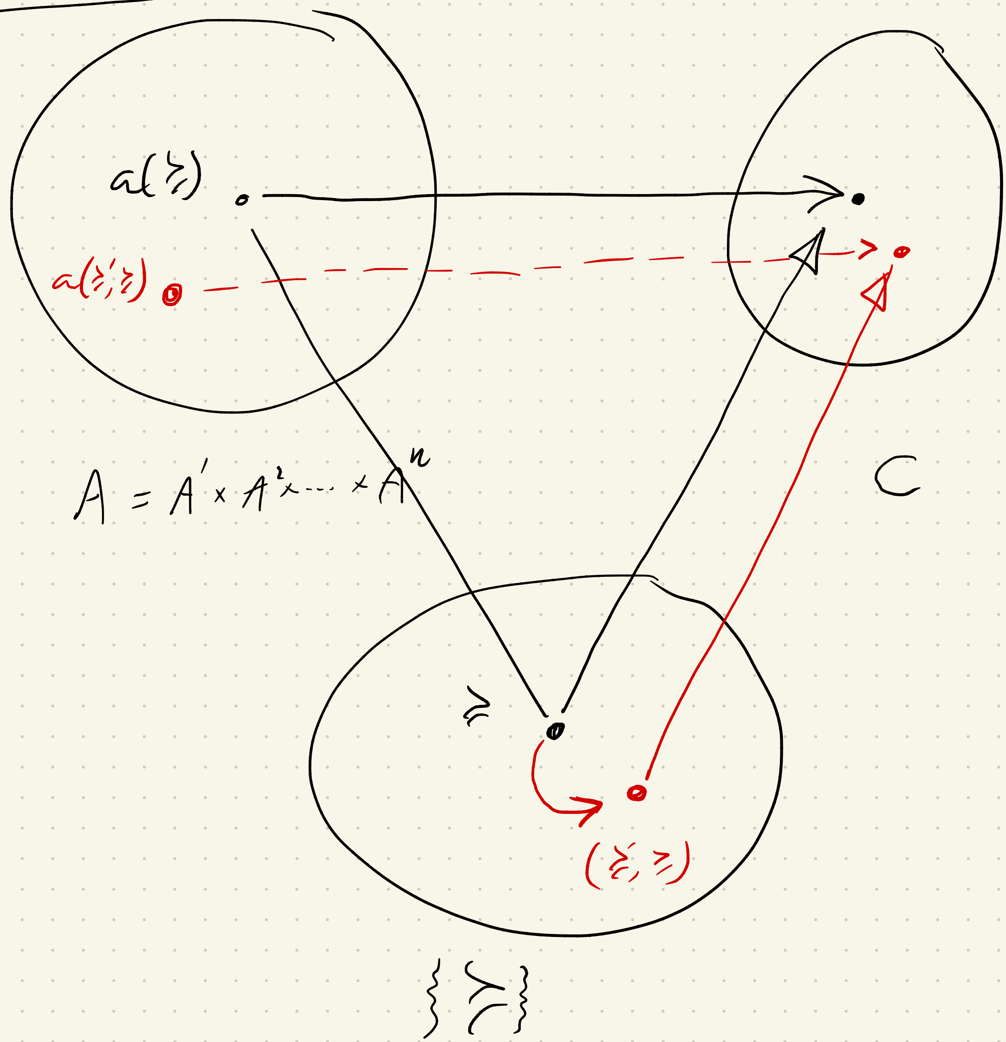
⟷ i.e., we can't restrict attention to games G, C where there are no violations

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 $b \succ'_i c$.

in my proof, players only report own preferences.

Illustration of proof



each player j announces a number a_j, interpreted as a declaration of his value of the project, and the project is executed if and only if the sum of these declarations is at least γ; the payment made by player j is equal to $h_j(a_{-j})$ (which is independent of his announcement), plus, if the project is carried out, an amount equal to the difference between the cost of the project and the sum of the announcements made by the other players. Formally, in this strategic game form $\langle N,(A_i),g\rangle$ we have $A_i = \mathbb{R}$ and $g(a) = (x(a), m(a))$ for each $a \in A$ where

$$\begin{cases} x(a) = 1 \text{ if and only if } \sum_{i\in N} a_i \geq \gamma \\ m_j(a) = x(a)(\gamma - \sum_{i\in N\setminus\{j\}} a_i) + h_j(a_{-j}) \text{ for each } j \in N. \end{cases} \tag{184.1}$$

Such a game form is called a *Groves mechanism*.

▪ PROPOSITION 184.2 *Let $\langle N, C, \mathcal{P}, \mathcal{G}\rangle$ be an environment in which $C = \{(x,m): x \in \{0,1\}$ and $m \in \mathbb{R}^N\}$, $\mathcal{P}$ is the set of profiles $(\succsim_i)$ in which each $\succsim_i$ is represented by a utility function of the form $\theta_i x - m_i$ for some $\theta_i \in \mathbb{R}$, and $\mathcal{G}$ is the set of strategic game forms; identify $\mathcal{P}$ with R^N. A choice function $f: \mathbb{R}^N \to C$ with $f(\theta) = (x(\theta), m(\theta))$ for which*

- *$x(\theta) = 1$ if and only if $\sum_{i\in N} \theta_i \geq \gamma$*
- *for each $j \in N$ there is a function h_j such that $m_j(\theta) = x(\theta)(\gamma - \sum_{i\in N\setminus\{j\}} \theta_i) + h_j(\theta_{-j})$ for all $\theta \in \mathbb{R}^N$*

is truthfully DSE-implemented by the Groves mechanism $\langle N,(A_i),g\rangle$ defined in (184.1).

Proof. Let $j \in N$ and let a_{-j} be an arbitrary vector of actions of the players other than j. We argue that when the players other than j choose a_{-j}, j's payoff when he chooses $a_j = \theta_j$ is at least as high as his payoff when he chooses any other action in A_j. There are three cases.

- If $x(a_{-j},\theta_j) = x(a_{-j},a'_j)$ then $m_j(a_{-j},a'_j) = m_j(a_{-j},\theta_j)$ and hence $g(a_{-j},a'_j) = g(a_{-j},\theta_j)$.
- If $x(a_{-j},\theta_j) = 0$ and $x(a_{-j},a'_j) = 1$ then j's payoff under (a_{-j},θ_j) is $-m_j(a_{-j},\theta_j) = -h_j(a_{-j})$, while his payoff under (a_{-j},a'_j) is $\theta_j - m_j(a_{-j},a'_j) = \theta_j - (\gamma - \sum_{i\in N\setminus\{j\}} a_i) - h_j(a_{-j}) < -h_j(a_{-j})$, since $x(a_{-j},\theta_j) = 0$ implies that $\sum_{i\in N\setminus\{j\}} a_i + \theta_j < \gamma$.
- If $x(a_{-j},\theta_j) = 1$ and $x(a_{-j},a'_j) = 0$ then j's payoff under (a_{-j},θ_j) is $\theta_j - m_j(a_{-j},\theta_j) = \theta_j - (\gamma - \sum_{i\in N\setminus\{j\}} a_i) - h_j(a_{-j})$, while his payoff under (a_{-j},a'_j) is $-m_j(a_{-j},a'_j) = -h_j(a_{-j}) \leq \theta_j - (\gamma - \sum_{i\in N\setminus\{j\}} a_i) - h_j(a_{-j})$, since $x(a_{-j},\theta_j) = 1$ implies that $\sum_{i\in N\setminus\{j\}} a_i + \theta_j \geq \gamma$.

Hence it is a dominant action for each player j to choose $a_j = \theta_j$. The outcome $g(\theta)$ is equal to $f(\theta)$, so that $\langle N,(A_i),g\rangle$ truthfully DSE-implements f. □

Note that the Groves mechanism (184.1) does *not* Nash-implement a choice function f satisfying the conditions of the proposition: for example, if $\gamma = 2$, $|N| = 2$ and $\theta_i = 1$ for both players then the associated game has also, in addition to $(1,1)$, an inefficient equilibrium $(-2,-2)$.

EXERCISE 185.1 In an environment like that in the previous proposition, show that if a choice function f with $f(\theta) = (x(\theta), m(\theta))$ and $x(\theta) = 1$ if and only if $\sum_{i\in N} \theta_i \geq \gamma$ is truthfully DSE-implementable then for each $j \in N$ there is a function h_j such that $m_j(\theta) = x(\theta)(\gamma - \sum_{i\in N\setminus\{j\}} \theta_i) - h_j(\theta_{-j})$ for all $\theta \in \mathbb{R}^N$. [You need to show that whenever $x(\theta_{-j},\theta_j) = 1$ and $x(\theta_{-j},\theta'_j) = 0$ then $m_j(\theta_{-j},\theta_j) - m_j(\theta_{-j},\theta'_j) = \gamma - \sum_{i\in N\setminus\{j\}} \theta_i$.]

10.4 Nash Implementation

We now turn to the case in which the planner, as in the previous section, uses strategic game forms, but assumes that for any game form she designs and for any preference profile the outcome of the game may be any of its Nash equilibria.

The first result is a version of the revelation principle (see also Lemma 181.4). It shows that any Nash-implementable choice rule is also truthfully Nash-implementable: there is a game form in which (*i*) each player has to announce a preference profile and (*ii*) for any preference profile truth-telling is a Nash equilibrium. This result serves two purposes. First, it helps to determine the boundaries of the set of Nash-implementable choice rules. Second, it shows that a simple game can be used to achieve the objective of a planner who considers truthful Nash equilibrium to be natural and is not concerned about the outcome so long as it is in the set given by the choice rule.

▪ LEMMA 185.2 (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. Let $G = \langle N,(A_i),g\rangle$ be a game form that Nash-implements the choice rule $f: \mathcal{P} \to C$ and for each $\succsim \in \mathcal{P}$ let $(a_i(\succsim))$ be a Nash equilibrium of the game $\langle G, \succsim\rangle$. Define a new game form $G^* = \langle N,(A^*_i),g^*\rangle$ in which $A^*_i = \mathcal{P}$ for each $i \in N$ and $g^*(p) = g((a_i(p_i)))$ for each

$p \in \times_{i \in N} A_i^*$. (Note that each p_i is a preference profile and p is a profile of preference profiles.) Clearly the profile p^* in which $p_i^* = \succsim$ for each $i \in N$ is a Nash equilibrium of $\langle G^*, \succsim \rangle$ and $g^*(p^*) \in f(\succsim)$. □

Note that it does *not* follow from this result that in an analysis of Nash implementation we can restrict attention to games in which each player announces a preference profile, since the game that truthfully Nash-implements the choice rule may have non-truthful Nash equilibria that generate outcomes different from that dictated by the choice rule. Note also that it is essential that the set of actions of each player be the set of preference *profiles*, not the (smaller) set of preference relations, as in part (b) of the revelation principle for DSE-implementation (Lemma 181.4).

We now define a key condition in the analysis of Nash implementation.

▶ DEFINITION 186.1 A choice rule $f: \mathcal{P} \to C$ is **monotonic** if whenever $c \in f(\succsim)$ and $c \notin f(\succsim')$ there is some player $i \in N$ and some outcome $b \in C$ such that $c \succsim_i b$ and $b \succ_i' c$.

That is, in order for an outcome c to be selected by a monotonic choice rule when the preference profile is $\succsim$ but not when it is $\succsim'$ the ranking of c relative to some other alternative must be worse under $\succsim'$ than under $\succsim$ for at least one individual.

An example of a monotonic choice rule f is that in which $f(\succsim)$ is the set of weakly Pareto efficient outcomes: $f(\succsim) = \{c \in C$: there is no $b \in C$ such that $b \succ_i c$ for all $i \in N\}$. Another example is the rule f in which $f(\succsim)$ consists of every outcome that is a favorite of at least one player: $f(\succsim) = \{c \in C$: there exists $i \in N$ such that $c \succsim_i b$ for all $b \in C\}$.

■ PROPOSITION 186.2 *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. Suppose that the choice rule $f: \mathcal{P} \to C$ is Nash-implemented by a game form $G = \langle N, (A_i), g \rangle$, $c \in f(\succsim)$, and $c \notin f(\succsim')$. Then there is an action profile a for which $g(a) = c$ that is a Nash equilibrium of the game $\langle G, \succsim \rangle$ but not of $\langle G, \succsim' \rangle$. That is, there is a player j and action $a_j' \in A_j$ such that $g(a_{-j}, a_j') \succ_j' g(a)$ and $g(a) \succsim_j g(a_{-j}, a_j')$. Hence f is monotonic. □

◇ EXAMPLE 186.3 (*Solomon's predicament*) The biblical story of the Judgment of Solomon illustrates some of the main ideas of implementation theory. Each of two women, 1 and 2, claims a baby; each of them knows

who is the true mother, but neither can prove her motherhood. Solomon tries to educe the truth by threatening to cut the baby in two, relying on the fact that the false mother prefers this outcome to that in which the true mother obtains the baby while the true mother prefers to give the baby away than to see it cut in two. Solomon can give the baby to either of the mothers or order its execution.

