Some Beautiful Theorems with Beautiful Proofs

Dan Quint

Spring, 2014

Introduction – Why Are We Here?

The basic goal of this semester is to prove five or so elegant results in micro theory. The list:

1. *The First Welfare Theorem* – any Walrasian equilibrium gives a Pareto-efficient allocation
2. *Arrow's Impossibility Theorem* – individual preferences don't "aggregate up" to societal preferences well
3. *A "No Trade" Theorem* – access to different information can't be the sole basis for trade
4. *Revenue Equivalence and the Optimal Auction* – under certain conditions, many standard auctions are equally good for the seller, and the revenue-maximizing auction (out of all possible formats) can be determined
5. *Impossibility of Efficient Bilateral Trade* – when both buyer and seller have private information, no mechanism can realize all possible gains from trade

Of course, doing this will require learning a bunch of different setups/environments that are standard in micro theory. We'll see an exchange economy with production; a simple voting world with ordinal preferences over policies; a world with an unknown true state of the world and differential private information about that state; and the benchmark environment for analyzing auctions and other trading mechanisms, the independent private values setup. And we'll deal with Bayesian Nash equilibrium – the standard extension of Nash equilibrium to settings with incomplete information.

The lectures are generally organized around the result to be proven. We'll typically start by defining the environment we're considering, and the goal – what problem we're interested in. Then once we've got everything defined, we'll state the theorem, and prove it. And hopefully briefly discuss its relevance.

Lecture 1

The First Welfare Theorem

Any Walrasian equilibrium allocation is Pareto-efficient.

- Basic exchange economy – lots of consumption goods, lots of individuals endowed with some of each good
- Money is a means of exchange, but has no consumption value and nobody's endowed with it, it just facilitates trade
- Firms are technologies for turning some goods into some other goods
- Each good has a "market price" – nobody knows where prices come from, but everyone's free to trade as much as they want at the market price
  - So you can sell some of your endowment and use that money to buy other stuff you want
- The key assumption is *price-taking behavior* – individuals, and firms, assume that market prices are fixed and outside their own control
  - All you can do is trade, or not trade, at the prices you see
- We're thinking about *general equilibrium*, or *competitive equilibrium*, or *Walrasian equilibrium*, which is when market prices equate supply and demand for each good
  - Given market prices, individuals demand the best consumption bundle they can afford
  - Given market prices, firms choose production to maximize profit (and pay that profit out to their shareholders)
  - And markets clear
- Today's result: any such equilibrium is a Pareto-efficient allocation.

1 What Environment Are We In?

We begin by defining a very general exchange economy with production.

- There are M different goods
  - We'll let p_m denote the price per unit of good m, and $p \in (\Re^+)^M$ the vector of all prices
- There are N consumers
  - Each consumer $i \in \{1, 2, \ldots, N\}$ has some endowment $e_i \in (\Re^+)^M$ of goods...
  - ...and a utility function $u_i : (\Re^+)^M \to \Re$ over how much he/she consumes of each good
  - We'll let $x_i \in (\Re^+)^M$ denote how much i consumes of each good, giving utility $u_i(x_i)$
- There are F firms
  - Each firm $j \in \{1, 2, \ldots, F\}$ is represented by a *production set* $Y_j \subset \Re^M$, which consists of all feasible production plans for that firm
  - For example, if $M = 5$ and the vector $(-1, -1, 1, 0, 0) \in Y_j$, then firm j has the ability to turn one unit each of goods 1 and 2 into one unit of good 3
  - (Technology need not scale – if $y \in Y_j$, it does not imply that $2y \in Y_j$)
  - We'll let $y_j \in Y_j$ denote the production plan that firm j chooses
- The firms are owned by the consumers
  - Consumer i owns a share θ_{ij} of firm j, with $\sum_{i=1}^{N} \theta_{ij} = 1$ for each j
  - If firm j produces y_j at prices p, it earns profits $\pi_j = p \cdot y_j$, and pays $\theta_{ij}\pi_j$ to consumer i
- Finally, one assumption on preferences:

  each consumer's preferences are **locally non-satiated**:

  for any i, any x_i, and any $\epsilon > 0$, there's some x_i' such that $\|x_i' - x_i\| \leq \epsilon$ and $u_i(x_i') > u_i(x_i)$

2 What Are We Trying To Do?

We're interested in two things: Pareto efficient consumption plans, and general equilibrium.

2.1 Pareto efficiency

- Let $x = (x_1, x_2, \ldots, x_N) \in (\Re^+)^{MN}$ be a consumption plan for each consumer
- x is *feasible* if it's technologically possible to produce that much of each good: that is, if there is some production plan $(y_1, y_2, \ldots, y_F) \in Y_1 \times Y_2 \times \ldots \times Y_F$ such that

$$\sum_{i=1}^{N} x_i \quad \leq \quad \sum_{i=1}^{N} e_i + \sum_{j=1}^{F} y_j$$

- A feasible consumption plan x is *Pareto efficient* if it's not Pareto-dominated by any other feasible consumption plan: that is, if there does not exist a feasible plan $x' = (x'_1, x'_2, \ldots, x'_M)$ with

$$u_i(x'_i) \quad \geq \quad u_i(x_i)$$

for every i, with strict inequality holding for at least one i.

2.2 Walrasian Equilibrium

Walrasian equilibrium is defined, basically, as two things happening:

1. Consumers and firms are "price takers" – they take market prices as being outside their control – and behave optimally given the prices they see
2. Market prices are such that the demand for each good equals the supply, i.e., markets clear

Formally, a Walrasian equilibrium is a vector of prices p^*, a consumption plan for each consumer $x^* = (x_1^*, \ldots, x_N^*)$, and a production plan for each firm $y^* = (y_1^*, \ldots, y_F^*)$ such that:

- Given prices p^*, each firm is maximizing profits: for each $j \in \{1, 2, \ldots, F\}$,

$$y_j^* \quad \in \quad \arg\max_{y_j \in Y_j} p^* \cdot y_j$$

- Given prices p^* and their wealth (from both their endowment of goods and their income from the firms they own), each consumer is maximizing utility: for each i,

$$x_i^* \quad \in \quad \arg\max_{x_i} \left\{ u_i(x_i) \quad : \quad p^* \cdot x_i \leq p^* \cdot e_i + \sum_j \theta_{ij}(p^* \cdot y_j^*) \right\}$$

- Each market clears: element by element,

$$\sum_i x_i^* \quad = \quad \sum_i e_i + \sum_j y_j^*$$

3 Today's Big Result

Theorem (The First Fundamental Theorem of Welfare Economics)**.** *If* (p^*, x^*, y^*) *is a Walrasian equilibrium, then* x^* *is Pareto efficient.*

4 But before we prove this...

- Before we prove this, one more preliminary – **proof by contradiction**
- Many of our results will be proven this way
- We want to show that some claim B is true – or really, that it follows from some set of assumptions A
- So we want to show $A \longrightarrow B$
- Very often, it's easier to show that contrapositive: $\neg B \longrightarrow \neg A$
- Which can be summarized: "Suppose B was false. Show this leads to a contradiction. That proves B must be true."
- Classic example of this: Euclid's proof there must be an infinite number of prime numbers.
  - Suppose there was a finite set of prime numbers
  - Label them $\{p_1, p_2, \ldots, p_r\}$
  - Let $P = p_1 p_2 \cdots p_r + 1$
  - If P is prime, we have a contradiction, because it's bigger than any of the numbers on our list
  - If P is not prime, then some prime number divides it, call that p
  - p can't be on our list, because $P = ap + 1$; but then p is a prime that wasn't on our list, another contradiction
  - So either way, we've found a prime that wasn't on our list
  - Thus, the number of primes is infinite
- Now we're ready to prove the First Welfare Theorem

5 Proof of the First Welfare Theorem

- The proof is surprisingly straightforward. We will prove this by contradiction.
- Suppose the theorem is false, which means there is some Walrasian equilibrium (p^*, x^*, y^*) which is not Pareto-efficient
- Then x^* must be Pareto-dominated by some other feasible consumption plan

  Let $x' = (x'_1, x'_2, \ldots)$ be that consumption plan,

  and $y' = (y'_1, y'_2, \ldots)$ a feasible production plan that generates it $(\sum_i x'_i = \sum_i e_i + \sum_j y'_j)$
- First thing to show: **at prices p^*, x' must be more expensive than x^***
  - Since x' Pareto-dominates x^*, at least one guy *strictly* prefers x'_i to x^*_i; call him Bob
    - Since Bob chose x^*_{Bob} at prices p^*, he can't afford anything better
    - Since x'_{Bob} is better, that means he can't afford it – or $p^* \cdot x'_{Bob} > p^* \cdot x^*_{Bob}$
  - For everyone else, x'_i is *at least as good* as x^*_i – and this means it must also be *at least as expensive* at prices p^*
    - **I did this in a kind of hand-wavy way in class**, so let me do it properly here
    - Suppose this was false, i.e., for some i, $u_i(x'_i) \geq u_i(x^*_i)$ but $p^* \cdot x'_i < p^* \cdot x^*_i$
    - Let $\delta = p^* \cdot x^*_i - p^* \cdot x'_i$, and let $\epsilon = \frac{\delta}{\|p^*\|}$
    - Local non-satiation means there exists an x''_i with $\|x''_i - x'_i\| \leq \epsilon$ and $u_i(x''_i) > u_i(x'_i)$
    - Now, $p^* \cdot x''_i = p^* \cdot (x''_i - x'_i) + p^* \cdot x'_i = p^* \cdot (x''_i - x'_i) + (p^* \cdot x^*_i - \delta)$
    - Recall from years ago that $a \cdot b = \|a\|\|b\| \cos\theta$, where θ is the angle between vectors a and b; or more significantly, $a \cdot b \leq \|a\|\|b\|$
    - This means $p^* \cdot (x''_i - x'_i) \leq \|p^*\|\epsilon = \|p^*\|\frac{\delta}{\|p^*\|} = \delta$
    - So $p^* \cdot x''_i \leq \delta + (p^* \cdot x^*_i - \delta) = p^* \cdot x^*_i$
    - But $u_i(x''_i) > u_i(x'_i) \geq u_i(x^*_i)$, which contradicts x^*_i being chosen in equilibrium
    - So the contradiction proves that $p^* \cdot x'_i \geq p^* \cdot x^*_i$
    - (Note that this is the only step of the proof where we need local non-satiatedness.)
  - So if we sum up over all consumers,

$$\sum_{i=1}^N (p^* \cdot x'_i) \quad > \quad \sum_{i=1}^N (p^* \cdot x^*_i)$$

- Now, we assumed the production plan y' would generate x', or $\sum_i x'_i = \sum_i e_i + \sum_j y'_j$
- And since markets cleared in equilibrium, $\sum_i x^*_i = \sum_i e_i + \sum_j y^*_j$

- Plugging these into our last equation,

$$p^* \cdot \left(\sum_i e_i + \sum_j y'_j\right) \quad > \quad p^* \cdot \left(\sum_i e_i + \sum_j y^*_j\right)$$

or, subtracting $p^* \cdot \sum_i e_i$ from both sides and pulling the dot product inside the summation,

$$\sum_j p^* \cdot y'_j \quad > \quad \sum_j p^* \cdot y^*_j$$

- But if the left-side sum is bigger than the right, then at least one of the summands must be bigger – there's some firm k such that

$$p^* \cdot y'_k \quad > \quad p^* \cdot y^*_k$$

- But by definition, firm k was supposed to be maximizing profits at p^* by choosing y^*_k, and there's our contradiction!