Formally, let a be the outcome in which the baby is given to mother 1, b that in which the baby is given to mother 2, and d that in which the baby is cut in two. Two preference profiles are possible:

$$\theta \text{ (1 is the real mother): } a \succ_1 b \succ_1 d \text{ and } b \succ_2 d \succ_2 a$$

$$\theta' \text{ (2 is the real mother): } a \succ_1' d \succ_1' b \text{ and } b \succ_2' a \succ_2' d.$$

Despite Solomon's alleged wisdom, the choice rule f defined by $f(\theta) = \{a\}$ and $f(\theta') = \{b\}$ is not Nash-implementable, since it is not monotonic: $a \in f(\theta)$ and $a \notin f(\theta')$ but there is no outcome y and player $i \in N$ such that $a \succsim_i y$ and $y \succ_i' a$. (In the biblical story Solomon succeeds in assigning the baby to the true mother: he gives it to the only woman to announce that she prefers that it be given to the other woman than be cut in two. Probably the women did not perceive Solomon's instructions as a strategic game form.)

The next result provides sufficient conditions for a choice rule to be Nash-implementable.

▶ DEFINITION 187.1 A choice rule $f: \mathcal{P} \to C$ has **no veto power** if $c \in f(\succsim)$ whenever for at least $|N| - 1$ players we have $c \succsim_i y$ for all $y \in C$.

■ PROPOSITION 187.2 *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. Let $f: \mathcal{P} \to C$ be a monotonic choice rule that has no veto power. We construct a game form $G = \langle N, (A_i), g \rangle$ that Nash-implements f as follows. The set of actions A_i of each player i is the set of all triples (p_i, c_i, m_i), where $p_i \in \mathcal{P}$, $c_i \in C$, and m_i is a nonnegative integer. The values $g((p_i, c_i, m_i)_{i \in N})$ of the outcome function are defined as follows.

- If for some $j \in N$ and some $(\succsim, c, m)$ with $c \in f(\succsim)$ we have $(p_i, c_i, m_i) = (\succsim, c, m)$ for all $i \in N \setminus \{j\}$ then

$$g((p_i, c_i, m_i)) = \begin{cases} c_j & \text{if } c \succsim_j c_j \\ c & \text{if } c \prec_j c_j. \end{cases}$$

$\succ_j$ $\precsim_j$?

~~seems to work better this way.~~

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).

Examples of Monotone Social Choice Rules.

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

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

from definition of monotonicity!

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.

monotonicity is necessary for Nash-implementation

Proof. Osborne and Rubinstein, p. 186. above!

Example (Solomon's trial as an implementation problem). Recall the setup of Solomon's trial as a truthful implementation problem.

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

and preferences are given by

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

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."

(Revelation Principle implies that impossibility of truthful implementation implies impossibility of implementation. The argument given here is a (simpler) direct proof

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.

(*) שאלה: אמינות!

האם יש דוגמא? לכל מנגנון. אם כבר וכו' אלא נימוק מינימלי?

Proof. Osborne and Rubinstein, pp. 187-8. [בקיצור, לכל מקרה אלא?!] ■

בקיצור, נימוק משמעותי [illegible]. וכו'.

below #

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. this implies that the sufficiency part of Maskin's result applies only to choice rules or 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. See below: #

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!

$p \in \times_{i \in N} A_i^*$. (Note that each p_i is a preference profile and p is a profile of preference profiles.) Clearly the profile p^* in which $p_i^* = \succsim$ for each $i \in N$ is a Nash equilibrium of $\langle G^*, \succsim\rangle$ and $g^*(p^*) \in f(\succsim)$. □

Note that it does *not* follow from this result that in an analysis of Nash implementation we can restrict attention to games in which each player announces a preference profile, since the game that truthfully Nash-implements the choice rule may have non-truthful Nash equilibria that generate outcomes different from that dictated by the choice rule. Note also that it is essential that the set of actions of each player be the set of preference *profiles*, not the (smaller) set of preference relations, as in part (b) of the revelation principle for DSE-implementation (Lemma 181.4).

We now define a key condition in the analysis of Nash implementation.

▶ DEFINITION 186.1 A choice rule $f: \mathcal{P} \to C$ is **monotonic** if whenever $c \in f(\succsim)$ and $c \notin f(\succsim')$ there is some player $i \in N$ and some outcome $b \in C$ such that $c \succsim_i b$ and $b \succ_i' c$.

That is, in order for an outcome c to be selected by a monotonic choice rule when the preference profile is $\succsim$ but not when it is $\succsim'$ the ranking of c relative to some other alternative must be worse under $\succsim'$ than under $\succsim$ for at least one individual.

An example of a monotonic choice rule f is that in which $f(\succsim)$ is the set of weakly Pareto efficient outcomes: $f(\succsim) = \{c \in C\colon$ there is no $b \in C$ such that $b \succ_i c$ for all $i \in N\}$. Another example is the rule f in which $f(\succsim)$ consists of every outcome that is a favorite of at least one player: $f(\succsim) = \{c \in C\colon$ there exists $i \in N$ such that $c \succsim_i b$ for all $b \in C\}$.

■ PROPOSITION 186.2 *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. Suppose that the choice rule $f: \mathcal{P} \to C$ is Nash-implemented by a game form $G = \langle N, (A_i), g\rangle$, $c \in f(\succsim)$, and $c \notin f(\succsim')$. Then there is an action profile a for which $g(a) = c$ that is a Nash equilibrium of the game $\langle G, \succsim\rangle$ but not of $\langle G, \succsim'\rangle$. That is, there is a player j and action $a_j' \in A_j$ such that $g(a_{-j}, a_j') \succ_j' g(a)$ and $g(a) \succsim_j g(a_{-j}, a_j')$. Hence f is monotonic. □

◇ EXAMPLE 186.3 (*Solomon's predicament*) The biblical story of the Judgment of Solomon illustrates some of the main ideas of implementation theory. Each of two women, 1 and 2, claims a baby; each of them knows

who is the true mother, but neither can prove her motherhood. Solomon tries to educe the truth by threatening to cut the baby in two, relying on the fact that the false mother prefers this outcome to that in which the true mother obtains the baby while the true mother prefers to give the baby away than to see it cut in two. Solomon can give the baby to either of the mothers or order its execution.

Formally, let a be the outcome in which the baby is given to mother 1, b that in which the baby is given to mother 2, and d that in which the baby is cut in two. Two preference profiles are possible:

$$\theta \text{ (1 is the real mother): } a \succ_1 b \succ_1 d \text{ and } b \succ_2 d \succ_2 a$$

$$\theta' \text{ (2 is the real mother): } a \succ_1' d \succ_1' b \text{ and } b \succ_2' a \succ_2' d.$$

Despite Solomon's alleged wisdom, the choice rule f defined by $f(\theta) = \{a\}$ and $f(\theta') = \{b\}$ is not Nash-implementable, since it is not monotonic: $a \in f(\theta)$ and $a \notin f(\theta')$ but there is no outcome y and player $i \in N$ such that $a \succsim_i y$ and $y \succ_i' a$. (In the biblical story Solomon succeeds in assigning the baby to the true mother: he gives it to the only woman to announce that she prefers that it be given to the other woman than be cut in two. Probably the women did not perceive Solomon's instructions as a strategic game form.)

The next result provides sufficient conditions for a choice rule to be Nash-implementable.

▶ DEFINITION 187.1 A choice rule $f: \mathcal{P} \to C$ has **no veto power** if $c \in f(\succsim)$ whenever for at least $|N| - 1$ players we have $c \succsim_i y$ for all $y \in C$.

■ PROPOSITION 187.2 *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. Let $f: \mathcal{P} \to C$ be a monotonic choice rule that has no veto power. We construct a game form $G = \langle N, (A_i), g\rangle$ that Nash-implements f as follows. The set of actions A_i of each player i is the set of all triples (p_i, c_i, m_i), where $p_i \in \mathcal{P}$, $c_i \in C$, and m_i is a nonnegative integer. The values $g((p_i, c_i, m_i)_{i \in N})$ of the outcome function are defined as follows.

- If for some $j \in N$ and some $(\succsim, c, m)$ with $c \in f(\succsim)$ we have $(p_i, c_i, m_i) = (\succsim, c, m)$ for all $i \in N \setminus \{j\}$ then

$$g((p_i, c_i, m_i)) = \begin{cases} c_j & \text{if } c \succsim_j c_j \\ c & \text{if } c \prec_j c_j. \end{cases}$$

j wants something that is dominated! makes no sense!

j wants something better for j: block!

everyone except possibly one player agrees...

if j is rational.. must be everyone else is lying! listen to j!

disagreement: shouting match!

- Otherwise $g((p_i, c_i, m_i)) = c_k$ where k is such that $m_k \geq m_j$ for all $j \in N$ (in the case of a tie the identity of k is immaterial).

This game form has three components. First, if all the players agree about the preference profile $\succsim$ and the outcome $c \in f(\succsim)$ to be implemented then the outcome is indeed c. Second, if there is almost agreement—all players but one agree—then the majority prevails unless the exceptional player announces an outcome that, under the preference relation announced by the majority, is not better for him than the outcome announced by the majority (which persuades the planner that the preference relation announced for him by the others is incorrect). Third, if there is significant disagreement then the law of the jungle applies: the player who "shouts loudest" chooses the outcome.

$f \subseteq NE$

We now show that this game form Nash-implements f. Let $c \in f(\succsim)$ for some $\succsim \in \mathcal{P}$. Let $a_i = (\succsim, c, 0)$ for each $i \in N$. Then (a_i) is a Nash equilibrium of the game $\langle G, \succsim \rangle$ with the outcome c: any deviation by any player j, say to $(\succsim', c', m')$, that affects the outcome has the property that the outcome is $c' \precsim_j c$. typo!

$NE \subseteq f$

Now let (a_i^*) be a Nash equilibrium of the game $\langle G, \succsim \rangle$ with the outcome c^*. We show that $c^* \in f(\succsim)$.

There are three cases to consider. First suppose that $a_i^* = (\succsim', c^*, m')$ for all $i \in N$ and $c^* \in f(\succsim')$. If $c^* \notin f(\succsim)$ then the monotonicity of f implies that there is a player $i \in N$ and $b \in C$ such that $c^* \succsim_i' b$ and $b \succ_i c^*$. But then the deviation by player i to the action $(\succsim, b, 0)$ changes the action profile to one that yields his preferable outcome b. Hence $c^* \in f(\succsim)$.

Second suppose that $a_i^* = (\succsim', c^*, m')$ for all $i \in N$ and $c^* \notin f(\succsim')$. If there is some $i \in N$ and outcome $b \in C$ such that $b \succ_i c^*$ then player i can deviate to $(\succsim', b, m'')$ for some $m'' > m'$, yielding the preferred outcome b. Thus c^* is a favorite outcome of every player; since f has no veto power we have $c^* \in f(\succsim)$.

Third suppose that $a_i^* \neq a_j^*$ for some players i and j. We show that for at least $|N| - 1$ players c^* is a favorite outcome, so that since f has no veto power we have $c^* \in f(\succsim)$. Since $|N| \geq 3$ there exists $h \in N \setminus \{i, j\}$; a_h^* is different from either a_i^* or a_j^*, say $a_h^* \neq a_i^*$. If there is an outcome b such that $b \succ_k c^*$ for some $k \in N \setminus \{i\}$ then k can profitably deviate by choosing $(\succsim', b, m'')$ for some $m'' > m_\ell$ for all $\ell \neq k$. Thus for all $k \in N \setminus \{i\}$ we have $c^* \succsim_k b$ for all $b \in C$. (Note that player i, unlike the other players, may not be able to achieve his favorite outcome by deviating since all the other players might be in agreement.) □

The interest of a result of this type, like that of the folk theorems in Chapter 8, depends on the reasonableness of the game form constructed in the proof. A natural component of the game form constructed here is that a complaint against a consensus is accepted only if the suggested alternative is worse for the complainant under the preference profile claimed by the other players. A less natural component is the "shouting" part of the game form, especially since shouting bears no cost here.

The strength of the result depends on the size of the set of choice rules that are monotonic and have no veto power. If there are at least three alternatives and $\mathcal{P}$ is the set of all preference profiles then *no* monotonic choice *function* has no veto power. (This follows from Muller and Satterthwaite (1977, Corollary on p. 417); note that a monotonic choice function satisfies Muller and Satterthwaite's condition SPA.) Thus the proposition is of interest only for either a nondegenerate choice *rule* or a choice function with a limited domain.

The game form in the proof of the proposition is designed to cover all possible choice rules. A specific choice rule may be implemented by a game form that is much simpler. Two examples follow.

Example 189.1 Suppose that an object is to be assigned to a player in the set $\{1, \ldots, n\}$. Assume first that for all possible preference profiles there is a single player who prefers to have the object than not to have it. The choice function that assigns the object to this player can be implemented by the game form in which the set of actions of each player is $\{Yes, No\}$ and the outcome function assigns the object to the player with the lowest index who announces *Yes* if there is such a player, and to player n otherwise. It is easy to check that if player i is the one who prefers to have the object than not to have it then the only equilibrium outcome is that i gets the object.

Now assume that in each preference profile there are two ("privileged") players who prefer to have the object than to not have it, and that we want to implement the choice rule that assigns to each preference profile the two outcomes in which the object is assigned to one of these players. The game form just described does not work since, for example, for the profile in which these players are 1 and 2 there is no equilibrium in which player 2 gets the object. The following game form does implement the rule. Each player announces a name of a player and a number. If $n - 1$ players announce the same name, say i, then i obtains the object unless he names a different player, say j, in which case j obtains the object. In any other case the player who names the largest number gets the could change to $(1,0,0)$?!