6 Where do we go from here?

- We just proved: if we "get to" an equilibrium, it's efficient
- But we don't yet even know for sure an equilibrium exists
- Turns out, we can prove it does, but only with some additional assumptions – it's pretty easy to find examples of environments that do not have a Walrasian equilibrium.
- If there are no firms, just consumers, existence of an equilibrium is guaranteed if...
  - each u_i is continuous
  - each u_i is increasing
  - each u_i is concave
  - and every consumer has a strictly positive endowment of every good ($e^i \gg 0$)

  But if any of these conditions is violated, you can generate examples with no equilibrium.
- For one simple example, take two goods, and let $u_1(x, y) = \min\{x, y\}$ and $u_2(x, y) = \max\{x, y\}$, which is convex

  Let $e^1 = e^2 = (1, 1)$

  The problem is that at any positive prices, the first guy demands equal amounts of the two goods, which means he refuses to trade; but the second guy demands to trade his endowment of one good for as much as he can afford of the other good; so markets can't clear at positive prices

  But if either price is 0, the second guy demands an infinite amount of that good, and again, markets can't clear

- For another example, suppose there are two goods and two consumers, and endowments are not strictly positive: $e_1 = (10, 0)$ and $e_2 = (0, 10)$. Utility functions are $u_1(x, y) = x$ and $u_2(x, y) = \sqrt{x} + y$. It turns out, no prices exist that clear markets:
  - If either price is zero, player 2 will demand an infinite amount of that good, so markets won't clear
  - If both prices are strictly positive, player 1 will demand exactly his endowment; but player 2 will demand some of good 1, so markets won't clear
- Even if we know a Walrasian equilibrium exists, we haven't said why we would expect it to be reached...
- But if we do believe that Walrasian equilibrium will naturally occur, then the First Welfare Theorem is sort of the competitive-markets analog of the Coase Theorem – let people trade, and an efficient outcome will inevitably be reached
- (The "price-takers" assumption can be thought of as the assumption that each person, and each firm, is small relative to the "market," so that everyone believes his own impact on market prices is negligible.)
- There's also the Second Welfare Theorem – that given an economy, *any* Pareto-efficient consumption plan is a Walrasian equilibrium for some set of prices and endowments – but this also requires some additional assumptions
  - Basically, we need continuous, concave utility functions and convex production sets (if we allow firms)
  - Under those assumptions, any Pareto-efficient allocation is a Walrasian equilibrium for some set of initial endowments

References

- Mas-Colell, Whinston and Green (1995), *Microeconomic Theory*, Oxford University Press (the standard first-year PhD micro textbook), ch. 16.

Lecture 2

Arrow's Impossibility Theorem

Aggregating individual preferences is hard.

1 What Environment Are We In?

- Finite set $\mathbf{A} = \{A, B, C, \ldots\}$ of at least three different policy options
- Finite number N of different individuals $i = 1, 2, \ldots, N$
- Each person i has preferences over the policy options $\succsim_i$, which are *complete* and *transitive*, and for simplicity, we'll assume they are *strict* as well (no indifferences)
  - so for any individual i and any two policies a and b, either $a \succ_i b$ or $b \succ_i a$
  - and for any three policies a, b, and c, if $a \succ_i b$ and $b \succ_i c$, then $a \succ_i c$
- And that's it.

2 What Are We Trying To Do?

- We're looking for a way to *aggregate preferences* – that is, we want a way to turn each possible set of individual preferences $\{\succ_1, \succ_2, \ldots, \succ_N\}$ into a preference relation $\succsim^*$ for "society"
  - We won't require $\succsim^*$ to be strict
- The mapping from individual preference relations into social preferences relations is alternatively called a *social welfare function*, or a *preference aggregation rule*, or a *Constitution*
- I'll use *social welfare function* – just remember that the SWF is not a set of preferences itself, but a rule for generating a set of preferences for society for each set of individual preferences
  - So if we let X be the set of all possible preference relations over the set of policies $\mathbf{A}$, a SWF is a mapping from X^N to X
- We want our social welfare function to satisfy certain properties

2.1 Social preferences need to be defined for any set of individual preferences.

- Our SWF has to specify some set of social preferences $\succsim^*$ for *any* given set of individual preferences $\{\succ_1, \succ_2, \ldots, \succ_N\}$
- This is also called "universal domain" – we're not ruling out any possible preferences.

2.2 Social preferences should be complete and transitive.

- Just like individual preferences $\succsim_i$ need to be complete and transitive to be "reasonable"...
- we want the social preferences $\succsim^*$ chosen by our SWF to be complete and transitive for each set of individual preferences
- (otherwise we won't have a coherent choice rule for society)
- This rules out something like pairwise majority voting

  Why? Suppose there are three people and three policies, with preferences

$$A \succ_1 B \succ_1 C$$
$$B \succ_2 C \succ_2 A$$
$$C \succ_3 A \succ_3 B$$

  If we go by majority rule, two out of three prefer A to B, and two out of three prefer B to C, and two out of three prefer C to A; so social preferences would have to be $A \succ^* B \succ^* C$ but $C \succ^* A$, which isn't transitive.

2.3 Social preferences should respect unanimity.

- If *everyone* in society agrees that policy a is strictly better than b, then the social preferences defined by our SWF should also strictly prefer a to b.
- If not, it's doing a pretty bad job of aggregating the individuals' preferences.
- This rules out stupid rules like "society ranks all options equally" or "regardless of individual preferences, society ranks policies alphabetically"

2.4 The SWF should satisfy *independence of irrelevant alternatives*

- Basically, this says that if we're trying to figure out whether society prefers a to b, what people think of c shouldn't matter.
- Formally, suppose we start with some set of individual preferences $\{\succ_1, \ldots, \succ_N\}$, and the SWF picks a social preference function under which $a \succsim^* b$

  Now modify one guy's preferences $\succ_i$ such that his preference between a and b stays the same, but his preferences for other things (including whether he prefers c to a or b) changes

  Then under the new preferences, the SWF should still pick a social preference function that prefers a to b.
- This rules out a rule like the Borda count.

  Suppose there are three policies, and we ask everyone for a rank-order list; and then we give a policy 3 points for each person who ranked it first, 2 points for each person who ranked it second, and 1 point for every person who ranked it third, and we say society prefers policies with more points.

  This rule violates IIA. For example, with two voters, if

  $$\begin{array}{ccccc} A & \succ_1 & B & \succ_1 & C \\ B & \succ_2 & C & \succ_2 & A \end{array}$$

  then the Borda count would prefer B to A. But if preferences were

  $$\begin{array}{ccccc} A & \succ_1 & C & \succ_1 & B \\ B & \succ_2 & A & \succ_2 & C \end{array}$$

  it would prefer A to B. But in both cases, $A \succ_1 B$ and $B \succ_2 A$ – all we did was change parts of preferences that "aren't supposed to matter" for choosing between A and B. So a Borda count rule violates IIA.

2.5 (Aside: note the ordinality of preferences)

- Note that preferences here are ordinal, not cardinal
- All we know is whether an individual prefers one policy to another – we have no language to even talk about *how much* he prefers one to another
- We don't have money, to allow people to barter (I prefer A to B, but I'll prefer B if you give me \$11), or anything like that
- All we have is ordinal pairwise preferences, and we need to aggregate them to preferences for society

2.6 What's Left?

- That's what Arrow's theorem tells us. Not much.
- One more thing to define: a SWF is a *dictatorship* if the social preference always just reflects the same one guy's preferences, that is, if there's some individual k such that regardless of anyone else's preferences, $a \succ^* b$ if and only if $a \succ_k b$.

3 The Result

Theorem (Arrow)**.** *Any SWF which respects transitivity, unanimity, and independence of irrelevant alternatives is a dictatorship.*

- So basically, you have four choices:

  1. you can violate transitivity (with something like majority voting)
  2. you can violate unanimity (with something like, social preferences are fixed regardless of individual preferences)
  3. you can violate independence of irrelevant alternatives (with something like the Borda count)
  4. or you can have a dictatorship

  But those are the only choices.

- The proof has several steps. We assume we have a SWF which satisfies transitivity, unanimity, and IIA, and then show that this means there must be some voter whose preferences always match the social preferences, who is therefore a dictator.

4 The Proof

- Throughout the proof, we will maintain the assumptions that social preferences $\succsim^*$ are always transitive, and that the SWF satisfies unanimity and IIA.

4.1 Part 1 – the Extremal Lemma

Lemma 1 (Extremal Lemma)**.** *For any policy b, if every individual i ranks b* either *strictly best or strictly worst, then $\succsim^*$ must rank b either strictly best or strictly worst as well.*

- We'll prove this by contradiction.
- Suppose the lemma were false. Then there would be some set of individual preferences and two other policies a and c such that $a \succsim^* b \succsim^* c$, even though each individual has b as either their favorite or least favorite policy.
- Now modify individual preferences as follows.

  For each individual i who has b at the top of their list: move c up to second on the list, so it is now strictly above a if it wasn't already.

  And for each individual i who has b at the *bottom* of their list, move c to the top of the list, so it's strictly above a if it wasn't already.
- This doesn't change any individual's preferences between b and a: if b was at the top of your list, it's still at the top; and if it was at the bottom, it's still at the bottom. So by IIA, society still prefers a to b
- And similarly, it doesn't change any individual's preferences between b and c, so by IIA, b is still preferred to c.
- And by transitivity, $a \succsim^* b$ and $b \succsim^* c$ imply $a \succsim^* c$
- But now, *everyone* has c ranked above a, so unanimity would require $c \succ^* a$, giving a contradiction
- So that proves that b must be either strictly best or strictly worst according to $\succsim^*$ – if it wasn't, we could generate this type of contradiction.

4.2 Part 2 – Find a pivotal guy and give him a name.

- Pick some random policy b
- We know (*unanimity*) that if *everyone in society* puts b last, then $\succsim^*$ puts b last too; and if everyone puts b first, $\succsim^*$ puts b first
- And we just showed that if some people put b first and the rest put b last, then $\succsim^*$ puts b *either* first or last
- So start out with a set of individual preferences where *everyone in society puts b last.* By unanimity, $\succsim^*$ must put b last as well.
- Now change voter 1's preferences by moving b from last to first. Since everyone in society either likes b the most or the least, b must be either first or last in $\succsim^*$.
- Now change voter 2's preferences by moving b from last to first. Again, since everyone in society either likes b the most or the least, b must be either first or last in $\succsim^*$.
- Keep going like this. By the time we've switched *everyone's* preferences to having b first, by unanimity, society must have b first as well.
- Now find the voter where the first switch happened. That is, the first time that the social preference $\succsim^*$ switched from having b at the bottom to b at the top. Call him Bob.

  That is, given the other preferences we started with, if everyone with a lower number than Bob has b first on their list, and Bob and everyone after has b last on their list, then society puts b last; but if Bob switches to having b first, then society puts b first.
- Call the first profile of preferences – where everyone up to Bob puts b first, and Bob and the rest put it last – profile I; and call the second set of preferences – where everyone up to Bob, and Bob himself, put b first, and the rest put it last – profile II.
- The rest of the proof involves showing that Bob is a dictator – that to satisfy IIA and transitivity, $\succsim^*$ has to *always* agree with Bob's preferences, regardless of what everyone else thinks

4.3 Part 3 – Proving Bob is a dictator

Part 3a: Bob is a dictator over policies that aren't b.

- That is, for any two policies a and c which aren't b, we'll show that if Bob prefers a to c, then no matter what everyone else's preferences are, $\succsim^*$ must put a strictly ahead of c as well.
- Start with arbitrary preferences where $a \succ_{Bob} c$, and call these (true) preferences Profile IV.
- Make the following changes to preferences:

  Move a to the top of Bob's preference list, and b to second on Bob's preference list

  For individuals 1 up to Bob, move b to the top of their list

  For individuals Bob +1 up to N, move b to the bottom of their list.

  Call this new set of preferences Profile III
- When we moved from IV to III, we didn't change anyone's ranking of a versus c – for Bob, a was above c, and we moved it to the top; for everyone else, all we did was move b

  So by IIA, the societal preference between a and c has to be the same at profile IV as at profile III

  So we'll show that at profile III, society has to prefer a to c
- Now, recall that at profile I, society put b last, which means $a \succ^* b$ at profile I.

  And everyone's ranking of a versus b is the same at profile I as at profile III.

  So by IIA, $a \succ^* b$ at profile III.
- At profile II, on the other hand, society put b first, which means $b \succ^* c$ at profile II

  And everyone's ranking of b versus c is the same at II as at III

  So by IIA, $b \succ^* c$ at profile III.
- Since preferences at profile III must be transitive, $a \succ^* c$.
- So at profile IV, $a \succ^* c$.
- So whenever $a \succ_{Bob} c$, $a \succ^* c$.