	Mine	Hers	Mine+
Mine	$(0,\epsilon,\epsilon)$	$(1,0,0)$	$(2,\epsilon,M)$
His	$(2,0,0)$	$(0,\epsilon,\epsilon)$	$(0,0,0)$
Mine+	$(1,M,\epsilon)$	$(0,0,0)$	$(0,2\epsilon,2\epsilon)$

Figure 190.1 A game form that implements the choice function considered in Example 190.1 in which the legitimate owner obtains the object. (Note that the entries in the boxes are outcomes, not payoffs.)

object. Any action profile in which all players announce the name of the same privileged player is an equilibrium. Any other action profile is not an equilibrium, since if at least $n-1$ players agree on a player who is not privileged then that player can deviate profitably by announcing somebody else; if there is no set of $n-1$ players who agree then there is at least one privileged player who can deviate profitably by announcing a larger number than anyone else.

◇ EXAMPLE 190.1 (*Solomon's predicament*) Consider again Solomon's predicament, described in Example 186.3. Assume that the object of dispute has monetary value to the two players and that Solomon may assign the object to one of the players, or to neither of them, and may also impose fines on them. The set of outcomes is then the set of triples (x, m_1, m_2) where either $x = 0$ (the object is not given to either player) or $x \in \{1,2\}$ (the object is given to player x) and m_i is a fine imposed on player i. Player i's payoff if he gets the object is $v_H - m_i$ if he is the legitimate owner of the object and $v_L - m_i$ if he is not, where $v_H > v_L > 0$; it is $-m_i$ if he does not get the object. There are two possible preference profiles, $\succsim$ in which player 1 is the legitimate owner and $\succsim'$ in which player 2 is.

King Solomon wishes to implement the choice function f for which $f(\succsim) = (1,0,0)$ and $f(\succsim') = (2,0,0)$. This function is monotonic: for example $(1,0,0) \succ_2 (2,0,(v_H+v_L)/2)$ and $(2,0,(v_H+v_L)/2) \succ_2' (1,0,0)$. Proposition 187.2 does not apply since there are only two players. However, the following game form (which is simpler than that in the proof of the proposition) implements f: each player has three actions, and the outcome function is that given in Figure 190.1, where $M = (v_H+v_L)/2$ and $\epsilon > 0$ is small enough. (The action "Mine+" can be interpreted

show first implement f
then show $\exists$ no other eq.

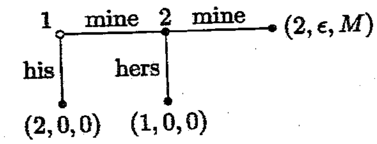


Figure 191.1 An extensive game form that implements the choice function given in Example 190.1. The vector near each terminal history is the outcome associated with that history.

as an impudent demand for the object, which is penalized if the other player does not dispute the ownership.)

Given our interest in the structure of the game forms that we construct, the fact that the game form in this example is simple and lacks a "shouting" component is attractive. In the next section (see Example 191.2) we show that the choice function in the example can be implemented by an even simpler scheme.

EXERCISE 191.1 Consider the case in which there are two individuals. Let $N = \{1,2\}$ and $C = \{a,b,c\}$, and suppose that there are two possible preference profiles, $\succsim$ with $a \succ_1 c \succ_1 b$ and $c \succ_2 b \succ_2 a$ and $\succsim'$ with $c \succ_1' a \succ_1' b$ and $b \succ_2' c \succ_2' a$. Show that the choice function f defined by $f(\succsim) = a$ and $f(\succsim') = b$ is monotonic but not Nash-implementable.

10.5 Subgame Perfect Equilibrium Implementation

Finally, we turn to the case in which the planner uses extensive game forms with perfect information and assumes that for any preference profile the outcome of the game may be any subgame perfect equilibrium (SPE). To motivate the possibilities for implementing choice rules in this case, consider Solomon's quandary once again.

◇ EXAMPLE 191.2 (*Solomon's predicament*) The choice function f given in the previous example (190.1) is SPE-implemented by the following game form. First player 1 is asked whether the object is hers. If she says "no" then the object is given to player 2. If she says "yes" then player 2 is asked if he is the owner. If he says "no" then the object is given to player 1, while if he says "yes" then he obtains the object and must pay a fine of M satisfying $v_L < M < v_H$ while player 1 has to pay a small fine $\epsilon > 0$. This game form is illustrated in Figure 191.1 (in which outcomes, not payoffs, are shown near the terminal histories).

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

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

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 (+ money)	$\Leftrightarrow$	every social choice rule

need to incorporate this into lit!!

implementation should be done w/ "acceptable" mechanisms!

If player 1 is the legitimate owner (i.e. the preference profile is $\succsim$) the game has a unique subgame perfect equilibrium, in which player 2 chooses "hers" and player 1 chooses "mine", achieving the desirable outcome $(1,0,0)$. If player 2 is the real owner then the game has a unique subgame perfect equilibrium, in which he chooses "mine" and player 1 chooses "his", yielding the outcome $(2,0,0)$. Thus the game SPE-implements the choice function given in Example 190.1.

The key idea in the game form described in this example is that player 2 is confronted with a choice that leads him to choose truthfully. If he does so then player 1 is faced with a choice that leads her to choose truthfully also. The tricks used in the literature to construct game forms to SPE-implement choice functions in other contexts are in the same spirit. In the remainder of the chapter we present a result that demonstrates the richness of the possibilities for SPE-implementation.

Let C^* be a set of deterministic consequences. We study the case in which the set C of outcomes has the form

$$C = \{(L,m): L \text{ is a lottery over } C^* \text{ and } m \in \mathbb{R}^N\}. \qquad (192.1)$$

If $(L,m) \in C$ then we interpret m_i as a fine paid by player i. (Note that m_i is *not* transferred to another player.)

We assume that for each player i there is a payoff function $u_i: C^* \to \mathbb{R}$ such that player i's preference relation over C is represented by the function $\mathrm{E}_L(u_i(c^*)) - m_i$; we identify a preference profile with a profile $(u_i)_{i\in N}$ of such payoff functions and denote $\mathrm{E}_L u_i(c^*)$ simply by $u_i(L)$. We assume further that $\mathcal{P} = U^N$, where U is a finite set that excludes the constant function. The set $\mathcal{G}$ of game forms that we consider is the set of extensive game forms with perfect information with consequences in C.

The notion of implementation that we explore is weaker than those studied previously: we construct a game form $\Gamma \in \mathcal{G}$ with the property that for any preference profile $u \in \mathcal{P}$ the game $\langle \Gamma, u\rangle$ has a unique subgame perfect equilibrium in which the desired alternative is realized with very high probability, though not necessarily with certainty. More precisely, we say that a choice function $f: \mathcal{P} \to C^*$ is **virtually SPE-implementable** if for any $\epsilon > 0$ there is an extensive game form $\Gamma \in \mathcal{G}$ such that for any preference profile $u \in \mathcal{P}$ the extensive game $\langle \Gamma, u\rangle$ has a unique subgame perfect equilibrium, in which the outcome is $f(u)$ with probability at least $1-\epsilon$.

(2) K k + 1

1 2 3 ... 1 2 ... |N| 1 2 ... |N|

PROPOSITION 193.1 *Let C^* be a set of deterministic consequences. Let $\langle N, C, \mathcal{P}, \mathcal{G}\rangle$ be an environment in which $|N| \geq 3$, C is given by (192.1), $\mathcal{P} = U^N$, where U is the (finite) set of payoff functions described above, and $\mathcal{G}$ is the set of extensive game forms with perfect information and consequences in C. Then every choice function $f: \mathcal{P} \to C^*$ is virtually SPE-implementable.*

Proof. First note that since for no payoff function in U are all outcomes indifferent, for any pair (v,v') of distinct payoff functions there is a pair $(L(v,v'), L'(v,v'))$ of lotteries over C^* such that $v(L(v,v')) > v(L'(v,v'))$ and $v'(L'(v,v')) > v'(L(v,v'))$. (A player's choice between the lotteries $L(v,v')$ and $L'(v,v')$ thus indicates whether his payoff function is v or v'.) For any triple (u,v,v') of payoff functions let $L^*(u,v,v')$ be the member of the set $\{L(v,v'), L'(v,v')\}$ that is preferred by u. Then for any pair (v,v') of payoff functions we have $u(L^*(u,v,v')) = \max\{u(L(v,v')), u(L'(v,v'))\}$, so that $u(L^*(u,v,v')) \geq u(L^*(u',v,v'))$ for any payoff function u'. Further, $u(L^*(u,u,u')) > u(L^*(u',u,u'))$.

Now suppose that for some pair (v,v') a player who announces the payoff function u is given the lottery $L^*(u,v,v')$. Let B be the minimum, over all pairs (u,u') of distinct payoff functions, of the average gain, over all pairs (v,v'), of any player with payoff function u from announcing u rather than u':

$$B = \min_{(u,u')\in\mathcal{W}} \left(\frac{1}{M} \sum_{(v,v')\in\mathcal{W}} \{u(L^*(u,v,v')) - u(L^*(u',v,v'))\} \right),$$

where $\mathcal{W}$ is the set of all pairs of distinct payoff functions and $M = |U|(|U|-1)$ (the number of members of $\mathcal{W}$). By the argument above we have $B > 0$.

For every $\epsilon > 0$ we construct a game form that has $K+1$ stages (K being defined below). Each stage consists of $|N|$ substages. Let $N = \{1,\ldots,n\}$. In substage i of each of the first K stages player i announces a preference *profile* (a member of U^N); in substage i of stage $K+1$ player i announces a payoff *function* (a member of U).

For any terminal history the outcome, which consists of a lottery and a profile of fines, is defined as follows. Each stage k for $k = 1,\ldots,K$ contributes to the lottery a consequence with probability $(1-\epsilon)/K$. If in stage k all the players except possibly one announce the same preference profile, say (u_i), then this consequence is $f((u_i))$; otherwise it is some fixed consequence $c^* \in C^*$.

(1) Show that can ~~induce~~ bound payoff from announcing preference truthfully above some $B>0$.

Each substage of the last stage contributes the probability $\epsilon/|N|$ to the lottery. This probability is split into M equal parts, each corresponding to a pair of payoff functions. The probability $\epsilon/|N|M$ that corresponds to $(i,(v,v')) \in N \times \mathcal{W}$ is assigned to the lottery $L^*(u_i', v, v')$, where u_i' is the payoff function that player i announces in stage $K+1$.

As for the fines, a player has to pay $\delta > 0$ if he is the last player in the first K stages to announce a preference profile different from the profile of announcements in stage $K+1$. In addition, a player has to pay a fine of δ/K for each stage in the first K in which all the other players announce the same profile, different from the one that he announces. (In order for the odd player out to be well-defined we need at least three players.)

Finally, we choose δ so that $\epsilon B/|N| > \delta$ and K so that $(1-\epsilon)D/K + \delta/K < \delta$, where

$$D = \max_{v,c,c'}\{v(c) - v(c')\colon v \in U,\ c \in C^*,\text{ and } c' \in C^*\}.$$

We now show that for any $(u_i) \in U^N$ the game $\langle \Gamma, (u_i) \rangle$, where Γ is the game form described above, has a unique subgame perfect equilibrium, in which the outcome is $f((u_i))$ with probability at least $1-\epsilon$. We first show that after every history in every subgame perfect equilibrium each player i announces his true payoff function in stage $K+1$. If player i announces a false payoff function at this stage then, relative to the case in which he announces his true payoff function, there are two changes in the outcome. First, the lotteries contributed to the outcome by substage i of stage $K+1$ change, reducing player i's expected payoff by at least $\epsilon B/|N|$. Second, player i may avoid a fine of δ if by changing his announcement in the last period he avoids being the last player to announce a preference profile in one of the first K stages that is different from the profile of announcements in the final stage. Since $\epsilon B/|N| > \delta$ the net effect is that the best action for any player is to announce his true payoff function in the final period, whatever history precedes his decision.

We now show that in any subgame perfect equilibrium all players announce the true preference profile (u_i) in each of the first K stages. Suppose to the contrary that some player does not do so; let player i in stage k be the last player not to do so. We argue that player i can increase his payoff by deviating and announcing the true preference profile (u_i). There are two cases to consider.

- If no other player announces a profile different from (u_i) in stage k then player i's deviation has no effect on the outcome; it reduces the fine he has to pay by δ/K, since he no longer announces a profile different from that announced by the other players, and may further reduce his fine by δ (if he is no longer the last player to announce a profile different from (u_i)).
- If some other player announces a profile different from (u_i) in stage k then the component of the final lottery attributable to stage k may change, reducing player i's payoff by at most $(1-\epsilon)D/K$. In addition he may become the odd player out at stage k and be fined δ/K. At the same time he avoids the fine δ (since he is definitely not the last player to announce a profile different from (u_i)). Since $(1-\epsilon)D/K + \delta/K < \delta$, the net effect is that the deviation is profitable.

We conclude that in every subgame perfect equilibrium every player, after every history at which he has to announce a preference profile, announces the true preference profile, so that the outcome of the game assigns probability of at least $1-\epsilon$ to $f((u_i))$. □

The game form constructed in this proof is based on two ideas. Stage $K+1$ is designed so that it is dominant for every player to announce his true payoff function. In the earlier stages a player may wish to announce a preference profile different from the true one, since by doing so he may affect the final outcome to his advantage; but no player wants to be the *last* to do so, with the consequence that no player ever does so.

Notes

The Gibbard–Satterthwaite theorem (181.2) appears in Gibbard (1973) and Satterthwaite (1975). For alternative proofs see Schmeidler and Sonnenschein (1978) and Barberá (1983). Proposition 184.2 is due to Groves and Loeb (1975); the result in Exercise 185.1 is due to Green and Laffont (1977). Maskin first proved Proposition 187.2 (see Maskin (1985)); the proof that we give is due to Repullo (1987). The discussion in Section 10.5 is based on Abreu and Matsushima (1992), who prove a result equivalent to Proposition 193.1 for implementation via iterated elimination of strictly dominated strategies in strategic game forms; the variant that we present is that of Glazer and Perry (1992). The analysis of Solomon's predicament in Examples 186.3, 190.1, and 191.2 first appeared in Glazer and Ma (1989).

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.

$C = \cancel{X} \times \mathbb{R}^n$

$u_i(x, \theta, t_i) = \cancel{v_i(x, \theta)} + t_i$

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. And (4) ~~Also~~, in mechanism design the 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.

market for kidneys, Roth

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 $P = \prod_{i \in N} P_i$ and $\Theta_i = [\underline{v_i}, \overline{v_i}]$. 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 $\langle N, (A_i)_{i \in N}, g \rangle$ where $g : (A_i)_{i \in N} \to X \times \mathbb{R}^n$ 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 $\langle N, (\Theta_i)_{i \in N}, (x, t) \rangle$, 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.

of mech. M

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

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

Another important definition is the following.

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

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

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.

the world takes place in the interim!! (we live in the interim)

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

(1)

$$\overset{2a}{\sum_k} V_k\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_k\right) + \cancel{\tau_i(\theta_{-i})} = V_i\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_i\right) + \sum_{j \neq i} V_j\left(x^*\left(\widehat{\theta}_i, \theta_{-i}\right), \theta_j\right) + \tau_i(\theta_{-i})$$

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

2b

or (2) 2a > 2b

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~~ would be identical to 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 in private values environments. 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:

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

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

The ex-post efficient decision rule is given by $x^*(\theta) = \left(0, \ldots, 0, \underset{i\text{th place}}{1}, 0, \ldots, 0\right)$ 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 $\max_{k \neq i}\left\{\widehat{\theta}_k\right\}$, 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:

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

and Ex-post budget balance requires that:

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

should redo analysis!!

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

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

$(1) + (5)$ 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

$$(1\cancel{4}5) \qquad (0, 0, 0, 0, 0, 0, 0, 0, c - 2L),$$

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

fortnote: why are Vickrey auctions ~~rarely used~~ rare? Roth hoff, Teisberg, Kahn, JPE 1990 - cheating (by seller!) - reluctance to reveal truth

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]

The seller acts as a "budget breaker" and so budget balance, 9 money, bidders, is not imposed.

[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.

If ~~the~~ the seller is introduced ~~to~~ as an additional agent, then because it

From: dolev bracha <dolev.br@gmail.com>
Subject: Re: תיקון-
Date: December 7, 2015 4:40:34 AM EST
To: Zvika Neeman <zvika@post.tau.ac.il>

1 Attachment, 341 KB

מצורף בpdf

On Sun, Dec 6, 2015 at 8:16 PM, Zvika Neeman <zvika@post.tau.ac.il> wrote:

Shalom Dolev,

Can you please email it to me in pdf ?

Thanks !

Zvika

Zvika Neeman
The Eitan Berglas School of Economics
Tel-Aviv University, Tel-Aviv 69978, ISRAEL

Email zvika@post.tau.ac.il
URL http://zvikaneeman.weebly.com/

On Dec 2, 2015, at 8:30 AM, dolev bracha wrote:

מצורף קובץ תיקון.

<התיקון לטעות ברשימות- צביקה.docx>

התיקון לטעות ברשימות

$$\begin{pmatrix} -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & L \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & L \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & -c \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & -c \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -c \end{pmatrix}$$

$$(9) = (1) + (5): (0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad L)$$

$$(10) = (2) + (6): (0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad -c)$$

$$(11) = (9) - (7): (0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad -1 \quad 0 \quad L + c)$$

$$(12) = (10) - (8): (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad -1 \quad 0)$$

$$(13) = (11) + (3): (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad -1 \quad 0 \quad 2L + c)$$

$$(14) = (12) - (13): (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad -1 \quad -c - 2L)$$

$$(15) = (14) + (4): (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad -c - 2L)$$

So $c = -2L$

סתירה לעובדה ששני הפרמטרים חיוביים ממש (אם הם שווים אפס, אין משמעות לסיפור)

this is a serious drawback compared to Groves mech. because it means they depend on the prior and on CPA!

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

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

and for all bidders

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

In the AAGV version of the 2nd price auction

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

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

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

θ_j

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 $\{x^* (\theta), t^* (\theta)\}$ satisfies budget balance, or such that $\sum_{i=1}^n t_i^* (\theta) = C_0 (x^* (\theta))$ for every θ.

III. SORTING

/ Adverse selection where the agent cannot effect this variable vs. Moral Hazard

Here we discuss a family of models in which there is interaction between an informed agent who has private information and an uninformed one who does not have private information. The models in this family have the special structure of the uninformed (the "principal") making a move ("offers a contract") to which the informed (the "agent") reacts. Many important economic models of asymmetric information belong to this family: models of second degree price discrimination, optimal income taxation and regulation.

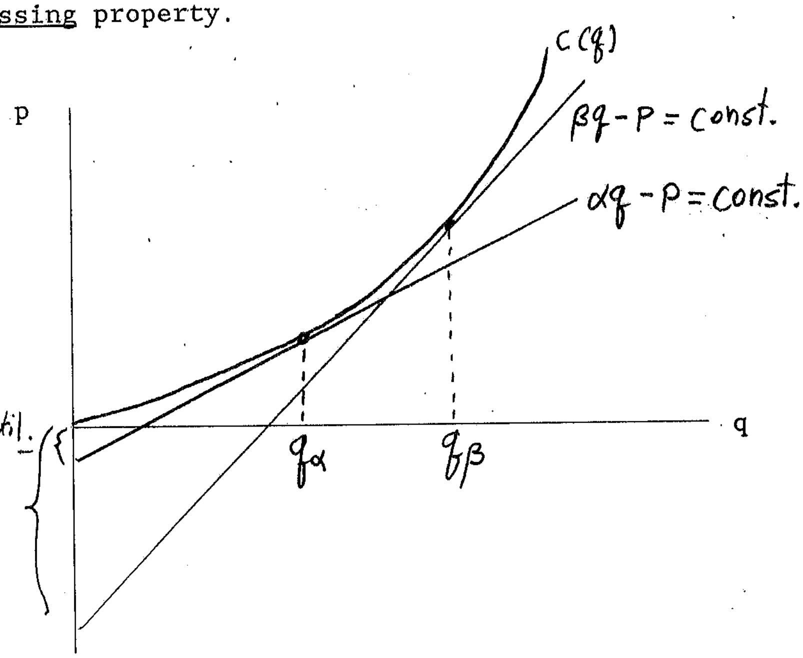
III.1. Second Degree Price Discrimination

III.1.1. The two types case

The example presented here is based on Mussa and Rosen (JET, 1978). A monopoly produces a product which can be of different quality levels. Quality will be denoted by q. Each consumer wants a unit. There are two types of consumer: α and β. Their utilities of buying a unit of quality q at the price p are $\alpha q-p$ and $\beta q-p$ respectively. Their respective numbers are N_α and N_β respectively. Thus, all consumers appreciate quality, but they differ with respect to their willingness to pay for it.

The cost of producing y units of quality q is $c(q)y$ where $c(0)=c'(0)=0$ and, for $q>0$, $c'(q)>0$ and $c''(q)>0$.

The following diagram depicts the technology and consumer preferences in the (q,p) plane. Note that an indifference curve of an α type crosses an indifference curve of a β type only once. This is an instance of what we call the <u>single crossing</u> property.



measures negative util. (linearity $\beta q-p=$ const.) $p=\beta q-$ const.)

$\beta q-p$ $\alpha q-p$

Let q_i be defined by $i=c'(q_i)$, where $i=\alpha,\beta$. The first best arrangement is that quality q_i is produced and distributed to type i consumes. If the

welfare max w.r.t. q_1, q_2 $\Rightarrow$ $c'(q_1)=\alpha$, $c'(q_2)=\beta$

consumer surplus $N_\alpha(\alpha q_1-p_1)+N_\beta(\beta q_2-p_2)$

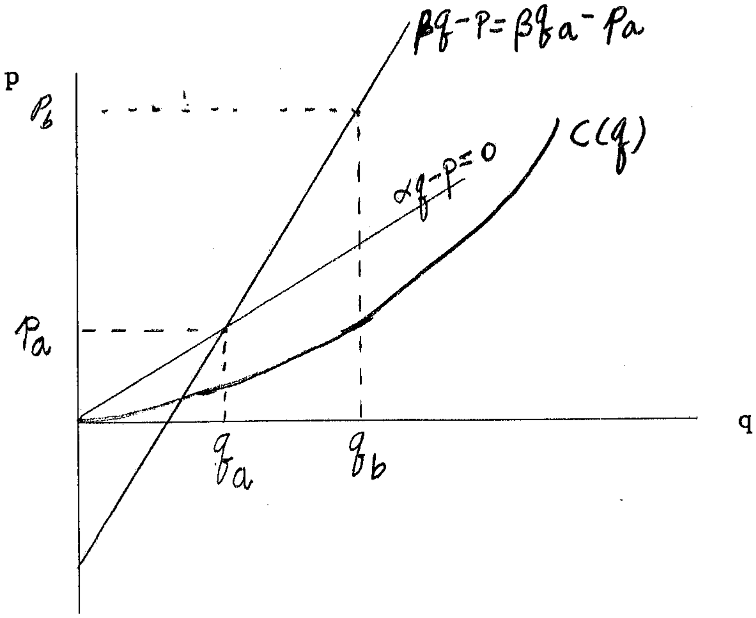
producers " $N_\alpha(p_1-c(q_1))+N_\beta(p_2-c(q_2))$

Total surplus $N_\alpha(\alpha q_1-c(q_1))+N_\beta(\beta q_2-c(q_2))$

monopoly problem perfect price discrimination → the same points

monopoly extracts all consumer surplus! $(q_\alpha, p_\alpha)(q_\beta, p_\beta)$!!

enough to consider 2 points since have only two types!



Start at q_α and q_β and look at the maximal prices that satisfy the constraints. By reducing q_a marginally and adjusting the prices, there is a zero first order effect on the profit made on each type α consumer, but positive first order effect (at the rate of β-α) on the profit made on each type β consumer.

The main insights of this example are: (i) The seller's effort to sort the customers creates a distortion of the menu of products relative to the first best solution; (ii) The distortion is concentrated at the lower end. (iii) The higher type gets some positive surplus, while the lower type gets a none. These characteristics of this solution, translated into the appropriate context, appear repeatedly in sorting problems: non-linear pricing, optimal taxation, regulation etc.

This problem discussed here is essentially identical to the problem of optimal non-linear pricing by a monopoly, which is also a second degree price discrimination problem where consumers are sorted by different price-quantity combinations.

III.1.2. Continuum of types (later version follows)

The above example captures many of the main ideas of sorting models. However, much of the literature deals with the multi-type case and this involves a characteristic technique which appears in similar forms in virtually all such analyses. The purpose of this part is to present this technique.

The buyers' set is a continuum with measure 1. Buyers's types are indexed by b and are distributed on [0,B] according to distribution function F. The von-Neumann-Morgenstern utility of a buyer with valuation b who gets quantity q for the payment t is v(q,b)-t and the utility of no purchase is 0. The seller has constant per unit cost c.

 $t = qp$

monotonicity of q() implies that it is differentiable a.e. Part (ii) then directly follows from application of the envelope theorem to v(q(b),b).

$U(b) = V(q(b),b) - t(b)$ $\qquad \frac{dU(b)}{db} = \underbrace{V_1(q(b),b)\,q'(b) - t'(b)}_{=0} + V_2(q(b),b)$ from maximization of buyer

From part (ii) of the proposition

$$(4) \qquad U(b)=U(0)+\int_0^b v_2(q(x),x)dx.$$

Max $V(q(b),b) - t(b)$

FONC $V_1(q(b),b)\,q'(b) - t'(b) = 0$

The interesting fact is that t(b) does not appear in this expression. Type b's surplus is expressed only in terms of the quantities bought by all lower types and of U(0). Put in other words, t(b) can also be expressed in terms of q() and U(0), since by the definition of U(b) and (4)

$$(5) \qquad t(b)=v(q(b),b)-U(b)=v(q(b),b)-\int_0^b v_2(q(x),x)dx-U(0)$$

The observation of (4) will be used next to express the seller's profit in terms of the schedule q(b) and the number U(0).