Part 3b: Bob is also dictator when it comes to b

- All that's left is to show is that Bob is also a dictator when one of the choices being considered is b
- This is the easy part. We need to show that if $b \succ_{Bob} a$, $b \succ^* a$; and if $a \succ_{Bob} b$, $a \succ^* b$.
- So now pick a policy that's different from those two – say, c – and repeat everything we already did
- Starting with the extremal lemma, move preferences from c worst to c best, find the pivotal guy, and prove that he has to be a dictator for *any two policies that aren't c*
  - So far, we don't know whether this new dictator is Bob or someone else
  - Just that *someone* is a dictator when c isn't involved, meaning, when choosing between a and b
- But we already know that Bob's preferences over b *sometimes* matter
  - When we moved from profile I to II, Bob's preferences for b were all that changed
    And that shifted society from putting b last to putting b first
- So if the new dictator isn't Bob, we have a contradiction; so the new dictator must also be Bob
- Which means Bob is dictator over any pair of policies that excludes b, *and* over any pair of policies that includes b
- So Bob's your dictator, and we're done. $\square$

5 So What?

- So the result is, if you want to aggregate individual preferences and get a transitive social preference that respects unanimity and IIA, all that's left is a dictatorship.
- One way to interpret this is, IIA is a really strong restriction
  - Basically, by assuming IIA, we're ruling out inferring anything "cardinal" about preferences from where you rank other alternatives
  - If there are 100 policy choices, and you have a and b ranked #45 and #46, chances are, you're pretty close to indifferent between them

    OTOH, if you have a ranked #1 and b ranked #100, you probably like a a lot more

    IIA says that for choosing between a and b, we have to treat those two cases the same – all we're allowed to consider is that you like a more than b

    By assuming IIA, we're really ruling out having any way to elicit cardinal preferences

    Which is what makes this hard
- Contrast that with the first welfare theorem
  - In some sense, markets behave well exactly because prices elicit *cardinal* information
  - How much you're willing to buy of something at a given price reveals exactly how much you like it – not just whether you prefer it to something else
- In Arrow's world, we can't use prices – or anything else – to make this type of judgment
- So we can't aggregate preferences in a well-behaved, coherent way
- One other thing to notice: we've completely ignored the problem of *figuring out* peoples' preferences
  - Depending on the social choice function, people might have an incentive to lie about their preferences
  - The rest of the semester will, in some sense, be focused on that part of the problem

References

- Original result: Kenneth Arrow (1951), *Social Choice and Individual Values*, New York: Wiley
- Our proof: John Geanakoplos (2005), "Three Brief Proofs of Arrow's Impossibility Theorem," *Economic Theory* 26(1)

Lecture 3

Common Knowledge

People with the same prior can't "agree to disagree,"
even on the basis of different information.

1 Context

- So far, we've ignored problems of *private information*
  - In Walrasian equilibrium, it doesn't matter whether other people know your preferences
    Prices come from wherever they come from,
    and all anyone can do is optimize given those prices
  - In Arrow's policy setting, we didn't worry about *learning* each person's preferences
    We assumed they were known,
    and focused on getting from individual preferences to social preferences.
- For the rest of the semester, we'll be looking at settings where people have *private information*
  - Thus, part of the challenge will be understanding their incentives to reveal that information
- First, though, we need a formal framework to think about private information.

- There are lots of types of private information
  - I could have private information about something that's only relevant to me – like private information about my own preferences.

    (This still matters strategically to others – it might tell a seller how much money he could try to get out of me, or a competing buyer how much I might bid in an auction – but it only affects others through my actions, not directly.)
  - I could also have private information that's directly relevant to other peoples' payoffs

    I might have inside information about a company – so if I'm trying to sell shares, buyers need to be worried about what I know
- For a lot of applications, we work with a model of private information that's suited to that particular environment.
- But for today, we're going to introduce a very general model of information – general enough to (more or less) nest all other models.
- For the next two lectures, we'll be thinking about interactions between two people – but everything here could be extended to more.

2 General Model of Information

2.1 Probability Spaces

- We begin with a **finite set Ω of states of the world ω**
- A "state of the world" ω is a complete description of the world – basically, a resolution of all uncertainty.
- So, what does it mean to know something? Basically, it means you can distinguish between two states of the world.
- What I know can be described as a **partition of** Ω – which states of the world I can tell apart, and which ones I can't.
- For example, let $\Omega = \{1, 2, 3\}$

  We define my **information partition** as a set of disjoint subsets that make up Ω, such as $\mathcal{P}_1 = \{\{1, 2\}, \{3\}\}$

  That means that if the actual state of the world is 3, I know it's 3; but if the actual state is either 1 or 2, I know that it's *either* 1 or 2, but I don't know which.

- At first glance, this sounds like it doesn't capture uncertainty well, but it turns out, it does

  We just need to put prior probabilities on each state, and define the states right

  Suppose we've caught a murder suspect and want to know whether he actually committed the murder, so we check to see whether his fingerprints match the prints at the scene

  There are four states of the world:

  ω_1: he's guilty, and and there are fingerprints

  ω_2: he's guilty, but there aren't any fingerprints

  ω_3: he's innocent, and there are fingerprints

  ω_4: he's innocent, and there aren't any fingerprints

  The "event" we care about is whether he's guilty – which means what we care about is the probability that the state of the world is *either* 1 or 2

  Our information partition, though, is $\{\{1,3\},\{2,4\}\}$ – we can tell whether there are fingerprints, but not whether he's guilty

  There's one other thing we need to make sense of this – a **prior probability distribution** on the different states of the world.

  So suppose that...

  - With probability 0.4, he's guilty and left fingerprints
  - With probability 0.1, he's guilty but there are no fingerprints
  - With probability 0.05, he's innocent but there are fingerprints anyway
  - With probability 0.45, he's innocent and there are no fingerprints

- Formally, this defines a **probability space** – a set of states Ω, a set of events that we might care about $\mathcal{B}$, and a prior probability over the states p. (Each event is just the set of states in which that event has happened – the set of states in which the guy is guilty, for example.)

2.2 Bayes' Rule and Posterior Probabilities

- So now suppose we found fingerprints, and we want to know how likely it is that the guy is guilty
- We use **Bayes' Law**
- Iterated expectations say that if A and B are two different events,

$$\Pr(A \ \textit{and} \ B) \quad = \quad \Pr(A)\Pr(B|A)$$

  and if we move this around, it gives Bayes' Law, which is

$$\Pr(B|A) = \frac{\Pr(A \ \textit{and} \ B)}{\Pr(A)}$$

- So now take a probability space $(\Omega, \mathcal{B}, p)$; an information partition $\mathcal{P}_1$; an element $P_1 \in \mathcal{P}_1$; and an event E; Bayes' Law says that

$$\Pr(E|P_1) = \frac{\Pr(E \cap P_1)}{\Pr(P_1)}$$

- Or in our context,

$$\Pr(guilty|fingerprints) = \frac{\Pr(guilty \ \textit{and} \ fingerprints)}{\Pr(fingerprints)}$$

- So if I know the prior probability of each state, I can calculate the posterior probability of an event, given what I've learned; in this case,

$$\Pr(guilty|fingerprints) = \frac{\Pr(\omega_1)}{\Pr(\omega_1 \cup \omega_3)} = \frac{.4}{.4 + .05} \approx 0.889$$

2.3 Who knows who knows who knows what

- Next, we want to think about environments with multiple people, and think not just about who knows what, but what people think about what each other know
- Suppose $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$, and there are two of us
- My partition is $\mathcal{P}_1 = \{\{1, 2, 3\}, \{4, 5\}, \{6, 7, 8\}\}$
- Your partition is $\mathcal{P}_2 = \{\{1, 2\}, \{3, 4\}, \{5\}, \{6, 7\}, \{8\}\}$
- Suppose the state of the world is 5. Let's think about who knows what.
- I know the state is either 4 or 5, but I don't know which.
- You know the state is 5.

  But obviously, I don't know that.

  If we think about what I know about what you know: I know that *either* the state is 5, in which case you know it's 5; or the state is 4, and you therefore know that the state is either 3 or 4.

  So if we think about *my beliefs about your beliefs* – I know that you know the state is either 3, 4, or 5.
- What about what I know about what you know *about what I know*?

  I know that you might think the state is either 3, 4, or 5.

  If you know the state is 5, then you know that I know it's either 4 or 5.

  But if the state is 4, then you know it's either 3 or 4; which means that you know that *either* I know it's either 4 or 5, *or* you know that I know it's either 1, 2, or 3.

  So if we think about what I know that you know that I know: all I know that you know that I know is that the state is either 1, 2, 3, 4, or 5.
- What about you? We already said, you know $\omega = 5$.
- So you know that I know it's either 4 or 5.
- So you know I know you know it's either 3, 4, or 5.
- So you know I know you know I know it's either 1, 2, 3, 4, or 5.

2.4 Common Knowledge

- Something is **common knowledge** if we both know it's true;

  and I know that you know it's true;

  and you know that I know it's true;

  and I know that you know that I know that you know that I know that you know it's true;

  and so on, for any string of beliefs we put together.

- So something being common knowledge is a pretty high bar – it's a lot more than just both of us knowing something is true.

- In the current example, when the state is 5, you know it's 5, and I know it's either 4 or 5

  So we both know the state is either 4 or 5

  But it's not common knowledge that the state is either 4 or 5 – because as far as I know, you might think the state is 3

  So if a particular event E occurs only in states 4 and 5, and the state of the world is 5, then the event has occurred, *and we both know the event has occurred*, but it is *not common knowledge* that the event has occurred

  Because I don't know whether you know that I know it's occurred!

- And the big result we'll show today is that common knowledge is very restrictive

- But first, a couple more tools

2.5 Refinements and Coarsenings of Information Partitions

- Suppose there are eight states of the world: $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$
- And suppose my information partition is $\mathcal{P}_1 = \{\{1, 2, 3\}, \{4, 5\}, \{6, 7, 8\}\}$
- We can make my information better by allowing me to distinguish between more states

  This is the same as breaking up some of the elements of my existing partition, leading to

  $$\mathcal{P}_1' = \{\{1, 2\}, \{3\}, \{4, 5\}, \{6, 7, 8\}\}$$

  This partition is *finer* than my old one – each element of $\mathcal{P}_1'$ is a subset of a single element of $\mathcal{P}_1$, so I still know everything I used to know, and then more

  So $\mathcal{P}_1'$ is a *refinement* of $\mathcal{P}_1$.
- We can go the other direction – make my information worse, by lumping together two or more elements of my partition

  The partition

  $$\mathcal{P}_1'' = \{\{1, 2, 3\}, \{4, 5, 6, 7, 8\}\}$$

  is a *coarsening* of $\mathcal{P}_1$

2.6 Meets and Joins

- Suppose my information partition continues to be $\mathcal{P}_1 = \{\{1,2,3\},\{4,5\},\{6,7,8\}\}$
- And suppose you also have an information partition, $\mathcal{P}_2 = \{\{1,2\},\{3,4\},\{5\},\{6,7\},\{8\}\}$
- Note that $\mathcal{P}_1$ is neither coarser or finer than $\mathcal{P}_2$ – they can't be compared in this way

  Even though your partition has more pieces, there are some things I know that you don't know – we can't unambiguously say your information is better than mine

  For example, consider the event $E = \{4,5,6\}$ – an event that occurs in states 4, 5, and 6, but not otherwise. When $\omega = 4$, I know that the event has occurred, but you don't – since for all you know, the state might be 3.

- Now, there are two binary operations we want to define on information partitions, sort of analogous to taking unions and intersections of sets
- Given two information partitions $\mathcal{P}_1$ and $\mathcal{P}_2$, we define their *join* $\mathcal{P}_1 \vee \mathcal{P}_2$ as the *coarsest common refinment* of the two partitions
  - basically, the set of intersections of one of your partition elements with one of my partition elements
  - in this example, $\mathcal{P}_1 \vee \mathcal{P}_2 = \{\{1,2\},\{3\},\{4\},\{5\},\{6,7\},\{8\}\}$
  - this represents what we would know if we shared our information.
- We also define the *meet* $\mathcal{P}_1 \wedge \mathcal{P}_2$ as the *finest common coarsening*
  - this is the finest partition such that each of my elements is contained in a single element of the meet, and each of your elements is contained in a single element of the meet
  - in this example, 1, 2 and 3 have to be in the same element of the meet, because they're in the same element of $\mathcal{P}_1$;

    3 and 4 need to be together, because they're in the same element of $\mathcal{P}_2$;

    and 4 and 5 need to be together, because they're together in $\mathcal{P}_1$

    It turns out, the meet is $\mathcal{P}_1 \wedge \mathcal{P}_2 = \{\{1,2,3,4,5\},\{6,7,8\}\}$
- It's not yet clear what this represents intuitively – but it will be soon

3 Today's Big Result

- Today's big result is from Aumann, "Agreeing to Disagree," and it says the following:

  Theorem (Aumann)**.** *If two people have the same priors, and their posteriors for an event E are common knowledge, then these posteriors are equal.*

- In other words, we can't agree to disagree
  - if it's *common knowledge* that you think the probability of E given your information is x,

    and it's *common knowledge* that I think the probability of E given my information is y,

    then $x = y$

- To prove this, we first have to figure out what it means for our posteriors to be common knowledge
- If it's common knowledge that you put probability x on event E, then this means...
  - Obviously, you have to believe $\Pr(E) = x$ – which means that $\Pr(E|P_2) = x$ at the element of $\mathcal{P}_2$ that contains ω
  - But also, I have to know that you believe $\Pr(E) = x$ – which means $\Pr(E|P_2) = x$ at every element of $\mathcal{P}_2$ that *I think you might be at*

    Which is every element of $\mathcal{P}_2$ that intersects the element of $\mathcal{P}_1$ that contains ω
  - And also, *you need to know* that I know that you believe $\Pr(E) = x$

    Which means $\Pr(E|P_2) = x$ at every element of $\mathcal{P}_2$ that *you might think* I might think you are at – which is every element of $\mathcal{P}_2$ that intersects an element of $\mathcal{P}_1$ that intersects the element of $\mathcal{P}_2$ that contains ω
  - And *I need to know* you know I know you believe $\Pr(E) = x$

    Which means $\Pr(E|P_2) = x$ at every element of $\mathcal{P}_2$ that *I might think* you might think I might think you are at

    Which is every element of $\mathcal{P}_2$ that intersects an element of $\mathcal{P}_1$ that intersects the element of $\mathcal{P}_2$ that intersects the element of $\mathcal{P}_1$ that contains ω
  - And so on
  - Remember our old example, with $\mathcal{P}_1 = \{\{1,2,3\},\{4,5\},\{6,7,8\}\}$ and

    $\mathcal{P}_2 = \{\{1,2\},\{3,4\},\{5\},\{6,7\},\{8\}\}$

    If $\omega = 5$, I don't know whether the state is 4 or 5, so I don't know whether you're at information partition element $\{3,4\}$ or $\{5\}$

Which means I don't know whether you think I'm at $\{1, 2, 3\}$ or $\{4, 5\}$

Which means I don't know whether you think I think you're at $\{1, 2\}$, $\{3, 4\}$, or $\{5\}$

So for your posterior belief to be common knowledge,

it has to be the same at each of those three information sets

- In general, for it to be common knowledge that your posterior belief is x,
  we need $\Pr(E|P_2) = x$ at every P_2 in the element of $\mathcal{P}_1 \wedge \mathcal{P}_2$ that contains ω
- And likewise, for it to be common knowledge that my posterior is y,
  then $\Pr(E|P_1) = y$ at every P_1 in the element of $\mathcal{P}_1 \wedge \mathcal{P}_2$ that contains ω

- Now we're ready to prove Aumann's result.

Proof of Aumann's Theorem

- fix a state of the world ω and an event E
- let P be the element of $\mathcal{P}_1 \wedge \mathcal{P}_2$ that contains ω – so in our recent example, P would be $\{1, 2, 3, 4, 5\}$
- For it to be *common knowledge* that my posterior is at q_1, my posterior must be q_1 at every element of $\mathcal{P}_1$ that's inside P.
- Now since $\mathcal{P}_1 \wedge \mathcal{P}_2$ is a coarsening of $\mathcal{P}_1$, write P as the union of some elements of $\mathcal{P}_1$,

$$P = \bigcup_i P^i$$

We just said my posterior probability must be q_1 at each of these elements P^i, or

$$\Pr(E|P^i) = \frac{\Pr(E \cap P^i)}{\Pr(P^i)} = q_1$$

or

$$\Pr(E \cap P^i) = q_1 \Pr(P^i)$$

But now since the different elements P^i of my partition are disjoint, and their union is P, we can sum over i and get

$$\begin{array}{rcl} \sum_i \Pr(E \cap P^i) & = & \sum_i q_1 \Pr(P^i) \\ & \downarrow & \\ \Pr(E \cap P) & = & q_1 \Pr(P) \\ & \downarrow & \\ \frac{\Pr(E \cap P)}{\Pr(P)} & = & q_1 \end{array}$$

- But if we let q_2 be your prior belief about the probability of E, and assume that *that* is common knowledge, then we can do the same steps, and show that

$$\frac{\Pr(E \cap P)}{\Pr(P)} = q_2$$

- So if our posterior probabilities q_1 and q_2 on E are both common knowledge, then $q_1 = q_2$, assuming we put the same prior probabilities p on each state of the world – even if my posterior is based on different information than yours!

Three additional things to note.

Just knowing each others' posteriors is not enough.

- From Aumann. Consider a set of states $\Omega = \{1, 2, 3, 4\}$, with equal prior on each state. My information partition is $\mathcal{P}_1 = \{\{1, 2\}, \{3, 4\}\}$, and yours is $\mathcal{P}_2 = \{\{1, 2, 3\}, \{4\}\}$. Consider the event $E = \{1, 4\}$, and suppose the state is 1.

- Then I know the state is either 1 or 2, with equal probability, and that the event $\{1, 4\}$ therefore has probability $\frac{1}{2}$.

- And you know that that's my posterior – because that would be my posterior in any state!

- You know the state is either 1, 2, or 3, so you put posterior $\frac{1}{3}$ on the event A.

- But since I know the state is either 1 or 2, I know that you know it's either 1, 2, or 3 – so I know your posterior too.

- So I know your posterior and you know mine, but they're not equal – and they don't have to be, because they're not common knowledge.

  Since you think the state might be 3, you think that I might think it's either 3 or 4 – which means you think that I might think your posterior is $\frac{1}{3}$, or I might think it's 1.

How our posteriors might become common knowledge

- Aumann says that we have to agree if our posteriors are common knowledge, but doesn't explain how that would come to happen
- A different paper – Geanakoplos and Polemarchakis, "We Can't Disagree Forever" – offers one way it could happen
- Suppose I learn whatever I learn and calculate my posterior probability, you learn whatever you learn and calculate yours – and then we do the following.
  - I announce my posterior probability.
  - Based on that new information, you revise your posterior probability, and then announce your new one.
  - Based on that new information, I revise my posterior, and announce my new one.
  - And so on.
- As long as Ω is finite, they show that in finite time, our posteriors will become common knowledge – at which point they must be equal.
- They also give a cute example of how, even if our "communication" just consists of saying the same thing over and over for a while, we still eventually converge.
- An example from their paper.

  $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $p(\omega) = \frac{1}{9}$ for each $\omega \in \Omega$

  $\mathcal{P}_1 = \{\{1, 2, 3\}, \{4, 5, 6\}, \{7, 8, 9\}\}$

  $\mathcal{P}_2 = \{\{1, 2, 3, 4\}, \{5, 6, 7, 8\}, \{9\}\}$

  Consider the event $E = \{3, 4\}$, and suppose $\omega = 1$
- Work it out – communication goes $p = \frac{1}{3}$, $q = \frac{1}{4}$, $p = \frac{1}{3}$, $q = \frac{1}{3}$.
- If instead $E = \{1, 5, 9\}$ and $\omega = 1$, it would go $p = \frac{1}{3}$, $q = \frac{1}{4}$, $p = \frac{1}{3}$, $q = \frac{1}{4}$, $p = \frac{1}{3}$, $q = \frac{1}{3}$.

Common knowledge of posteriors does not imply we know everything.

- Also from Geanakoplos and Polemarchakis
- Since we communicate until our beliefs converge, one might suspect that I learn "all there is to know" from your information, and vice versa
- But this turns out not to be the case.
- Suppose we each get to flip a coin, and the event we're interested in is that the two coins matched – either both heads or both tails.
- Suppose we each flip our coin, and they both come up heads.
- I believe the probability the two coins match is $\frac{1}{2}$.

  And you know that. And I know you know it.

  It's common knowledge my posterior is $\frac{1}{2}$, because my posterior would be $\frac{1}{2}$ at either information set.

  Same for you – your posterior is $\frac{1}{2}$, and that's common knowledge.
- And of course, since both posteriors are common knowledge, they have to match (Aumann).

  And since they're already common knowledge, when I announce that my posterior is $\frac{1}{2}$, it doesn't change your beliefs, and vice versa – our beliefs have already converged.
- But if we shared information, we would know that the two coins matched with probability 1, not probability $\frac{1}{2}$.

References

- Robert Aumann (1976), "Agreeing to Disagree," *Annals of Statistics*
- Geanakoplos and Polemarchakis (1982), "We Can't Disagree Forever," *Journal of Economic Theory*

Lecture 4

Bayesian Games and No Trade

The receipt of private information cannot create any incentives to trade.

Today is really about understanding games of incomplete information and Bayesian Nash equilibrium; the no-trade result is sort of a corollary.

But first, a little tangent on Bayes' Law.

1 Bayes' Law and Odds Ratios

- Last week, we saw Bayes' Law: $\Pr(E|P_1) = \frac{\Pr(E \cap P_1)}{\Pr(P_1)}$
- We can think of this as

$$\Pr(E|data) \quad = \quad \frac{\Pr(E)\Pr(data|E)}{\Pr(data)} \quad = \quad \frac{\Pr(E)\Pr(data|E)}{\Pr(E)\Pr(data|E)+\Pr(\neg E)\Pr(data|\neg E)}$$

- Given a prior and data, this will always allow us to compute a posterior probability
- But if we're in a setting where we expect to get multiple "new" pieces of information, it's nice to have an easier way to update beliefs
- For example, suppose there are two possible states of the world, "GOOD" and "BAD", and we're going to get repeated noisy signals about the true state
- In each period, we get a signal that matches the true state with probability $\frac{2}{3}$
- That is, if the state is GOOD, we get a good signal with probability $\frac{2}{3}$ and a bad signal with probability $\frac{1}{3}$; and if the state is BAD, that's reversed
- Suppose our initial prior is $\Pr(G) = \frac{3}{4}$, and we get the string of signals good, bad, good, bad, bad, or *gbgbb*

- What is our posterior?
- Well, by Bayes' Law,

$$\Pr(G|gbgbb) = \frac{\Pr(G)\Pr(gbgbb|G)}{\Pr(gbgbb)} = \frac{\Pr(G)\Pr(gbgbb|G)}{\Pr(G)\Pr(gbgbb|G)+\Pr(B)\Pr(gbgbb|B)}$$

  Instead of using this, though, let's look at the *odds ratio* – the ratio of p to $1-p$ – which is

$$\frac{\Pr(G|gbgbb)}{\Pr(B|gbgbb)} = \frac{\frac{\Pr(G)\Pr(gbgbb|G)}{\Pr(G)\Pr(gbgbb|G)+\Pr(B)\Pr(gbgbb|B)}}{\frac{\Pr(B)\Pr(gbgbb|B)}{\Pr(G)\Pr(gbgbb|G)+\Pr(B)\Pr(gbgbb|B)}} = \frac{\Pr(G)\Pr(gbgbb|G)}{\Pr(B)\Pr(gbgbb|B)}$$

  We can even decompose this further – since the signals are independent in each period, the probability of $gbggb$ given a particular state is just the probability of g in that state, times the probability of b in that state, times, etc. So we can write this as

$$\frac{\Pr(G|gbgbb)}{\Pr(B|gbgbb)} = \frac{\Pr(G)\Pr(g|G)\Pr(b|G)\Pr(g|G)\Pr(b|G)\Pr(b|G)}{\Pr(B)\Pr(g|B)\Pr(b|B)\Pr(g|B)\Pr(b|B)\Pr(b|B)}$$

$$= \frac{\Pr(G)}{\Pr(B)}\cdot\frac{\Pr(g|G)}{\Pr(g|B)}\cdot\frac{\Pr(b|G)}{\Pr(b|B)}\cdot\frac{\Pr(g|G)}{\Pr(g|B)}\cdot\frac{\Pr(b|G)}{\Pr(b|B)}\cdot\frac{\Pr(b|G)}{\Pr(b|B)}$$

- So in the example I mentioned above, we would start with the prior odds ratio $\frac{3/4}{1/4} = 3$, and then multiply it by $\frac{2/3}{1/3} = 2$ every time we get a good signal, and multiply it by $\frac{1/3}{2/3} = \frac{1}{2}$ every time we get a bad signal, to get an updated odds ratio
- The point: if we work in odds ratios, rather than probabilities, then we can just start with the prior odds ratio, and each time we get a new signal, multiply the odds ratio by the relative likelihood of the new data under the two states, to get the posterior odds ratio; and we can keep doing this over and over with each new signal that we receive.

- OK, enough tangent, back to work
- Today's real task is to update our notion of Nash equilibrium to include private information

2 Review – (Static) Nash Equilibrium

- The standard setup for a static game of complete information requires three elements:
  - A set of players $N = \{1, 2, \ldots, n\}$
  - A set of available actions A_i for each player
  - A set of payoff functions
    $$U_i : A_1 \times A_2 \times \cdots \times A_n \to \Re$$
    giving the payoff to each player given each profile of strategies
- A (pure strategy) Nash equilibrium is an action profile $a = (a_1, a_2, \ldots, a_n)$ such that each player is best-responding to the remaining players' actions – that is, such that for all i,
  $$a_i \in \arg\max_{a \in A_i} U_i(a, a_{-i})$$
  where a_{-i} is a vector of the other $n - 1$ players' actions.
- A key maintained assumption is that the entire environment is *common knowledge*
  - everyone knows the set of players, actions and payoffs,
    everyone knows everyone knows it,
    and so on.
- So what happens if we *don't* have common knowledge about the entire environment?
- For example, suppose you know your own payoff function, but you're not sure of mine.
- That's a game of *incomplete information*, and that's what the rest of the semester will be about.