<u>Proposition 2</u>: The seller's profit SH is given by

$$(6) \qquad SH = \int_0^B[v(q(b),b)-cq(b)-(1-F(b)/f(b)v_2(q(b),b)]f(b)db - U(0).$$

<u>Proof</u>: Total buyers' surplus, SU, is

$$(7)\ SU = \int^B U(b)dF(b) \overset{\text{integ by parts}}{=} U(B)-\int^B[dU(b)/db]F(b)db \overset{(4)\ \text{integ by parts}}{=}$$

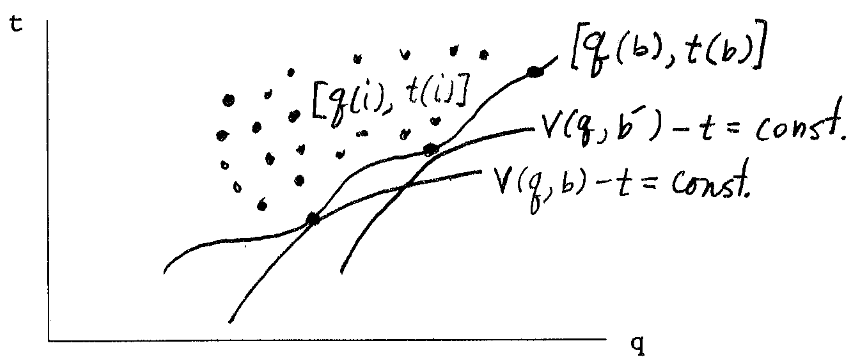
$$U(0)+\int^B(v_2(q(b),b))db-\int^B v_2(q(b),b)F(b)db \overset{(4)}{=} U(0)+\int^B v_2(q(b),b)[1-F(b)]db$$

The seller's profit on a valuation b buyer is H(b)=t(b)-cq(b). The surplus associated with the sale to type b is U(b)+H(b)=v(q(b),b)-cq(b). Therefore, the total surplus, SU+SH, is

$$(8) \qquad SU+SH = \int^B[v(q(b),b)-cq(b)]f(b)db$$

Subtract (7) from (8) to get SH=(SU+SH)-SU as given by (6). QED

So far we have only shown that, given an arbitrary schedule of quantities and required payments, higher consumer types will choose to buy at least as much as lower types, and that the seller's profit can be expressed in terms of selected quantities, q(b), b∈[0,B], and the lowest type's utility, U(0).



$$V_2(q(b'),b') = \frac{\int_{b'}^{b} V_2(q(b'),x)\,dx}{b-b'} \le \frac{U(b)-U(b')}{b-b'} \le \frac{\int_{b'}^{b} V_2(q(b),x)\,dx}{b-b'} = V_2(q(b),b)$$

q(b) increasing → at most countable points of discontinuity

take a point at which U(b) is cont. (assume U is cont. ?)

∃ an interval where U(b) is cont.

One aspect of the model that was crucial for the elegant treatment of this problem is the separability and linearity of the utility function in the payment.

qualitative assumption $v(q,s)-t$ linearity + separability (allows to cancel t!)
v_{12} sep

Problem Set III

1. Let W={a,b,c}. P(a)=P(c)={a,b,c}, P(b)={b,c}. A={B,N}. (N,w)=0 for all w. u(B,a)=-2, u(B,b)=-2, u(B,c)=3. q(w)=1/3, for all w.
(i) Which conditions does P satisfy?
(ii) What is the optimal decision rule for the problem (W,P,A,u,q)?
(iii) Find a partition Q which is coarser than P and such that the optimal decision rule with respect to it will give higher expected utility than the optimal rule you found in (ii).

2. Consider a market for a homogeneous product. The cost of production is C(y)=cy. Each buyer is interested in getting one unit of the product. There are two types of consumer: 1 and 2. Their utilities of buying a unit at the price p are a_i-p respectively, where $a_2>a_1>c$. Their respective numbers are N_1 and N_2 respectively.
(i) Characterize in terms of the parameters the monopoly price and quantity in this market.
(ii) Find the prices and quantities sold by the perfectly discriminating monopoly.

Suppose that the firm cannot identify buyers' types, but can offer different packages of the form (p,t), where p is a price and t waiting time for the product. Let w_i denote the cost incurred by buyer type i=1,2 by waiting a time unit. Assume $w_2>w_1$.
(iii) If it turns out that it is profitable for the firm to make only a single offer (p,t), what would be the value of t in such an offer?
(iv) Show that, if it is profitable for the firm to offer more than one (p,t) package, it has to be that $N_2(w_2-w_1)>N_1w_1$. Explain why.
(v) Assume that the parameters satisfy $a_1-w_1(a_2-a_1)/(w_2-w_1)>c$ and $N_2(w_2-w_1)>N_1w_1$. Describe the profit maximizing menu of (p,t) packages and the resulting consumers' behavior.
(vi) Compare the profit and welfare associated with cases (i) and (v). In what sense does (v) involve distortion?

3. Modify the continuum types model discussed in class as follows. Each buyer is interested in getting only one unit. The VNM utility of a type b from getting a unit at price p is b-p and the utility of no purchase is 0. Assume that [1-F(b)]/f(b) is a decreasing function of b.
(i) Suppose the seller is constrained to charge just one price. Show that the profit maximizing price satisfies p = c + [1-F(p)]/f(p).
(ii) Suppose that the seller can commit to a menu of offers [q(i),p(i)], where q(i) is the probability with which a consumer who chooses offer i will get a unit and p(i) is the price he will pay in the event that he gets a unit. Modify the analysis done in class to deal with this case. Prove that the menu that maximizes the seller's profit consists of a single price, which is the one you found in (i), and that any buyer can get the good at this price with probability 1.

III.1.2. Continuum of types

The two-types example already captures many of the main ideas of sorting models. However, much of the literature deals with the multi-type case and this involves a certain technique which appears in similar forms in virtually all such analyses. The purpose of this part is to present this technique.

The buyers' set is a continuum with measure 1. Buyers's types are indexed by v and are distributed on [0,V] according to distribution function F. The von-Neumann-Morgenstern utility of a buyer with valuation v who gets quantity q for the payment t is $u(q,v)-t$, where $u(0,v)=0$. The seller has constant per unit cost c.

Technical assumptions:
(i) F is twice differentiable. Let f denote its density.
(ii) u is twice continuously differentiable $u_1>0$, $u_2>0$, $u_{11}<0$ and $u_{12}>0$.

Thus, u is increasing and concave in q and a consumer with a higher v has both a higher willingness to pay for a given quantity and a higher marginal willingness to pay. The condition $u_{12}>0$ is the single crossing condition which appears repeatedly in this literature.

The seller can commit to a menu of offers from which the buyers can choose. We shall denote the menu by $[q(i),t(i)]$, $i\in I$, where $q(i)$ and $t(i)$ are the quantity received and payment made by a consumer who chooses offer i. The purpose of the analysis is to characterize the menu that maximizes the monopoly's profit under the assumption that consumers will react to it optimally. We shall consider only menus which form compact sets so that the consumers' problem would be well defined.

Given a menu of such offers, let $[q(v),t(v)]$ denote the offer selected optimally by type v. Let $U(v)$ denote buyer v's utility from his choice

$$U(v)\equiv u(q(v),v)-t(v).$$

The seller's profit on a type v buyer is $t(v)-cq(v)$. The seller's total profit is then

$$H = \int_0^\infty [t(v)-cq(v)]f(v)dv. \tag{1}$$

The fact that $[q(v),t(v)]$ is optimally selected by type v implies that, for any v and v',

$$u(q(v),v)-t(v) \geq u(q(v'),v)-t(v'), \tag{2}$$

and also

$$U(v)\geq 0, \qquad \text{for all v.} \tag{3}$$

Conditions (2) is the incentive compatibility (IC) constraint, and condition (3) is the individual rationality (IR)or voluntary participation constraint. Since, for all v, $U(v)\geq U(0)$, condition (3) can be replaced by where the last inequality follows from $u_{12}>0$. First rearrangement and then substitution from (7) yield

$$U(v) \geq u(q(v'),v) - [u(q(v'),v') - U(v')] = u(q(v'),v) - t(v'),$$

which means that [q(),t()] satisfy the IC constraints (2).

(ii) The monotonicity of q() implies that it is continuous a.e. Let v be a continuity point of q(v), and rewrite (6) as

$$u_2(q(v'),\theta')(v-v')=\int_{v'}^{v} u_2(q(v'),x)\,dx \leq U(v)-U(v') \leq \int_{v'}^{v} u_2(q(v),x)\,dx = u_2(q(v),\theta)(v-v'),$$

where θ,θ' are between v' and v. Divide through by $(v-v')$ and take the limits as v' approaching v. Since v is a continuity point of q(), $\lim_{v'\to v} u_2(q(v),\theta)=\lim_{v'\to v} u_2(q(v'),\theta')=u_2(q(v),v)$. Thus, U() is differentiable at any continuity point of q() and $dU(v)/dv=u_2(q(v),v)$. Also, from (6), U() is continuous at all points. Therefore, for all v,

(*) actually, need ↑q(), cont. here – but will talk about it later in course!

$$U(v)=U(0)+\int_0^v u_2(q(v),v). \tag{8}$$

It follows that

$$t(v)=u(q(v),v)-U(v)=u(q(v),v)-\int_0^v u_2(q(x),x)\,dx-U(0). \tag{9}$$

QED

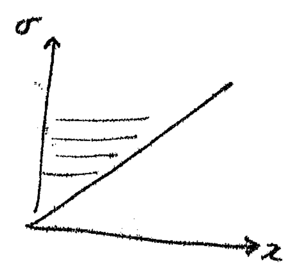
<u>Remarks</u>: (1) The single crossing property of the preferences was used in establishing monotonicity of any q() which satisfies the constraints (2).
(2) Another way to derive part (ii) is to notice that the monotonicity of q() implies that it is differentiable a.e. Part (ii) then directly follows from application of the envelope theorem to u(q(v),v).

Proposition 1 simplifies the maximization problem in two important ways. First, it shows that the set of menus which satisfy the constraints has a relatively simple form: it consists of all the menus in which q() is non-decreasing, t() is given by (9) and $U(0)\geq 0$. Second, it allows to substitute t() out and express the maximand and the constraints only in terms of q() and U(0).

Substitute (9) into (1) to get

$$H = \int_0^\infty \Big[u(q(v),v)-\int_0^v u_2(q(x),x)\,dx-cq(v)\Big]f(v)\,dv-U(0) \tag{10}$$

Integration by parts of the middle term yields



$$\int_0^\infty\int_0^v u_1(q(x),x)\,dx\,dv \quad f(v) =$$

$$\int_0^\infty\int_x^\infty u_2(q(x),x)f(v)\,dv\,dx = \int_0^\infty u_2(q(x),x)(1-F(x))\,dx$$

בכמה, גידול כזה מגדיל תועלת של טיפוס v מאוד

seller's problem as he contemplates increasing $q(v)$ by a small increment dq for some mass of buyer types around v. This will increase the revenue from each buyer type in that neighborhood by approximately $u_1(q(v),v)dq$ and hence the expected revenue by $[u_1(q(v),v)dq]f(v)$. On the other hand, this will increase the informational rent of all higher types by approximately $u_{12}(q(v),v)dq$ and hence decrease the expected revenue by $[u_{12}(q(v),v)dv][1-F(v)]$. Thus, the tradeoff facing the seller is exactly captured by the derivative of the virtual utility.
(3) The term $[1-F(v)]/f(v)$ is the inverse of the hazard rate and the condition that it is a decreasing function of v is often invoked in this literature for the purpose it serves here.
(4) The separability and linearity of the utility function in the payment helps the presentation, but is not crucial for deriving this type of results. If the buyer's utility is given by $u(q,t,v)$, then the single crossing condition requires $d[u_1(q,t,v)/u_2(q,t,v)]/dv \leq 0$. See Fudenberg-Tirole's text for analysis of this case.

Problem Set III

1. Let $W=\{a,b,c\}$. $P(a)=P(c)=\{a,b,c\}$, $P(b)=\{b,c\}$. $A=\{B,N\}$. $(N,w)=0$ for all w. $u(B,a)=-2$, $u(B,b)=-2$, $u(B,c)=3$. $q(w)=1/3$, for all w.
(i) Which conditions does P satisfy?
(ii) What is the optimal decision rule for the problem (W,P,A,u,q)?
(iii) Find a partition Q which is coarser than P and such that the optimal decision rule with respect to it will give higher expected utility than the optimal rule you found in (ii).

2. Consider a market for a homogeneous product. The cost of production is $C(y)=cy$. Each buyer is interested in getting one unit of the product. There are two types of consumer: 1 and 2. Their utilities of buying a unit at the price p are a_i-p respectively, where $a_2>a_1>c$. Their respective numbers are N_1 and N_2 respectively.
(i) Characterize in terms of the parameters the monopoly price and quantity in this market.
(ii) Find the prices and quantities sold by the perfectly discriminating monopoly.