3 Games of Incomplete Information

- The way we deal with incomplete information is to assume that there are different possible *types* of each player
- For example, suppose the incomplete information is that you don't know how willing I am to fight
- We assume this is due to you not knowing which type of player I am – a tough player, or a weak player
- I know which one I am, but you don't – so you just assign some probability to me being strong (having one payoff function), and some probability to me being weak (having a different payoff function)
- Formally, we assume each player's type is his own private information – I know which type I am, but not which types my opponents turned out to be
- So a static game of incomplete information consists of...
  1. A set of players $N = \{1, 2, \ldots, n\}$, same as before
  2. A set of actions for each player A_i, same as before
  3. A set of possible types T_i for each player
  4. A probability distribution p over the set of type profiles $T_1 \times T_2 \times \cdots \times T_n$
  5. A set of payoff functions, that may depend on every player's type:

$$U_i : A_1 \times \cdots \times A_n \times T_1 \times \cdots \times T_n \to \Re$$

- And *all of that* is common knowledge – if there are lots of players in the game, everyone agrees on the probability that I'm tough, and I know what they think the probability is, and so on – I just also know what my type "turned out" to be
- Several things to note about this setup:
  - It doesn't matter whether we allow the set of available actions A_i to depend on player i's type; if we wanted to "disallow" a particular action a_i for a particular type t_i of player i, we could keep a_i in A_i and just set $U_i(a_i, \cdot, t_i, \cdot) = -\infty$. So for simplicity, we assume A_i is fixed across types.
  - Information partitions are defined by the type space – player i is assumed to know the exact value of T_i, but know nothing about T_{-i} beyond what he infers from T_i. We very often, for simplicity, assume that different players' types are independent, but we don't have to.

– In many models, we assume that player i's payoff does not depend on other players' types, only his own (and the action profile). That is, we often (but not always) assume that I care about other players' types *only because it may influence their actions*, not because it directly influences my payoffs. But there are exceptions – in an adverse selection model, for example.

– Like before, we assume that the entire setup above is common knowledge – everyone agrees on the basic universe we live in, but only player i knows which type he turned out to be.

– It's customary to think of the game happening in two stages: in the first stage, "nature moves" by randomly assigning a type to each player; in the second stage, players play the game given their realized types

4 Bayesian Nash Equilibrium

The solution concept for a static game of incomplete information is Bayesian Nash equilibrium, which is really just a generalization of Nash equilibrium to accommodate types. Specifically...

- A *strategy* for player i is now a type-dependent choice of action, that is, a mapping

$$s_i : T_i \to A_i$$

  specifying an action $s_i(t_i) \in A_i$ that I plan to take for each type I might turn out to be

- A Bayesian Nash equilibrium is a profile of strategies $(s_1, s_2, \ldots, s_n)$ such that for every player i and type t_i,

$$s_i(t_i) \in \arg\max_{s \in A_i} E_{T_{-i}|t_i} U_i(s, s_{-i}(t_{-i}), t_i, t_{-i})$$

  That is, each type of each player is maximizing his expected payoff, given his correct beliefs about the probabilities of different opponent types (given his own) and given his correct beliefs about his opponents' strategies

5 An Example of BNE: First Price Auctions

- Let's do a simple example: a private-values, sealed-bid first-price auction

  There is a single object for sale

  Each player's type is the value he would get from winning it; everyone simultaneously submits a bid in writing, and the player with the highest bid pays his bid and receives the object

  Let's suppose types are independent and uniform over the interval $[0, 1]$

  So formally:

  - The set of players is $1, 2, \ldots, n$
  - The set of available actions is $A_i = \Re^+$ – any positive bid is allowed
  - The set of possible types is $T_i = [0, 1]$, and the probability distribution over type profiles is uniform over $[0, 1]^n$ (meaning it has density 1 everywhere)
  - The payoff from winning the auction is your type minus your bid, and the payoff from losing is 0, so if we ignore ties, we can write

$$U_i(a, t) = \begin{cases} t_i - a_i & if \quad a_i > \max_{j \neq i} a_j \\ 0 & if \quad a_i < \max_{j \neq i} a_j \end{cases}$$

  - To complete the model, let's assume they're broken randomly; so if $a_i = \max_{j \neq i} a_j$, the payoff to player i is

$$(t_i - a_i) \frac{1}{1 + \#\{j \neq i : a_j = a_i\}}$$

- Note that we've assumed *private values* – other players' types don't enter directly into my payoff function, they affect me only through their effect on my opponents' bids.

- What does the Bayesian Nash equilibrium of this game look like? It turns out, it's an equilibrium for everyone to bid $\frac{n-1}{n}$ times their type.

- Why?

- To show this, we need to show that if all my opponents are playing this strategy, then this strategy maximizes my expected payoff.

- So suppose my type is t_i, my opponents are all bidding $\frac{n-1}{n}$ times their values, and I bid b.

- First of all, note that if I bid more than $\frac{n-1}{n}$, I'll win with probability 1 – I outbid any opponent with type $t_j < 1$, and tie opponents with type $t_j = 1$, which occurs with prob 0

  So bidding more than $\frac{n-1}{n}$ is a bad idea – it drives up the price I pay, without making me more likely to win. So the only strategies that might be best-responses are in the range $[0, \frac{n-1}{n}]$.

- Now, for b within that range, my expected payoff is

$$\begin{aligned} E_{T_{-i}}U_i &= (t_i - b) \cdot \Pr(win|b) + 0 \cdot \Pr(lose|b) \\ &= (t_i - b) \Pr(\max_{j \neq i} s_j(t_j) < b) \\ &= (t_i - b) \Pr(\max_{j \neq i} \tfrac{n-1}{n} t_j < b) \\ &= (t_i - b) \Pr(\max_{j \neq i} t_j < \tfrac{n}{n-1} b) \\ &= (t_i - b)(\tfrac{n}{n-1} b)^{n-1} \\ &= (\tfrac{n}{n-1})^{n-1}(t_i - b)b^{n-1} \end{aligned}$$

- So now let's maximize this thing: $(t_i - b)b^{n-1} = t_i b^{n-1} - b^n$ has derivative

$$(n-1)t_i b^{n-2} - nb^{n-1} \quad = \quad nb^{n-2}\left[\frac{n-1}{n} t_i - b\right]$$

  This is increasing on $b < \frac{n-1}{n} t_i$ and decreasing on $b > \frac{n-1}{n} t_i$ – so it's maximized at $b = \frac{n-1}{n} t_i$.

- So if everyone else bids $\frac{n-1}{n}$ times their type, my best-response is to bid $\frac{n-1}{n}$ times my type; so everyone bidding $\frac{n-1}{n}$ times their type is an equilibrium.

6 An Application of BNE: A No-Trade Theorem

- Last week, we saw Aumann's result that if both our posterior beliefs are common knowledge, we can't disagree
- A nice analogy is to a trading problem: basically, if the entire environment is common knowledge, then differential private information on its own can't lead to trade
- We'll just prove a "simple" version, then describe how the result extends
- For the simple version...
  - there are just two traders, me and my bookie.
  - there's one event we care about – whether or not my team wins this Sunday
  - my bookie and I are both strictly risk-averse – but we still might want to bet because we have different information about my team's chances
  - my bookie and I share a common prior, but have different information partitions – he might have some information I don't have, and I might have some information he doesn't have
  - I've got wealth w_1, he's got wealth w_2, my utility for wealth is u_1 and his is u_2, and both strictly concave
- So here's the result:

Theorem. *In equilibrium, we never trade.*

- The basic logic is that, if I decide to place a bet, I need to condition not only on my own information, *but also on the fact that my bookie wants to bet with me*; and on the fact that he wants to trade with me *even knowing I want to trade with him*; and so on
- And similarly, he needs to condition on the fact that I want to bet with him
- In essence, this means it has to be common knowledge that I believe the bet is skewed in my favor, and also common knowledge that he believes the bet is skewed in his favor – which an analogy to Aumann suggests is impossible
- Which means we can't both be willing to bet if it's common knowledge we're both rational

Let's prove it formally

- As always, we have a set of states Ω,

  my information partition $\mathcal{P}_1 = \{t_1^1, t_1^2, \ldots, t_1^m\}$,

  and my bookie's information partition $\mathcal{P}_2 = \{t_2^1, \ldots, t_2^{m'}\}$

- Let's suppose there is a finite number of possible bets we might consider making

  (This is WLOG, if we limit ourselves to bets with non-fractional number of cents, in amounts less than the total wealth of the entire earth)

- And suppose that if we do make a bet, that occurs within the Bayesian Nash equilibrium of some sort of negotiation process

  We don't have to specify what that process is; we just need the environment we're in (including the prior) to be common knowledge, and that in order for a bet to happen, we both have to agree to it.

- What I want to show is that with probability 1, we don't bet

- So suppose that weren't true – that with positive probability, we made some nonzero bet

  Since there are a finite number of bets, that means there is some bet (x, y) we make with positive probability

  (Let x be the amount I receive if my team wins, and y the amount I receive if my team loses – so if I'm betting on my team, $x > 0 > y$)

- Let $q(t_1, t_2)$ be the probability that that bet occurs,

  given my information $t_1 \in \mathcal{P}_1$ and my bookie's $t_2 \in \mathcal{P}_2$

- For this to happen in equilibrium, two things need to be true:

  - For every information set t_1 at which I agree to this bet with positive probability, I must believe that in expectation over my bookie's possible information sets, and conditional on him also being willing to accept this bet, the bet does not decrease my expected payoff
  - For every information set t_2 at which my bookie agrees to this bet with positive probability, he must believe that in expectation over my possible information sets, and conditional on me also being willing to accept this bet, the bet does not decrease his expected payoff

  We'll show that these can't both hold unless the trade is $(0, 0)$.

- Suppose I'm at information set t_1, and I'm willing to agree to the bet (x, y)

- Conditional on you agreeing to the bet, I put some posterior probability $p(t_2^j|t_1, bet)$ on you being at each information set t_2^j
- Which means that I evaluate my expected payoff, *if we both agree to the bet*, as...

$$\sum_{t_2^j \in \mathcal{P}_2} p(t_2^j|t_1, bet) \left(p(W|t_1, t_2^j) u_1(w_1 + x) + (1 - p(W|t_1, t_2^j)) u_1(w_1 + y) \right)$$

$$= \left(\sum_{t_2^j \in \mathcal{P}_2} p(t_2^j|t_1, bet) p(W|t_1, t_2^j) \right) u_1(w_1+x) + \left(\sum_{t_2^j \in \mathcal{P}_2} p(t_2^j|t_1, bet)(1 - p(W|t_1, t_2^j)) \right) u_1(w_1+y)$$

Now, in order for me to be willing to make the bet, this has to be at least as good as $u_1(w_1)$, my utility from not betting at all:

$$\left(\sum_{t_2^j \in \mathcal{P}_2} p(t_2^j|t_1, bet) p(W|t_1, t_2^j) \right) u_1(w_1+x) + \left(\sum_{t_2^j \in \mathcal{P}_2} p(t_2^j|t_1, bet)(1 - p(W|t_1, t_2^j)) \right) u_1(w_1+y) \geq u_1(w_1)$$

Now let's multiply both sides by the probability that I have type t_1 and we both agree to (x, y), so this becomes

$$\left(\sum_{t_2^j \in \mathcal{P}_2} p(t_1, bet) p(t_2^j|t_1, bet) p(W|t_1, t_2^j) \right) u_1(w_1 + x)$$

$$+ \left(\sum_{t_2^j \in \mathcal{P}_2} p(t_1, bet) p(t_2^j|t_1, bet)(1 - p(W|t_1, t_2^j)) \right) u_1(w_1 + y) \geq p(t_1, bet) u_1(w_1)$$

Now let's sum over all the different values of t_1 that I might have, giving

$$\left(\sum_{t_1^i \in \mathcal{P}_1} \sum_{t_2^j \in \mathcal{P}_2} p(t_1^i, bet) p(t_2^j|t_1^i, bet) p(W|t_1^i, t_2^j) \right) u_1(w_1 + x)$$

$$+ \left(\sum_{t_1^i \in \mathcal{P}_1} \sum_{t_2^j \in \mathcal{P}_2} p(t_1^i, bet) p(t_2^j|t_1, bet)(1 - p(W|t_1^i, t_2^j)) \right) u_1(w_1 + y) \geq \sum_{t_1^i} p(t_1^i, bet) u_1(w_1)$$

Now, iterated expectations says that $\Pr(A \ \ and \ \ B) = \Pr(A) \Pr(B|A)$. Applying this a couple of times, we simplify the last expression to

$$\Pr(bet) \Pr(W|bet) u_1(w_1 + x) + \Pr(bet) (1 - \Pr(W|bet)) u_1(w_1 + y) \geq \Pr(bet) u_1(w_1)$$

or, since $\Pr(bet) > 0$,

$$\Pr(W|bet) u_1(w_1 + x) + (1 - \Pr(W|bet)) u_1(w_1 + y) \geq u_1(w_1)$$

Let $p^* = \Pr(W|bet)$. Since I'm risk-averse, I'd strictly prefer a sure thing with the same expected value as the bet: that is,

$$u_1(p^*(w_1 + x) + (1 - p^*)(w_1 + y)) > p^* u_1(w_1 + x) + (1 - p^*)u_1(w_1 + y) \geq u_1(w_1)$$

meaning $p^*(w_1 + x) + (1 - p^*)(w_1 + y) > w_1$, or $p^* x + (1 - p^*)y > 0$. So if I only bet when I want to bet, it has to be the case that, averaged over all the different cases where we bet, the bet has strictly positive expected value for me.

But if we did the same analysis from my bookie's point of view, we would have to conclude that, averaged over all the different cases where we bet, the bet has strictly positive expected value for him.

And those can't both be true.

7 Milgrom and Stokey

Milgrom and Stokey, "Information, Trade, and Common Knowledge," prove this result in a much more general setting.

- They allow $N > 2$ traders
- They allow any finite number of states of the world, which they break up into two components – a *payoff-relevant* state, and a part of the state that just conveys information about the payoff-relevant part
- They allow l different goods, which are each consumed in continuous quantities
- They allow *endowments* of the different goods to vary across payoff-states – so each trader has an endowment $e_i \in (\Re^+)^l$ in each payoff-state, and consumes a vector $x_i \in (\Re^+)^l$ in each payoff state
- They allow each trader's *utility function* to vary across payoff-states – all they require is each state's utility function is strictly concave
- They don't even require that we all have a common prior over the payoff-states – they only require that we have the same beliefs about how the information-states depend on the payoff-states
- And they allow us to make any state-contingent trades we want.

So for example, there could be one (payoff-relevant) state of the world where you lose your job. So in that state, you have less money. So you'd probably be willing to make a trade where you get more money in that state, and give up money in states where you still have your job. You and I can disagree on how likely you are to get fired – I might think it's 10% likely, you think it's 5% likely – but we both agree that *if* you're going to get fired, there's a 50% chance you'll be told you're on probation at your performance review.

Note that this model gives three different reasons we may want to trade:

- To reallocate goods within a state. In some state of the world, I'm endowed with a lot of one good, and want to consume other goods, so I want to trade goods within that state.
- To reallocate goods across states. I may want to trade stuff into one state (where I'm endowed with less, or have a steeper utility functino), out of another state (where I'm endowed with more). Note that my utility function within each state has to be strictly concave, but there's no restriction on how my utility function varies across states.

- To arbitrage different beliefs. If we have different beliefs about the likelihood of an event, I use my belief to evaluate my expected payoff, and you use your belief to evaluate yours; so we can both be better off by betting against each other.

And of course, in addition to that, we may want to trade because we received different information.

Milgrom and Stokey basically show that the last piece alone is not sufficient to generate trade. In a sense, they let us trade twice. First, they let us make whatever state-contingent trades we want to, and assume that we reach an allocation which is Pareto-efficient. *Then*, they assume each of us gets new information; and they show that there can't be any *new* trades based on that new information.

But the proof isn't really interesting – basically, they show that if there was ever a trade we'd both want to make after receiving additional information, we should have been willing to make that trade ex ante, just making it contingent on different realizations of that later information. It's not quite that simple, but that's the gist of it; so if we had already traded to a Pareto-efficient allocation, new information can't make us trade more.

References

- Paul Milgrom and Nancy Stokey (1982), "Information, Trade and Common Knowledge," *Journal of Economic Theory* 26(1)

Lecture 5

Optimal Auctions – Preliminaries

Lots of auction formats are equally good...
and we can find the best possible one.

1 Before we get started...

1.1 Second-price auctions

- Last week, as an example of Bayesian Nash equilibrium, we solved for the equilibrium of a sealed-bid, first-price auction, under the assumption that n risk-neutral bidders have private values which are independently uniform on $[0, 1]$
- Before we start today, I want to talk about one other common auction format
- Description of open oral ascending, or English, auctions
- A modeling convenience: for the private values case, think of a sealed-bid *second*-price auction
  - Bidders simultaneously submit written bids, as in a first-price
  - High bidder wins, but his payment is the *second*-highest bid, not his own
- With private values, turns out it's a dominant strategy to bid your valuation!
- Think of y a random variable being the highest opponent bid; you have a choice between bidding above y (for payoff $t_i - y$) and bidding below y (and getting 0)
- So when $t_i > y$, you want to overbid, and when $t_i < y$, you want to underbid – which you accomplish by bidding t_i
- So it's a Bayesian Nash equilibrium for everyone to bid their valuation

1.2 Revenue Equivalence

- Last week, we determined that in a first-price auction, it's an equilibrium for each bidder to bid $\frac{n-1}{n}$ times his or her valuation
- So in a first-price auction, the seller's expected revenue is the expected value of the highest bid, which is $\frac{n-1}{n}$ times the expected value of the highest valuation
- Turns out, with n independent uniform random variables, the expected value of the highest is $\frac{n}{n+1}$
- So expected revenue in the first-price auction is $\frac{n-1}{n} \times \frac{n}{n+1} = \frac{n-1}{n+1}$

  (Expected value of the highest of n independent $U[0,1]$ random variables is calculated by taking the CDF, x^n; calculating the PDF, nx^{n-1}, and taking the integral $\int_0^1 x \cdot nx^{n-1}dx = \int_0^1 nx^n dx = \frac{n}{n+1}$.)
- What about the second-price auction?
- In the second-price auction, everyone bids his own type, and the payment to the seller is the second-highest bid
- So revenue is the second-highest valuation
- Which has expected value $\frac{n-1}{n+1}$

  (The CDF of the second-highest of n is $nx^{n-1}(1-x) + x^n = nx^{n-1} - (n-1)x^n$, so the PDF is $n(n-1)(x^{n-2} - x^{n-1})$, and our expected value is $\int_0^1 n(n-1)(x^{n-1} - x^n)dx = (n-1)x^n - \frac{n(n-1)}{n+1}x^{n+1}\Big|_0^1 = n - 1 - \frac{n^2-n}{n+1} = \frac{n^2-1-n^2+n}{n+1} = \frac{n-1}{n+1}$.)
- So the two auction formats give the same expected payoff to the seller
- This turns out to be a much more general result, which we'll prove soon

2 Another Preliminary: The Envelope Theorem

- Put aside strategic concerns, and think of a one-agent decision problem

$$\max_{a \in A} h(a, \theta)$$

  where a is the agent's chosen action and θ an exogenous parameter

- In an auction, θ could be your valuation, and a your choice of bid
- In another setting, θ could be the outdoor temperature, and A is the set of jackets and coats you own – given today's weather, you pick an outfit to maximize your comfort level
- (DRAW IT for discrete A)
- A could be either discrete or continuous, but let θ be continuous
- Let $a^*(\theta)$ be the set of optimal choices, and $V(\theta)$ the value function; and let h_a and h_θ denote partial derivatives of h

Theorem (The Envelope Theorem)**.** *Suppose $\forall \theta$, $a^*(\theta)$ is nonempty, and $\forall (a, \theta)$, h_θ exists. Let $a(\theta)$ be any selection from $a^*(\theta)$.*

1. *If V is differentiable at θ, then*
$$V'(\theta) = h_\theta(a(\theta), \theta)$$
2. *If V is absolutely continuous, then for any $\theta' > \theta$,*
$$V(\theta') - V(\theta) = \int_\theta^{\theta'} h_\theta(a(t), t) dt$$

- This says: the derivative of the *value function* (or maximum) is the derivative of the *objective function*, evaluated *at the maximizer*
- Absolute continuity says that $\forall \epsilon > 0$, $\exists \delta > 0$ such that for any finite, disjoint set of intervals $\{[x_k, y_k]\}_{k=1,2,\ldots,M}$ with $\sum_k |y_k - x_k| < \delta$, $\sum_k |V(y_k) - V(x_k)| < \epsilon$.

  Absolutely continuity is equivalent to V being differentiable almost everywhere and being the integral of its derivative, so the second part follows directly from the first.

  (For the types of problems we'll be dealing with, value functions will generally be absolutely continuous.)
- Let's prove the first part

2.1 Proof of the Envelope Theorem

- If V is differentiable at θ, then

$$V'(\theta) \quad = \quad \lim_{\epsilon \to 0} \frac{V(\theta+\epsilon) - V(\theta)}{\epsilon} \quad = \quad \lim_{\epsilon \to 0} \frac{V(\theta) - V(\theta-\epsilon)}{\epsilon}$$

- Now, pick $a^* \in a^*(\theta)$, so $V(\theta) = h(a^*, \theta)$
- And $V(\theta+\epsilon) = \max_a h(a, \theta+\epsilon) \geq h(a^*, \theta+\epsilon)$
- So

$$V'(\theta) \quad = \quad \lim_{\epsilon \to 0} \frac{V(\theta+\epsilon) - V(\theta)}{\epsilon} \quad \geq \quad \lim_{\epsilon \to 0} \frac{h(a^*, \theta+\epsilon) - h(a^*, \theta)}{\epsilon} \quad = \quad h_\theta(a^*, \theta)$$

- And by the same token, $V(\theta-\epsilon) = \max_a h(a, \theta-\epsilon) \geq h(a^*, \theta-\epsilon)$
- So

$$V'(\theta) \quad = \quad \lim_{\epsilon \to 0} \frac{V(\theta) - V(\theta-\epsilon)}{\epsilon} \quad \leq \quad \lim_{\epsilon \to 0} \frac{h(a^*, \theta) - h(a^*, \theta-\epsilon)}{\epsilon} \quad = \quad h_\theta(a^*, \theta)$$

- So

$$h_\theta(a^*, \theta) \quad \leq \quad V'(\theta) \quad \leq \quad h_\theta(a^*, \theta)$$

  and we're done

- The second half of the theorem is just the fact that if V is absolutely continuous, it's the integral of its derivative (wherever the derivative exists)
- Typically, for the types of problems we'll be looking at, V will be absolutely continuous
- (For example, if h_θ exists everywhere and is bounded above, that's sufficient.)
- So even if V isn't differentiable everywhere, the integral form of the theorem will be valid.

2.2 Example – using the envelope theorem to solve for equilibrium bids in FPA

- For our setup last week (PV $i.i.d. \sim U[0,1]$), we can use the envelope theorem to recover equilibrium bid functions
- Suppose a symmetric equilibrium exists, and uses a strictly-increasing bid function, so the bidder with the highest type always wins in equilibrium
- Equilibrium bids solve $\max_b (t_i - b) \Pr(b > \max_{j \neq i} b_i)$, call maximand $h(t_i, b)$ and U its max
- The partial of h with respect to t_i is $\Pr(b > \max_{j \neq i} b_i)$
- If we assume the equilibrium is symmetric, then at the maximizer, this equals the probability my type is the highest, which is t_i^{n-1}
- Since the lowest-type bidder always loses, $U(0) = 0$, so

$$U(t_i) \quad = \quad \int_0^{t_i} s^{n-1} ds \quad = \quad \frac{1}{n} t_i^n$$

- But if we calculate expected payoff based on bid and probability of winning,

$$U(t_i) \quad = \quad (t_i - \beta(t_i))\, t_i^{n-1}$$

- And if we equate these,

$$\frac{1}{n} t_i^n \quad = \quad (t_i - \beta(t_i)) t_i^{n-1}$$

  implies $\beta(t_i) = \frac{n-1}{n} t_i$
- So if a symmetric, strictly-increasing equilibrium exists, bids must be $\frac{n-1}{n} t_i$ – exactly like we found last week!

2.3 Another example – the all-pay auction

- In class, we also did the example of an *all-pay* auction – an auction where the highest bidder wins, but *every* bidder pays their bid, even the losers
- Again, if we suppose a symmetric, strictly-monotonic equilibrium exists, we can use it to calculate the expected payoff to each type of bidder, which is the same $U(t_i) = \frac{1}{n} t_i^n$
- We then write $\frac{1}{n} t_i^n = t_i^{n-1} t_i - b$ and calculate equilibrium bids as $b = \frac{n-1}{n} t_i^n$
- Finally, we can calculate the seller's expected revenue as $n E_{t_i} \frac{n-1}{n} t_i^n$ and see that it's the same $\frac{n-1}{n+1}$ as before – more anecdotal evidence of revenue equivalence!