Suppose that the firm cannot identify buyers' types, but can offer different packages of the form (p,t), where p is a price and t waiting time for the product. Let w_i denote the cost incurred by buyer type $i=1,2$ by waiting a time unit. Assume $w_2>w_1$.
(iii) If it turns out that it is profitable for the firm to make only a single offer (p,t), what would be the value of t in such an offer?
(iv) Show that, if it is profitable for the firm to offer more than one (p,t) package, it has to be that $N_2(w_2-w_1)>N_1w_1$. Explain why.
(v) Assume that the parameters satisfy $a_1-w_1(a_2-a_1)/(w_2-w_1)>c$ and $N_2(w_2-w_1)>N_1w_1$. Describe the profit maximizing menu of (p,t) packages and the resulting consumers' behavior.
(vi) Compare the profit and welfare associated with cases (i) and (v). In what sense does (v) involve distortion?

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 $\sum_{i=1}^n t_i^*\left(\widehat{\theta}\right) = \sum_{i=1}^n \left(T_i\left(\widehat{\theta}_i\right) - \frac{1}{n-1}\sum_{j\neq i} T_j\left(\widehat{\theta}_j\right)\right) = 0$ for every $\widehat{\theta}$ 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. !! $T_i(\theta_i)$ so what?

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~~ 4.5 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.

short introduction to Sec

(1) market design

(2) Coase Theorem.

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 $\langle q_1, q_2, q_3, q_4, t_1, t_2, t_3, t_4 \rangle$:

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

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

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

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:

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

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

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

and seller

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

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.

M a S ignored regularity on F & G? just that densities are bounded away from 0!

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 is optimal in the sense of maximizing ex-ante efficiency subject to IC and IR. The DA also has a continuum of other, less efficient eq. It also has a trivial "no-trade" eq.

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 $\beta : [0,\omega] \longrightarrow \mathbb{R}$ 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 $z = (\beta^{\mathrm{I}})^{-1}(b)$.

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

[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 $E[Y_1 | Y_1 \leq x]$, the expected revenue to the seller in a first price auction is given by

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

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

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

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

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

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:

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

subject to incentive compatibility and individual rationality.

Proposition 1 implies

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

By changing the order of integration,

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

which implies that

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

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

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

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

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

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.

positive/normative   Voltaire vs. Leibnitz (?)
Candid vs. Panglos (opera L. Bernstein)
"This is the best of all possible worlds"

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 (ECM 1985, 1988).

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 (JPE, 1989, pp. 1078-80).

space

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:

$$\delta(1a, 2a) = A, \qquad \delta(1a, 2b) = B$$
$$\delta(1b, 2a) = C, \qquad \delta(1b, 2b) = B$$

All ex post eff except for

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.

This would be another criticism of durability in addition to any implied by 2nd example

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]

research question

[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.

In this example, the efficiency loss that is due to durability is "small" (only 1/2 of ex-ante expected surplus)

Is it possible to argue that this efficiency loss is always small!!??

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~~ some *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.

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

italics 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 because it is dominated by the mechanism

δ^*	$2a$	$2b$
$1a$	B	A
$1b$	A	B

But δ 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. after alt. is accepted

The question of what environments admit ~~such~~ interim renegotiation-proof mechanisms according to the stronger definition that is suggested here, 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.

babbling eq. under δ^* generates payoff $\frac{1}{2} \cdot \frac{3}{2} + \frac{1}{2} \cdot 2 = 1\frac{3}{4}$

check relationship + Aumann's "strong eq" w/ coalition-proofness

4.11.2. Ex-Post Renegotiation

whinston + Peleg?

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.

check Fudenberg + Tirole?

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, → ex-post eff.!

(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.

⇒ it implies ex-post efficiency

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 $\widehat{p} - p_0 \in (1, 2)$. $1 \le \hat{p} - p_0 \le 2$
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.

notice that this implies that the "effective price" that the buyer pays for object is $p_2 - p_0$

4.12. Robust Mechanism Design

4.13. Collusion

Bergemann & Morris (2005)
Cremer & McLean (1988)
Neeman (2004)

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 w/ cost c each 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.

mechanism should also be [illegible] IR for seller

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.

common to all bidders!

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.]] $v_i \searrow a$!

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.

notice that in my example the social choice function also imposes restrictions on transfers (FSE), which is not usually imposed in lit.

(1) Wilson critique

(2) Robust MD

payoff type
belief type } type

Given a dist. over payoff types BaM say that a social choice function be implemented in a "robust way" if it can be implemented in all type spaces that induce the same given dist. over payoff types.

Econometrica, Vol. 73, No. 6 (November, 2005), 1771–1813

ROBUST MECHANISM DESIGN

BY DIRK BERGEMANN AND STEPHEN MORRIS[1]

The mechanism design literature assumes too much common knowledge of the environment among the players and planner. We relax this assumption by studying mechanism design on richer type spaces.

We ask when ex post implementation is equivalent to interim (or Bayesian) implementation for all possible type spaces. The equivalence holds in the case of *separable* environments; examples of separable environments arise (1) when the planner is implementing a social choice function (not correspondence) and (2) in a quasilinear environment with no restrictions on transfers. The equivalence fails in general, including in some quasilinear environments with budget balance.

In private value environments, ex post implementation is equivalent to dominant strategies implementation. The private value versions of our results offer new insights into the relationship between dominant strategy implementation and Bayesian implementation.

(3) give my example

(4) BaM have following result:

KEYWORDS: Mechanism design, common knowledge, universal type space, interim equilibrium, ex post equilibrium, dominant strategies.

"Wilson critique"

(1)

> Game theory has a great advantage in explicitly analyzing the consequences of trading rules that presumably are really common knowledge; it is deficient to the extent it assumes other features to be common knowledge, such as one player's probability assessment about another's preferences or information.
>
> I foresee the progress of game theory as depending on successive reductions in the base of common knowledge required to conduct useful analyses of practical problems. Only by repeated weakening of common knowledge assumptions will the theory approximate reality. Wilson (1987).

ex-post IC ⟹ robust implementation

⇍

provide sufficient conditions for ⇐ to hold and examples that show when it doesn't.

1. INTRODUCTION

THE THEORY OF MECHANISM DESIGN helps us understand institutions ranging from simple trading rules to political constitutions. We can understand institutions as the solution to a well-defined planner's problem of achieving some objective or maximizing some utility function subject to incentive constraints. A common criticism of mechanism design theory is that the optimal mechanisms solving the well-defined planner's problem seem unreasonably complicated. Researchers have often therefore restricted attention to mechanisms that are "more robust" or less sensitive to the assumed structure of the environment.[2] However, if the optimal solution to the planner's problem is too

explain sense in which ex-post IC is stronger than BIC but weaker than dominant strat. IC

[1]This research is supported by NSF Grant SES-0095321. We would like to thank the co-editor, three anonymous referees, and seminar participants at many institutions for helpful comments. We thank Bob Evans for pointing out errors in earlier examples and Sandeep Baliga, Matt Jackson, Jon Levin, Bart Lipman, Eric Maskin, Zvika Neeman, Andrew Postlewaite, Ilya Segal, and Tomas Sjöström for valuable discussions.

[2]Discussions of this issue are an old theme in the mechanism design literature. Hurwicz (1972) discussed the need for "nonparametric" mechanisms (independent of parameters of the model).

no SCF can be implemented in a way that is

that argue that ex-post IC!

check Jehiel, Meyer-ter-Vehn, Moldovanu ECM, 200

Zame

ceptable outcomes. The planner (*partially*) *implements*[6] the social choice correspondence if there exists a mechanism and an equilibrium strategy profile of that mechanism such that equilibrium outcomes for every payoff type profile are acceptable according to the SCC.[7] This is sometimes referred to as Bayesian implementation, but since we do not have a common prior, we will call it interim implementation.

While holding this environment fixed, we can construct many type spaces, where an agent's type specifies both his payoff type and his belief about other agents' types. Crucially, there may be many types of an agent with the same payoff type. The larger the type space, the harder it will be to implement the social choice correspondence, and so the more "robust" the resulting mechanism will be. The smallest type space we can work with is the "payoff type space," where we set the possible types of each agent equal to the set of payoff types and assume a common knowledge prior over this type space. This is the usual exercise performed in the mechanism design literature. The largest type space we can work with is the union of all possible type spaces that could have arisen from the payoff environment. This is equivalent to working with a "universal type space," in the sense of Mertens and Zamir (1985). There are many type spaces in between the payoff type space and the universal type space that are also interesting to study. For example, we can look at all payoff type spaces (so that the agents have common knowledge of a prior over payoff types but the mechanism designer does not) and we can look at type spaces where the common prior assumption holds.

In the face of a planner who does not know about agents' beliefs about other players' types, a recent literature has looked at mechanisms that implement the SCC in *ex post equilibrium* (see references in footnote 10). This requires that in a payoff type direct mechanism, where each agent is asked to report his payoff type, each agent has an incentive to tell the truth if he expects others to tell the truth, whatever their types turn out to be. In the special case of private values, ex post implementation is equivalent to dominant strategies implementation. If an SCC is ex post implementable, then it is clearly interim implementable on *every* type space, since the payoff type direct mechanism can be used to implement the SCC.

The converse is not always true. In Examples 1 and 2, ex post implementation is impossible. Nonetheless, interim implementation is possible on every type space. The gap arises because the planner may have the equilibrium outcome depend on the agents' higher order belief types, as well as their realized pay-

[6]"Partial implementation" is sometimes called "truthful implementation" or incentive compatible implementation. Since we look exclusively at partial implementation in this paper, we will write "implement" instead of "partially implement."

[7]In companion papers (Bergemann and Morris (2005a, 2005b)), we use the framework of this paper to look at full implementation, i.e., requiring that every equilibrium delivers an outcome consistent with the social choice correspondence.

counterparts in private values environments. In particular, we (1) identify conditions when Bayesian implementation on all type spaces is equivalent to dominant strategies implementation, (2) give examples where the equivalence does not hold, and (3) show how and when the equivalence may depend on type spaces richer than the payoff type space. While related questions have long been discussed in the implementation literature (e.g., Ledyard (1978) and Dasgupta, Hammond, and Maskin (1979))—we discuss the relationship in detail in the concluding Section 6—our questions have not been addressed even under private values.

The paper is organized as follows. Section 2 provides the setup, introduces the type spaces, and provides the equilibrium notions. In Section 3 we present in some detail three examples that illustrate the role of type spaces in the implementation problem and point to the complex relationship between ex post implementation on the payoff type space and interim implementation on larger type spaces. In Section 4 we present equivalence results for separable social choice environments. The separable environment includes as special cases all social choice *functions* and the quasilinear environment without a balanced budget requirement. Section 5 investigates the quasilinear environment with a balanced budget requirement. We conclude with a discussion of further issues in Section 6.

2. SETUP

2.1. *Payoff Environment*

We consider a finite set of agents $1, 2, \ldots, I$. Agent i's *payoff type* is $\theta_i \in \Theta_i$, where Θ_i is a finite set. We write $\theta \in \Theta = \Theta_1 \times \cdots \times \Theta_I$. There is a set of outcomes Y. Each agent has utility function $u_i : Y \times \Theta \to \mathbb{R}$. A social correspondence is a mapping $F : \Theta \to 2^Y \setminus \emptyset$. If the true payoff type profile is θ, the planner would like the outcome to be an element of $F(\theta)$. (1) "payoff types"

An important special case—studied in some of our examples and results—is a *quasilinear environment* where the set of outcomes Y has the product structure $Y = Y_0 \times Y_1 \times \cdots \times Y_I$, where $Y_1 = Y_2 = \cdots = Y_I = \mathbb{R}$, and a utility function

$$u_i(y, \theta) = u_i(y_0, y_1, \ldots, y_I, \theta) \triangleq v_i(y_0, \theta) + y_i,$$

which is linear in y_i for every agent i. The planner is concerned only about choosing an "allocation" $y_0 \in Y_0$ and does not care about transfers. Thus there is a function $f_0 : \Theta \to Y_0$ and

$$F(\theta) = \{(y_0, y_1, \ldots, y_I) \in Y : y_0 = f_0(\theta)\}.$$

Throughout the paper, this environment is fixed and informally understood to be common knowledge. We allow for interdependent types: one agent's payoff from a given outcome depends on other agents' payoff types. The payoff mechanism design. The foundations of this formalism are discussed in some detail in Section 2.5.

2.3. *Solution Concepts*

Fix a payoff environment and a type space $\mathcal{T}$. A mechanism specifies a message set for each agent and a mapping from message profiles to outcomes. Social choice correspondence F is interim implementable if there exists a mechanism and an interim (or Bayesian) equilibrium of that mechanism such that outcomes are consistent with F. However, by the revelation principle, we can restrict attention to truth-telling equilibria of direct mechanisms.[9] A direct mechanism is a function $f : T \to Y$.

DEFINITION 1: *A direct mechanism $f : T \to Y$ is interim incentive compatible on type space $\mathcal{T}$ if*

$$\int_{t_{-i}\in T_{-i}} u_i\big(f(t_i, t_{-i}), \widehat{\theta}(t_i, t_{-i})\big)\, d\widehat{\pi}_i(t_i) \geq \int_{t_{-i}\in T_{-i}} u_i\big(f(t_i', t_{-i}), \widehat{\theta}(t_i, t_{-i})\big)\, d\widehat{\pi}_i(t_i)$$

for all i, $t \in T$ and $t_i' \in T_i$.