3 Now on to Myerson's Optimal Auctions

First, we define the environment, and our goal.

- Our environment is as follows.
  - Players $N = \{1, 2, \ldots, n\}$
  - Independent types $T_i \perp T_j$
  - Player i's type T_i has distribution F_i with support $[a_i, b_i]$
  - F_i is strictly increasing on $[a_i, b_i]$, with density f_i
  - One object to be allocated, value to player i is his type t_i, and players value money linearly, so i's payoff is $t_i - x$ if he receives object and pays x, and $-x$ if he doesn't receive object but still pays x
  - (Players are risk-neutral and maximize expected payoff)
  - Seller values keeping the good at t_0
- Note that this is the environment we looked at above for first- and second-price auctions, except that type distributions F_i can be arbitrary (not just uniform) and don't have to be the same
- We're still assuming:
  - Independent private values
  - Risk-neutrality (linear valuation for money)
  - Whole environment is common knowledge
- Our challenge: design an auction (or other) protocol that maximizes the seller's expected revenue, given that the buyers play a Bayesian Nash equilibrium, over all conceivable sales mechanisms.
- That is, we basically want to solve

$$\max_{all\ \ conceivable\ \ auction\ \ formats} E_{t_1,\ldots,t_n} \{Revenue + t_0 \Pr(seller\ \ keeps\ \ object)\}$$

  subject to the constraints that the buyers have to willingly participate and play an equilibrium
- The surprising thing: this is actually doable.

4 Mechanism Design

- The framework we'll use to answer this question is referred to as **mechanism design**
- This basically puts us in the role of a seller, or government, or whatever, who asks, suppose I have complete freedom to design a game – an auction format, or a voting rule, or whatever – and people will play it; how much can I accomplish?
- The only constraints we put on the mechanism designer are the following:
  1. You can't force people to play – they have to be willing to play. Which generally means their expected payoff from the game can't be less than 0.
  2. You need to assume people will play a Bayesian Nash equilibrium within whatever game you define – that is, you can't trick people into doing things that aren't in their own interest.
- We also assume the mechanism designer has **full commitment power** – once he defines the rules of the game, the players have complete confidence that he'll honor those rules. (This is important – you'd bid differently in an auction if you thought that, even if you won, the seller might demand a higher price or mess with you some other way.)
- Broadly speaking, mechanism design takes the *environment* as given – the *players*, their *type distributions*, and their *preferences over the different possible outcomes* – and designs a game for the players to play in order to select one of the outcomes. Outcomes can be different legislative proposals, different allocations of one or more objects, etc.
- We'll be focusing on the auction problem – designing a mechanism to sell a single object, and trying to maximize expected revenue
- So the set of possible outcomes X consists of who (if anyone) gets the object, and how much each person pays
- A *mechanism* is a strategy set $S_1, S_2, \ldots, S_N$ for each player, and a mapping from strategy profiles to outcomes: $\omega : S_1 \times \cdots \times S_N \to X$
- Actually, given a set of strategies played, the outcome selected by the mechanism can be deterministic or stochastic, so really $\omega : S_1 \times \cdots \times S_N \to \Delta(X)$
- For example, the first-price auction we considered last week is one mechanism: each player's strategy set is the positive reals, the mechanism allocates the good to whoever plays the highest strategy and demands a payment from that player equal to his strategy, and there are no payments to or from anyone else

- We'll assume the players play a Bayesian Nash equilibrium within the mechanism, but in general, if there are multiple equilibria, the mechanism designer is assumed to be able to pick which equilibrium gets played
- A *performance* is a mapping of types to outcomes; we say that a given performance is *implemented* by a mechanism if that mechanism has an equilibrium where the players' equilibrium strategies lead to that mapping of types to outcomes
  - For example, in a symmetric IPV world, the first-price auction is a mechanism that implements the efficient allocation (the highest type being allocated the object), along with a particular transfer from that player
  - In general, in mechanism design, we don't worry about there being multiple equilibria, just that the one that we want is indeed one of the equilibria
  - Implicitly, we're kind of assuming that, in addition to setting up the game, the mechanism designer gets to select the equilibrium played if there is more than one
  - By using this approach, we're also kind of assuming that "all that matters" to the players is what outcome is reached, not the exact process by which it is chosen

5 Direct Revelation Mechanisms and the Revelation Principle

- Much of the time, we are able to restrict our attention to a particular class of mechanisms called *direct revelation mechanisms*
- Informally, a direct revelation mechanism consists of the mechanism designer specifying a mapping from types directly to outcomes, and asking each player (in private) to tell him their type
- Formally, this is just a mechanism where the strategy set for each player matches his type space, and given the mapping of types to outcomes, we expect there to be an equilibrium where every player reveals his type truthfully
- We'll consider only mechanisms that promise possible outcomes at every strategy profile (or reported type profile) – that is, if there is no combination of actions/types at which it promises the same object to multiple players, or anything stupid like that.

Lemma 2. *(The Revelation Principle.) Given any equilibrium of any auction mechanism, there exists an equivalent direct revelation mechanism in which truthful revelation is an equilibrium and in which the seller and all the bidders get the same expected utilities.*

- Proof is basically this: take, for example, the first-price auction
- Instead of running the first-price auction, the mechanism designer approaches each of the players and says, "tell me your type, and then I'll calculate the equilibrium bids of you and the other bidders in the BNE of the first-price auction, and if yours is highest, I'll give you the object and charge you that much"
- So any deviation (to reporting another type) in the direct-revelation mechanism, would be a deviation (to that type's equilibrium bid) in the first-price auction
- By the same logic, a direct-revelation mechanism can implement any equilibrium of any feasible auction.
- For this reason, we will focus only on direct-revelation mechanisms (since they're easier to analyze).
- Again, we do not require truthful revelation to be the only equilibrium, just that it be an equilibrium.

6 Myerson's Optimal Auctions

- Jump back to our IPV setting
  - N bidders, each with independent type $t_i \sim F_i$ with support $[a_i, b_i]$
  - We don't need symmetry, but we do need independence.
  - Seller values the object at t_0
- An *outcome* is a choice of who (if anyone) gets the object, and transfers to/from each player
- So we can summarize any direct-revelation mechanism in this world as a set of mappings

$$p_i : T_1 \times T_2 \times \cdots \times T_N \to [0, 1]$$

and

$$x_i : T_1 \times T_2 \times \cdots \times T_N \to \Re$$

where at a profile of reported types $t = (t_1, t_2, \ldots, t_n)$, player i gets the object with probability $p_i(t)$ and pays in expectation $x_i(t)$

- (When the allocation is nondeterministic, it doesn't matter whether each player pays only when he gets the object or not; since the buyers are risk-neutral, all that matters is their expected payment at each profile.)
- Define $U_i(p, x, t_i)$ to be player i's equilibrium expected payoff under mechanism (p, x), assuming he has type t_i and everyone reports their type truthfully
  - This is $E_{t_{-i}} \{t_i p_i(t_i, t_{-i}) - x_i(t_i, t_{-i})\}$ – for each opponent profile t_{-i}, player i wins with probability $p(t)$ a prize worth t_i and pays $x_i(t)$

Definition 1. *A direct revelation mechanism is FEASIBLE if it satisfies...*

1. *Plausibility: for any profile of reported types t, $\sum_{j \in N} p_j(t) \leq 1$, and $p_i(t) \geq 0$ for all $i \in N$.*
2. *Individual rationality: $U_i(p, x, t_i) \geq 0$ for all $i \in N$, all $t_i \in [a_i, b_i]$.*
3. *Incentive compatibility:*

$$U_i(p, x, t_i) \quad \geq \quad E_{t_{-i}} \left\{ t_i p_i(t_{-i}, t_i') - x_i(t_{-i}, t_i') \right\}$$

for all $t_i' \in [a_i, b_i]$

- The first condition is just that the mechanism never promises the impossible, like giving the object to two people at once
- The second condition is that nobody would choose to opt out of the game
- The third is that given the game, everyone truthfully revealing their type is a BNE
  - That is, if everyone else reports their type truthfully, I can't gain by lying about my type – it's a best-response for me to report truthfully too
- Note that, we assume the bidders already know their types before deciding whether or not to participate
  - That is, individual rationality has to hold for each type of bidder, not just in expectation over the bidders' types
  - (If bidders had to commit to play before learning their types, mechanism design would actually be much easier – basically, pick any mechanism that allocates the object efficiently, and then calculate each bidder's expected payoff within that mechanism, and charge them that much as an entry fee. You get efficient outcomes and full surplus extraction by the seller. But that isn't realistic. We assume bidders can opt out after learning their types, so we can't charge an entry fee that low-type bidders would refuse to pay.)
- Define $Q_i(p, t_i) = E_{t_{-i}} p_i(t_i, t_{-i})$, that is, given a mechanism with allocation rule p, $Q_i(t_i)$ is the probability of bidder i getting the object given type t_i
- Lemma 2 of Myerson:

Lemma 3. *A mechanism (p, x) is feasible if and only if:*

4. *Q_i is weakly increasing in t_i*
5. *$U_i(p, x, t_i) = U_i(p, x, a_i) + \int_{a_i}^{t_i} Q_i(p, s)ds$ for all i, all t_i*
6. *$U_i(p, x, a_i) \geq 0$ for all i, and*
7. *$\sum_j p_j(t) \leq 1$, $p_i \geq 0$*

To prove this, we need to show $\{1,2,3\} \leftrightarrow \{4,5,6,7\}$.

First, $\{1,2,3\} \rightarrow \{4,5,6,7\}$

- First, 1 and 7 are the same condition, so $1 \rightarrow 7$
- Second, 2 says $U_i(p,x,t_i) \geq 0$ for all t_i, so it must hold for $t_i = a_i$, which is 6, so $2 \rightarrow 6$
- So what's left to show is that 3 implies 4 and 5.
- Let's show (3) $\rightarrow$ (4), that is, IC implies Q_i increasing
  - Pick $t_i' > t_i$
  - Let $x_i = E_{t_{-i}} x_i(t_i, t_{-i})$ and $x_i' = E_{t_{-i}} x_i(t_i', t_{-i})$
  - Incentive compatibility (3) requires

$$\begin{aligned} t_i Q_i(t_i) - x_i &\geq t_i Q_i(t_i') - x_i' \\ t_i' Q_i(t_i') - x_i' &\geq t_i' Q_i(t_i) - x_i \end{aligned}$$

  - Rearranging gives

$$t_i'(Q_i(t_i') - Q_i(t_i)) \quad \geq \quad x_i' - x_i \quad \geq \quad t_i(Q_i(t_i') - Q_i(t_i))$$

    and therefore

$$(t_i' - t_i)(Q_i(t_i') - Q_i(t_i)) \quad \geq \quad 0$$

    so since $t_i' > t_i$, $Q_i(t_i') \geq Q_i(t_i)$
  - Meaning Q_i is weakly increasing, which is condition 4
- Finally, we need to show that $3 \rightarrow 5$
  - This is basically just the envelope theorem
  - By construction, everyone is truthfully announcing their types in equilibrium, so fix everyone elses' strategy (truthful reporting) and consider bidder i's problem
  - He has type t_i, and can pick a type t_i' to report, which will give him payoff $t_i Q_i(t_i') - x_i(t_i')$
  - So his expected payoff is $U_i(t_i) = \max_{t_i'} t_i Q_i(t_i') - x_i(t_i')$
  - So the envelope theorem says that $U_i'(t_i) = \frac{\partial}{\partial t_i} \left(t_i Q_i(t_i') - x_i(t_i')\right)\big|_{t_i'=t_i} = Q_i(t_i)$
  - So integrating, $U_i(t_i) - U_i(a_i) = \int_{a_i}^{t_i} U_i'(s)ds = \int_{a_i}^{t_i} Q_i(s)ds$
  - (If you don't like relying on the envelope theorem, we can also prove it directly – see next page)

Parenthetical: If We Wanted To Prove This Without the Envelope Theorem

IC implies

$$U_i(t_i') - U_i(t_i) = t_i'Q_i(t_i') - x_i' - (t_iQ_i(t_i) - x_i) \geq t_i'Q_i(t_i) - x_i - (t_iQ_i(t_i) - x_i) = (t_i' - t_i)Q_i(t_i)$$

and likewise

$$U_i(t_i') - U_i(t_i) = t_i'Q_i(t_i') - x_i' - (t_iQ_i(t_i) - x_i) \leq t_i'Q_i(t_i') - x_i' - (t_iQ_i(t_i') - x_i') = (t_i' - t_i)Q_i(t_i')$$