The notion of interim incentive compatibility is often referred to as Bayesian incentive compatibility. We use the former terminology as there need not be a common prior on the type space.

DEFINITION 2: *A direct mechanism $f : T \to Y$ on $\mathcal{T}$ achieves F if*

$$f(t) \in F(\widehat{\theta}(t))$$

for all $t \in T$.

It should be emphasized that a direct mechanism f can prescribe varying allocations for a given payoff profile θ as different types, t and t', may have an identical payoff profile $\theta = \widehat{\theta}(t) = \widehat{\theta}(t')$.

DEFINITION 3: *A social choice correspondence F is interim implementable on $\mathcal{T}$ if there exists $f : T \to Y$ such that f is interim incentive compatible on $\mathcal{T}$ and f achieves F.*

[9] See Myerson (1991, Chapter 6).

(4)

2.4. *Questions*

Our main question is, When is F interim implementable on all type spaces? By requiring that F be interim implementable on all type spaces, we are asking for a mechanism that can implement F with no common knowledge assumptions beyond those in the specification of the payoff environment. In Sections 4 and 5, we provide sufficient conditions for ex post implementability to be equivalent to interim implementability on all type spaces, but Examples 1 and 2 in the next section show that it is possible to find social choice correspondences that are interim implementable on any type space but are not ex post implementable.

We also consider the implications of interim implementability on different type spaces. To describe these results, we must introduce some important properties of type spaces. A type space $\mathcal{T}$ is a *payoff type space* if each $T_i = \Theta_i$ and each $\widehat{\theta}_i$ is the identity map. Type space $\mathcal{T}$ is *finite* if each T_i is finite. Finite type space $\mathcal{T}$ has *full support* if $\widehat{\pi}_i(t_i)[t_{-i}] > 0$ for all i and t. Finite type space $\mathcal{T}$ satisfies the *common prior assumption* (with prior p) if there exists $p \in \Delta(T)$ such that

$$\sum_{t_{-i}\in T_{-i}} p(t_i, t_{-i}) > 0 \quad \text{for all } i \text{ and } t_i$$

and

$$\widehat{\pi}_i(t_i)[t_{-i}] = \frac{p(t_i, t_{-i})}{\sum_{t'_{-i}\in T_{-i}} p(t_i, t'_{-i})}.$$

The standard approach in the mechanism design literature is to restrict attention to a common prior payoff type space (perhaps with full support). Thus it is assumed that there is common knowledge among the agents of a common prior over the payoff types. A payoff type space can be thought of as the smallest type space embedding the payoff environment described above. Restricting attention to a full support, common prior, payoff type space is *with* loss of generality. We can relax the implicit common knowledge assumptions embodied in those restrictions by asking the following progressively tougher questions about interim implementability:

- Is F interim implementable on all full support common prior payoff type spaces?
- Is F interim implementable on all common prior payoff type spaces?
- Is F interim implementable on all common prior type spaces?
- Is F interim implementable on all type spaces?

We will see that relaxing common knowledge assumptions makes a difference. In particular, we will show that while the common prior assumption is of beliefs $(t_i^0, t_i^1, t_i^2, \dots)$. We want to require that high level types, which intuitively contain more information than lower level types, are consistent with lower levels. Formally, an infinite hierarchy is *coherent* if all higher level types have the same payoff-relevant type as lower level types and if the projection of their beliefs over other players' types onto lower level type spaces is consistent with lower level types' beliefs. We can let player i's possible types, T_i, be the set of all coherent infinite hierarchies of beliefs. The universal type space literature[14] shows that—under some topological assumptions—the set of types, i.e., infinite hierarchies, can be identified with pairs of payoff-relevant types and beliefs, so that, for each i, there exists a homeomorphism $f_i : T_i \to \Theta_i \times \Delta(T_{-i})$. Since each Θ_i is finite, such a construction is possible in our case. Now letting $\widehat{\theta}_i$ be the projection of f_i onto Θ_i and letting $\widehat{\pi}_i$ be the projection of f_i onto $\Delta(T_{-i})$, this canonical "known own payoff type" universal type space is an example of a type space $\mathcal{T} = (T_i, \widehat{\theta}_i, \widehat{\pi}_i)_{i=1}^I$, as described is Section 2.2, with the special property that for each $\theta_i \in \Theta_i$ and $\pi_i \in \Delta(T_{-i})$, there exists $t_i \in T_i$ such that $\widehat{\theta}_i(t_i) = \theta_i$ and $\widehat{\pi}_i(t_i) = \pi_i$.[15]

What is the connection between the explicit universal type space and the implicit type spaces we described above? An implicit type space has no "redundant types" if every pair of types differs at some level in their higher order belief types. Mertens and Zamir (1985, Property 5 and Proposition 2.16) show that any implicit type space that has no "redundant" types and satisfies some topological restrictions is a belief-closed subset of the universal type space (and the same result will be true in our setting). Thus modulo the redundancy and topological provisos, the union of all type spaces is the same as the universal type space.

How significant are the redundancy and topological restrictions required by Mertens and Zamir to show the equivalence of explicit and implicit type spaces? Heifetz and Samet (1999) show that—without topological restrictions—it is possible to find types that cannot be embedded in the universal type space.[16] In general, the no redundant types restriction is not innocuous either. To illustrate this point, consider the type space

$$T_1 = \{t_1, t_1'\},$$
$$T_2 = \{t_2, t_2'\},$$

[14]Mertens and Zamir (1985), Brandenburger and Dekel (1993), Heifetz (1993), Mertens, Sorin, and Zamir (1994).

[15]This "known own payoff type" universal type space has built in the feature that there is common knowledge that each agent i knows his payoff type θ_i. If instead we had allowed agents also to be uncertain about their own θ_i, we would be back to the standard universal type space concerning Θ, as constructed by Mertens and Zamir (1985).

[16]Heifetz and Samet (1998) provide a nonconstructive proof of the existence of a universal type space without topological restictions.

Inequalities (10) and (11) have a very simply structure. With very few exceptions, the payoffs that appear on the left- and right-hand sides of the inequalities are identical and only the transfers differ. These inequalities are generated either by true or misreported types, which induce only different transfer decisions but identical allocational decisions. The exceptions are the second and fifth inequality of agent 1, where a misreported type also leads to a different allocational decision. Rearranging the inequalities, we obtain

$$\begin{aligned}
&0 \geq y_{21} - y_{11}, \quad 0 \geq y_{11} - y_{16},\\
&-1 \geq y_{32} - y_{22}, \quad 0 \geq y_{22} - y_{21},\\
&0 \geq y_{43} - y_{33}, \quad 0 \geq y_{33} - y_{32},\\
&0 \geq y_{54} - y_{44}, \quad 0 \geq y_{44} - y_{43},\\
&-1 \geq y_{65} - y_{55}, \quad 0 \geq y_{55} - y_{54},\\
&0 \geq y_{16} - y_{66}, \quad 0 \geq y_{66} - y_{65}.
\end{aligned}$$

When we sum these twelve constraints, the transfers on the right-hand side of the inequalities cancel out and we are left with the desired contradiction for any arbitrary choice of probabilities, namely $-2 \geq 0$. The transfers cancelled out because the set of incentive constraints for agent 1 and agent 2 jointly formed a cycle through the type space.

4. SEPARABLE ENVIRONMENTS

We now present general results about the relationship between ex post implementability and interim implementability on different type spaces. The first result is an immediate implication from the definition of ex post equilibrium.

PROPOSITION 1: *If F is ex post implementable, then F is interim implementable on any type space.*

PROOF: If F is ex post implementable, then by hypothesis there exists $f^*: \Theta \to Y$ with $f^*(\theta) \in F(\theta)$ for all θ, such that for all i, all θ, and all θ_i',

$$u_i(f^*(\theta), \theta) \geq u_i(f^*(\theta_i', \theta_{-i}), \theta).$$

Consider then an arbitrary type space $\mathcal{T}$ and the direct mechanism $f: T \to Y$ with $f(t) = f^*(\widehat{\theta}(t))$. Incentive compatibility now requires true payoff types!

$$\begin{aligned}
t_i &\in \underset{t_i' \in T_i}{\arg\max} \int_{t_{-i} \in T_{-i}} u_i\big(f(t_i', t_{-i}), \big(\widehat{\theta}_i(t_i), \widehat{\theta}_{-i}(t_{-i})\big)\big)\, d\widehat{\pi}_i(t_i)\\
&= \underset{t_i' \in T_i}{\arg\max} \int_{t_{-i} \in T_{-i}} u_i\big(f^*\big(\widehat{\theta}_i(t_i'), \widehat{\theta}_{-i}(t_{-i})\big), \big(\widehat{\theta}_i(t_i), \widehat{\theta}_{-i}(t_{-i})\big)\big)\, d\widehat{\pi}_i(t_i).
\end{aligned}$$

~~reports about~~ true payoff types associated w. t!

There are two subsets of separable environments in which we are particularly interested.[20] First, there is the case of the single-valued private component where $Y_i = \{\overline{y}_i\}$ is a single allocation for all i. In this case, there exists a representation of the utility function $\widetilde{u}_i : Y_0 \times \Theta \to \mathbb{R}$ such that $\widetilde{u}_i$ depends only on the common component y_0 and the payoff type profile θ. Thus any social choice *function* is separable. Second, there is the case of the classic quasilinear environment (described in Section 2). In this case, we set, for each agent i,

$$Y_i = \mathbb{R},$$
$$\widetilde{u}_i(y_0, y_i, \theta) = v_i(y_0, \theta) + y_i,$$
$$F_i(\theta) = Y_i.$$

In the quasilinear environment, the common component $f_0(\theta)$ will often represent the problem of implementing an efficient allocation, so that

$$f_0(\theta) = \arg\max_{y_0 \in Y_0} \sum_{i=1}^{I} v_i(y_0, \theta).$$

Whereas the designer is only interested in maximizing the social surplus and the utilities are quasilinear, there are no further restriction on the private components, here the monetary transfers, offered to the agents. In contrast, in the next section, we shall investigate the quasilinear environment *with* a balanced budget requirement as a canonical example of a nonseparable environment. By requiring a balanced budget, the SCC contains an element of interdependence in the choice of the private components as the transfers have to add up to zero.

PROPOSITION 2: *In separable environments, if F is interim implementable on every common prior payoff type space $\mathcal{T}$, then F is ex post implementable.*

PROOF: Suppose that F can be interim implemented on all type spaces. Then, in particular, it must be possible to interim implement F on the type space where agents other than i have type profile θ_{-i}. Thus for each i and $\theta_{-i} \in \Theta_{-i}$, there must exist $g^{i,\theta_{-i}} : \Theta_i \to Y$ such that i has an incentive to truthfully report his type,

$$\widetilde{u}_i\big(g^{i,\theta_{-i}}(\theta_i), (\theta_i, \theta_{-i})\big) \geq \widetilde{u}_i\big(g^{i,\theta_{-i}}(\theta_i'), (\theta_i, \theta_{-i})\big) \tag{12}$$

for all $\theta_i, \theta_i' \in \Theta_i$, and such that F is implemented, so that

$$g^{i,\theta_{-i}}(\theta_i) \in F(\theta). \tag{13}$$

[20]We would like to thank an anonymous referee for suggesting that we incorporate these two special cases in the unified language of a separable environment.

Example: A Private Value Auction with 2 Bidders

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

in dominant strat.

- It is possible to extract full surplus with a Vickrey auction that is preceded by a lottery (λ_1, λ_2) that satisfies

$$\begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{3} \end{pmatrix}$$

- This can be done generically in any model with a **given** finite number of types as long as the number of any player's types is smaller or equal than the number of all other players' types.

However,

- As shown by Neeman (2004), a necessary condition for full-surplus extraction is that once each buyer's belief is identified, this belief pins down the buyer's valuation (almost surely) (Neeman called this property "beliefs determine preferences" or BDP)

$$type \mapsto \text{(preferences, beliefs)}$$

- BDP is satisfied in the previous example, but if we change it to,

1/3

	$v = 1$	$v = 1$	$v = 2$
$v = 1$	$\frac{13}{48}$	$\frac{1}{48}$	$\frac{1}{12}$
$v = 1$	$\frac{1}{48}$	$\frac{1}{48}$	$\frac{1}{12}$
$v = 2$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{3}$

1/6

" 1/6

then full surplus cannot be extracted.

- The example shows a failure of robustness against "higher-order-beliefs" (Bergemann and Morris, 2004)

Econometrica, Vol. 74, No. 4 (July, 2006), 1063–1107

ROBUSTLY COLLUSION-PROOF IMPLEMENTATION

BY YEON-KOO CHE AND JINWOO KIM [1]

A contract with multiple agents may be susceptible to collusion. We show that agents' collusion imposes no cost in a large class of circumstances with risk neutral agents, including both uncorrelated and correlated types. In those circumstances, any payoff the principal can attain in the absence of collusion, including the second-best level, can be attained in the presence of collusion in a way robust to many aspects of collusion behavior. The collusion-proof implementation generalizes to a setting in which only a subset of agents may collude, provided that noncollusive agents' incentives can be protected via an ex post incentive compatible and ex post individually rational mechanism. Our collusion-proof implementation also sheds light on the extent to which hierarchical delegation of contracts can optimally respond to collusion.

KEYWORDS: Robustly collusion-proof implementation, pairwise identifiability, subgroup collusion, hierarchical delegation.

1. INTRODUCTION

THERE HAS BEEN A GROWING INTEREST in studying collusion among asymmetrically informed agents, in various settings ranging from internal organization, regulation, and auctions, to oligopolistic competition.[2] Although most of these studies explore how agents can effectively collude against exogenously given institutions, a few recent studies have begun to investigate an *optimal* organizational/contractual response to agents' collusion. In particular, Laffont and Martimort (1997, 2000) have developed a modeling framework that integrates collusion as part of the general mechanism design analysis.[3] An important insight gained from this framework is that agents' asymmetric information imposes transaction costs on their abilities to carry out collusive arrangements.

[1]The authors are grateful for comments and suggestions from a co-editor and three referees, and from Gorkem Celik, Jacques Crémer, Ian Gale, Lucía Quesada, Larry Samuelson, Bill Sandholm, Sergei Severinov, Guofu Tan, and seminar participants at USC, UCSD, Korea University, and Yonsei University. The authors acknowledge warm hospitality they received from Yonsei University and the University of Wisconsin, respectively, during their visits. The authors have benefited from financial support from the Shoemaker Fellowship and from Yonsei University.

[2]Tirole (1986), Baliga and Sjöström (1998), Celik (2004), Faure-Grimaud, Laffont, and Martimort (2003), Severinov (2003), and Mookherjee and Tsumagari (2004) study collusion in internal organization and the value of delegation. Graham and Marshall (1987), McAfee and McMillan (1992), Mailath and Zemsky (1991), Marshall and Marx (2004), Brusco and Lopomo (2002), Caillaud and Jehiel (1998), and Esö and Schummer (2004) study collusion in one-shot auctions of various formats, while Aoyagi (2003), Blume and Heidhues (2002), Skrzypacz and Hopenhayn (2004), and Abdulkadiroğlu and Chung (2003) study collusion in repeated auctions.

[3]Earlier literature concerned about coalition formation in Groves' mechanisms includes Green and Laffont (1979) and Crémer (1996). The former paper envisions a coalition of symmetrically informed agents, whereas the latter allows for their possible asymmetric information. Although the latter framework resembles that of Laffont and Martimort, and even considers subgroup collusion, it restricts attention to dominant strategy implementation (at both grand and coalitional mechanism design) and does not consider participation constraints.

that gives the principal an ex post constant payoff equal to the original expected payoff. This mechanism forces the (grand) coalition to become a residual claimant of the entire surplus, after paying off the principal an ex post constant surplus, when it manipulates the outcome. That such a mechanism is implementable in the adverse selection setting is not obvious and will be an important part of our analysis. Also not obvious is that such a mechanism, if implementable, is immune to collusion. In fact, being the residual claimant, the coalition would prefer the first-best allocation over the intended allocation in case the latter involves distortion, so it will try to manipulate so that the former allocation arises. Yet, such a manipulation never succeeds. The reason is that the coalition faces an asymmetric information problem just like the principal in the original noncollusive mechanism design. This informational asymmetry means that an appropriate amount of information rent must be given to the members of the coalition to implement a particular allocation. However, since the principal is paid off to realize a desired level of surplus irrespective of the induced allocation, implementing any other allocation by the coalition would violate budget balancing.[4] (This intuition will become more transparent in Section 5, with the aid of a figure.) In short, by making the agents residual claimants, our mechanism forces them to internalize precisely the same amount of informational cost that the principal faces in noncollusive mechanism design, and in this sense exploits the coalitional transaction cost fully.

This idea of collusion-proof implementation does not rely on the agents' types being uncorrelated, although making the agents residual claimants while preserving their incentives proves more challenging in a correlated type environment. If there are only two agents, our method of collusion-proofing indeed does not work, much consistent with LM's (2000) finding in their two agents model. With more than two agents, however, given a reasonable type structure, our collusion-proof implementation works quite generally, implying again that the principal can attain any noncollusive payoff in a robustly collusion-proof fashion even with correlation. An important corollary of this result is that the principal can typically implement the first-best allocation and extract the entire rents from the agents even in the presence of collusive agents.

We then extend our analysis to consider a mechanism that would prevent collusion by a subgroup of agents. Although the issue of preventing collusion by a subgroup has rarely been analyzed before, it is practically relevant because in many settings, only a subgroup of agents is often in a position to collude. Collusion-proofing in this environment poses a new challenge because the coalition may prey on noncollusive agents as much as on the principal. Protecting the interests of noncollusive agents thus becomes an important consideration for the principal. Our collusion-proof implementation idea generalizes in a remarkable way to this problem: If *at least two* collusive agents are

[4]The intuition is the same as the one showing that implementing the first-best allocation would run a budget deficit in Myerson–Satterthwaite (1983) bargaining. The difference is that this problem is endogenously/deliberately created by our design to prevent collusion from being feasible.

(1) optimal collusion

Suppose now the agents can collude. It is easily seen that the second-price auction is susceptible to collusion. Prior to bidding, the firms can organize a knockout auction wherein the agents bid for the right to participate in the second-price auction uncontested; i.e., the loser bids 1 and the winner bids his cost.[5] Hence, with collusion, the buyer essentially pays the price of 1 to the winner of the knockout auction.

(3) anticipating collusion

Now consider a different mechanism. The buyer holds an auction in which the agents bid for a payment b_i and again the low bidder wins. The mechanism differs in the payment arrangement: The buyer pays a fixed amount, 2/3, to the losing (high) bidder, say j, who then pays the winning bidder its bid b_i to perform the job. Intuitively, the losing bidder is a "prime contractor" who "outsources" the job to the winning bidder and in the process finances the difference, $b_i - 2/3$.

Absent collusion, the bidding game has a unique equilibrium in which the agents adopt a symmetric increasing bidding strategy $\frac{1}{2} + \frac{1}{3}\theta$ for each type $\theta \in [0, 1]$. Consequently, the job is allocated efficiently as in the optimal mechanism and the buyer procures the good at the fixed price of 2/3. Since the allocation is the same and the buyer pays the same on average as in the (noncollusive) second-price auction, the revenue equivalence theorem implies that the interim payoffs of both firms are the same as in that game. Hence, it is equilibrium for both agents to participate in the auction game. In sum, the proposed mechanism implements the optimal procurement policy, in the absence of collusion. More importantly, the new mechanism is not susceptible to collusion. In the bidding game, the agents become residual claimants of the social surplus after paying a fixed amount of 2/3 to the buyer. Since the allocation is efficient, they have no incentive to collude in that bidding game. (*)

Greg's criticism: Buyers can collude before agreeing to participate ~~share info~~ Agree to not participate if both have costs > 2/3.

This example illustrates the main idea of preventing collusion, namely that of "selling the firm" to the agents. In what follows, this idea will be used to construct a general collusion-proof mechanism that works in a more complicated environment. The example also illustrates another feature of our collusion-proof mechanism, distinguished from the existing literature (e.g., LM (1997, 2000)). Unlike the traditional approach, our mechanism guarantees the buyer a desired level of ex post surplus, whether collusion actually occurs or not. Hence, in the example, the buyer could achieve the same outcome by delegating the procurement job to a "consortium" of agents (run by some uninformed third party) at a fixed price of 2/3; the consortium will then organize its own auction to allocate the job efficiently. Such delegation may provide a more practically relevant implementation of our mechanism.

[5]More precisely, they can organize a knockout auction in which the agents bid to pay their rivals for "uncontested bidding" in the official auction. This knockout auction game has a unique symmetric equilibrium in which an agent with cost θ bids $\frac{1}{3} - \frac{1}{3}\theta$. This equilibrium implements the direct revelation (strong) cartel mechanism studied by McAfee and McMillan (1992). A similar problem arises with the first-price auction.

Che and Kim (2006) show that this mechanism renders collusion ineffective and achieves the Myerson revenue if collusion takes place after the agents have agreed to participate in the mechanism. Indeed, under such a scenario collusion occurs "too late": by agreeing to participate the agents have already committed to pay the fixed fees that provide the principal with the Myerson revenue regardless of the agents' subsequent actions. But the only way for the cartel to fulfill this commitment, while still inducing the agents' participation in the collusive scheme and without breaking the budget, is to implement the Myerson allocation.

However, this mechanism fails to be collusion-proof if collusion takes place before the agents have agreed to participate in the mechanism. If the agents share the information about their valuations beforehand, then they can refuse to participate in the mechanism if both valuations are smaller than the total fixed fee Π^*. The principal's expected revenue then falls below the Myerson level, because she receives the total fixed fee Π^* only in case of a sale.

In the third mechanism the agents simultaneously decide whether to bid for the good or to stay out. If agent i stays out, then he does not get the good and his payment is zero. If agent i bids and the other agent stays out, then agent i gets the good and pays the price $\frac{5}{9}$ to the principal regardless of his bid. If both agents submit bids, then the highest bidder gets the good and pays his bid to his opponent, while the loser pays $\frac{5}{9}$ to the principal. There is a symmetric equilibrium where each agent i bids $\frac{1}{3}\theta_i + \frac{13}{36}$ if his valuation θ_i exceeds $\frac{1}{2}$, and stays out otherwise. This equilibrium results in the Myerson allocation, and the principal receives $\frac{5}{9}$ if she makes the sale, and 0 otherwise.

In this paper we show that the third mechanism renders collusion ineffective and achieves the Myerson revenue even when collusion takes place before the agents have agreed to participate in the mechanism. The Myerson allocation turns out to be *cartel interim efficient* when the cartel is facing such a mechanism: any alternative feasible allocation necessarily makes some types of some agents worse off and thus is vetoed. For example, consider a collusive mechanism that maximizes the sum of the agents' ex ante expected payoffs: the agents buy the good from the seller at the price $\frac{5}{9}$ and allocate the good to the agent with the highest valuation if and only if the highest valuation exceeds the price $\frac{5}{9}$. One can prove that such a collusive mechanism provides agents who have sufficiently high valuations with expected payoffs lower than those they expect to get through noncooperative play in the principal's mechanism and thus is vetoed.

3. Model

There is one principal who owns a good, and $n \geq 2$ agents. Each agent i has a valuation θ_i for the good, which is known only to him. Valuations are identically and independently distributed according to a continuous cumulative distribution function F with support $[\underline{\theta}, \overline{\theta}]$, where $0 \leq \underline{\theta} < \overline{\theta} < \infty$, and an everywhere positive differentiable density f. This distribution is common knowledge. We require the distribution to satisfy a standard condition on the hazard rates.

Example in Che and Kim (2006)

$$u_i\left(b_i,b_j;\theta_i\right)=\begin{cases} b_i-\theta_i & \text{if} \quad b_i<b_j \\ \frac{2}{3}-b_j & \text{if} \quad b_i>b_j \end{cases}$$

Suppose both bidders use the same strictly increasing bidding strategy $b:[0,1]\rightarrow\mathbb{R}_+$.

Bidder i of type θ_i problem:

$$\max_{b_i} E\left[\left(b_i-\theta_i\right)\cdot\mathbf{1}\left[b_i<b\left(\theta_j\right)\right]+\left(\tfrac{2}{3}-b\left(\theta_j\right)\right)\cdot\mathbf{1}\left[b_i>b\left(\theta_j\right)\right]\right]$$

$$=E\left[\left(b_i-\theta_i\right)\cdot\mathbf{1}\left[b^{-1}\left(b_i\right)<\theta_j\right]+\left(\tfrac{2}{3}-b\left(\theta_j\right)\right)\cdot\mathbf{1}\left[b^{-1}\left(b_i\right)>\theta_j\right]\right]$$

$$=\left(1-b^{-1}\left(b_i\right)\right)\left(b_i-\theta_i\right)+\int_0^{b^{-1}(b_i)}\left(\tfrac{2}{3}-b\left(\theta_j\right)\right)d\theta_j$$

FOC (evaluated at equilibrium $\theta_i=b^{-1}\left(b_i\right)$):

$$-\frac{\partial b^{-1}(b_i)}{\partial b_i}\left(b_i-b^{-1}\left(b_i\right)\right)+\left(1-b^{-1}\left(b_i\right)\right)+\left(\tfrac{2}{3}-b\left(b^{-1}\left(b_i\right)\right)\right)\frac{\partial b^{-1}(b_i)}{\partial b_i}=0$$

$$=>\left(1-b^{-1}\left(b_i\right)\right)=\left(2b_i-\tfrac{2}{3}-b^{-1}\left(b_i\right)\right)\frac{\partial b^{-1}(b_i)}{\partial b_i}=>$$

Try linear solution: $b^{-1}\left(b_i\right)=kb_i+c$

$$=>\left(1-kb_i-c\right)=\left(2b_i-\tfrac{2}{3}-kb_i-c\right)k$$

$$=>\begin{cases} -k=(2-k)\,k \\ (1-c)=\left(-\tfrac{2}{3}-c\right)k \end{cases} =>\begin{cases} k=3 \\ c=-\tfrac{3}{2} \end{cases}$$

$$=>b\left(\theta\right)=\tfrac{1}{3}\theta+\tfrac{1}{2}.$$