So for any $t_i' > t_i$,

$$(t_i' - t_i)Q_i(t_i') \geq U_i(t_i') - U_i(t_i) \geq (t_i' - t_i)Q_i(t_i)$$

- So now, we calculate $U_i(t_i) - U_i(a_1)$
- We can break the interval $[a_i, t_i]$ up into M smaller intervals, each of size $\epsilon = \frac{t_i - a_i}{M}$, and write

$$U_i(t_i) - U_i(a_i) = (U_i(a_i + \epsilon) - U_i(a_i)) + (U_i(a_i + 2\epsilon) - U_i(a_i + \epsilon)) + \cdots + (U_i(t_i) - U_i(t_i - \epsilon))$$

Using the result we just showed,

$$\begin{array}{ccccccccc} U_i(t_i) - U_i(a_i) & = & (U_i(a_i + \epsilon) - U_i(a_i)) & + & (U_i(a_i + 2\epsilon) - U_i(a_i + \epsilon)) & + & \cdots & + & (U_i(t_i) - U_i(t_i - \epsilon)) \\ & & \geq & & \geq & & & & \geq \\ & & \epsilon Q_i(a_i) & & \epsilon Q_i(a_i + \epsilon) & & & & \epsilon Q_i(t_i - \epsilon) \end{array}$$

and also

$$\begin{array}{ccccccccc} U_i(t_i) - U_i(a_i) & = & (U_i(a_i + \epsilon) - U_i(a_i)) & + & (U_i(a_i + 2\epsilon) - U_i(a_i + \epsilon)) & + & \cdots & + & (U_i(t_i) - U_i(t_i - \epsilon)) \\ & & \leq & & \leq & & & & \leq \\ & & \epsilon Q_i(a_i + \epsilon) & & \epsilon Q_i(a_i + 2\epsilon) & & & & \epsilon Q_i(t_i) \end{array}$$

- If we make M big, so each interval (and ϵ) get small, we have an upper and lower bound on $U_i(t_i) - U_i(a_i)$ that go to the same limit – which is the Riemann integral of Q_i over $[a_i, t_i]$
- So
$$U_i(t_i) - U_i(a_i) = \int_{a_i}^{t_i} Q_i(s)ds$$
which is exactly property 5, so we're done with the first direction.
- (Note that Q_i is Riemann-integrable because it's bounded (always within $[0, 1]$) and, since it's monotonically increasing, it must be continuous almost everywhere (if it had uncountably many discontinuities, it couldn't be bounded).)
- (Or if we want to see it "more directly," the upper and lower bound have the same limit because the two sums differ by exactly $\epsilon Q_i(t_i) - \epsilon Q_i(a_i)$, but $Q_i(t_i)$ and $Q_i(a_i)$ are both bounded within $[0, 1]$, so the two sums differ by at most ϵ, which goes to 0.)

Back to Work – still need to show $\{4,5,6,7\} \to \{1,2,3\}$

- Again, 1 and 7 are the same
- 5 implies U_i is increasing; along with $U_i(a_i) \geq 0$, this implies IR holds everywhere (2)
- All that's left is to show that 4 and 5 together imply 3 – that is, that monotonicity and the envelope condition together imply incentive compatibility
- Given a true type t_i, we can write the gain from deviating and reporting a type t_i' as

$$\begin{aligned} gain &= t_i Q_i(t_i') - x_i' - U_i(t_i) \\ &= t_i' Q_i(t_i') - (t_i' - t_i) Q_i(t_i') - x_i' - U_i(t_i) \\ &= U_i(t_i') - U_i(t_i) - (t_i' - t_i) Q_i(t_i') \end{aligned}$$

- For the case where $t_i' > t_i$, using 5, this is

$$gain = \int_{t_i}^{t_i'} Q_i(s) ds - (t_i' - t_i) Q_i(t_i') = \int_{t_i}^{t_i'} \left[Q_i(s) - Q_i(t_i') \right] ds$$

- But since (4) Q_i is weakly increasing, the integrand is nonpositive, so the gain is nonpositive
- For the case where $t_i' < t_i$,

$$gain = -\int_{t_i'}^{t_i} Q_i(s) ds + (t_i - t_i') Q_i(t_i') = \int_{t_i'}^{t_i} \left[Q_i(t_i') - Q_i(s) \right] ds$$

  which is once again nonpositive because Q_i is nondecreasing
- So together, 4 and 5 imply 3 – and we're done.

That proves the lemma.

7 Formulating the Seller's Problem

- We'll, we've now established that:

  1. we can accomplish something with any mechanism if and only if we can accomplish with a direct mechanism
  2. we can accomplish something with a direct mechanism if and only if we can accomplish it with a mechanism satisfying conditions (4), (5), (6), and (7)

- So our task is now, very literally, to solve the seller's problem: to maximize expected revenue (or really, expected seller profit) over all possible mechanisms (p, x), subject to the constraints (4), (5), (6), and (7).

- We can write the seller's expected payoff as

$$E_t \left\{ \sum_{i \in N} x_i(t) + t_0 \left(1 - \sum_{i \in N} p_i(t) \right) \right\}$$

  and so our problem is to maximize this, over all direct-revelation mechanisms, subject to (4), (5), (6), and (7)

- And that's where we'll start next week.

Lecture 6

Optimal Auctions – Conclusions

1 Recap

Last week, we...

- Set up the Myerson auction environment:
  - n risk-neutral bidders
  - independent types $t_i \sim F_i$ with support $[a_i, b_i]$
  - residual valuation of t_0 for the seller
- Named our goal: maximize expected seller payoff over all conceivable auction/sales formats, subject to two constraints: bidders participate voluntarily and play equilibrium strategies
- Defined mechanisms, and showed the Revelation Principle – that without loss of generality, we can restrict attention to direct revelation mechanisms
- Showed that feasibility of a mechanism is equivalent to four conditions holding:
  4. Q_i is weakly increasing in t_i
  5. $U_i(t_i) = U_i(a_i) + \int_{a_i}^{t_i} Q_i(p, s)ds$ for all i, all t_i
  6. $U_i(a_i) \geq 0$ for all i, and
  7. $\sum_j p_j(t) \leq 1$, $p_i \geq 0$
- So we defined our goal as solving

$$\max_{direct\ \ revelation\ \ mechanisms} E_t \left\{ \sum_{i \in N} x_i(t) + t_0 \left(1 - \sum_{i \in N} p_i(t) \right) \right\} \qquad s.t. \qquad (4), (5), (6), (7)$$

- So... onward!

2 Rewriting our Objective Function

By adding and subtracting $E_i \sum_i p_i(t)t_i$, we can rewrite the seller's objective function as

$$\begin{aligned} U_0(p,x) &= E_t\left\{\sum_{i\in N} x_i(t) + t_0\left(1 - \sum_{i\in N} p_i(t)\right)\right\} \\ &= t_0 + \sum_{i\in N} E_t p_i(t)(t_i - t_0) - \sum_{i\in N} E_t(p_i(t)t_i - x_i(t)) \end{aligned}$$

Note that...

- the first term is a constant (the payoff to doing nothing)
- the second term is the total surplus created by selling the object
- the third term, which is being subtracted, is the expected payoff going to the bidders

Next, we do some work expressing the last term, expected bidder surplus, in a different form:

$$\begin{aligned} E_t(p_i(t)t_i - x_i(t)) &= \textstyle\int_{a_i}^{b_i} U_i(t_i) f_i(t_i) dt_i \\ &= \textstyle\int_{a_i}^{b_i} \left(U_i(a_i) + \int_{a_i}^{t_i} Q_i(s_i) ds_i\right) f_i(t_i) dt_i \\ &= U_i(a_i) + \textstyle\int_{a_i}^{b_i}\int_{a_i}^{t_i} Q_i(s_i) f_i(t_i) ds_i dt_i \\ &= U_i(a_i) + \textstyle\int_{a_i}^{b_i}\int_{s_i}^{b_i} f_i(t_i) Q_i(s_i) dt_i ds_i \\ &= U_i(a_i) + \textstyle\int_{a_i}^{b_i} (1 - F_i(s_i)) Q_i(s_i) ds_i \\ &= U_i(a_i) + \textstyle\int_{a_i}^{b_i} (1 - F_i(s_i)) \left(\int_{t_{-i}} p_i(s_i, t_{-i}) f_{-i}(t_{-i}) dt_{-i}\right) ds_i \\ &= U_i(a_i) + \textstyle\int_{a_i}^{b_i} \frac{1-F_i(s_i)}{f_i(s_i)} \left(\int_{t_{-i}} p_i(s_i, t_{-i}) f_{-i}(t_{-i}) dt_{-i}\right) f_i(s_i) ds_i \\ &= U_i(a_i) + \textstyle\int_T \frac{1-F_i(s_i)}{f_i(s_i)} p_i(t_{-i}, s_i) f(t_{-i}, s_i) dt_{-i} ds_i \\ &= U_i(a_i) + \textstyle\int_T \frac{1-F_i(t_i)}{f_i(t_i)} p_i(t) f(t) dt \\ &= U_i(a_i) + E_t p_i(t) \frac{1-F_i(t_i)}{f_i(t_i)} \end{aligned}$$

- So we can rewrite the objective function as

$$\begin{aligned} U_0(p,x) &= t_0 + \sum_{i\in N} E_t p_i(t)(t_i - t_0) - \sum_{i\in N} U_i(a_i) - \sum_{i\in N} E_t p_i(t)\frac{1-F_i(t_i)}{f_i(t_i)} \\ &= t_0 + \sum_{i\in N} E_t p_i(t)\left(t_i - \frac{1-F_i(t_i)}{f_i(t_i)} - t_0\right) - \sum_{i\in N} U_i(a_i) \end{aligned}$$

- So the seller's problem amounts to choosing an allocation rule p and expected payoffs for the low types $U_i(a_i)$ to maximize

$$t_0 + E_t\left\{\sum_{i\in N} p_i(t)\left(t_i - \frac{1-F_i(t_i)}{f_i(t_i)} - t_0\right)\right\} - \sum_{i\in N} U_i(a_i)$$

  subject to feasibility – which just means p plausible, Q_i increasing in type, and $U_i(a_i) \geq 0$

- (The envelope formula for bidder payoffs is no longer a constraint – we've already imposed it.)

- Once we find a mechanism we like, each U_i is uniquely determined by the envelope formula, and so the rest of the transfers x_i are set to satisfy those required payoffs.

- Once we've phrased the problem in this way, Myerson points out, **revenue equivalence** becomes a straightforward corollary:

Corollary 1. *For a given environment, the expected revenue of an auction depends only on the equilibrium allocation rule and the expected payoffs of the lowest possible type of each bidder.*

The Revenue Equivalence Theorem is usually stated in this way:

Corollary 2. *Suppose bidders have symmetric independent private values and are risk-neutral. Define a* standard auction *as an auction where the following two properties hold:*

1. *In equilibrium, the bidder with the highest valuation always wins the object*
2. *The expected payment from a bidder with the lowest possible type is* 0

Any two standard auctions give the same expected revenue.

Two standard auctions also give the same expected surplus to each type of each bidder $U_i(t_i)$.

So this means with symmetry, independence, and risk-neutrality, *any* auction with a symmetric, strictly-monotone equilibrium gives the same expected revenue. (Examples.)

Now, back to maximizing expected revenue

- We've redefined the problem as choosing p and $U_i(a_i)$ to maximize

$$t_0 + E_t\left\{\sum_{i\in N} p_i(t)\left(t_i - \frac{1-F_i(t_i)}{f_i(t_i)} - t_0\right)\right\} - \sum_{i\in N} U_i(a_i)$$

- Clearly, to maximize this, we should set $U_i(a_i) = 0$ for each i
- This leaves Myerson's Lemma 3:

Lemma 4. *If p maximizes*

$$t_0 + E_t\left\{\sum_{i\in N} p_i(t)\left(t_i - \frac{1-F_i(t_i)}{f_i(t_i)} - t_0\right)\right\}$$

subject to Q_i increasing in t_i and p possible, and

$$x_i(t) = t_i p_i(t) - \int_{a_i}^{t_i} p_i(t_{-i}, s_i)ds_i$$

then (p, x) is an optimal auction.

- The transfers are set to make $U_i(a_i) = 0$ and give payoffs required by the envelope theorem
- To see this, fix t_i and take the expectation over t_{-i}, and we find

$$E_{t_{-i}} x_i(t) = t_i Q(t_i) - \int_{a_i}^{t_i} Q_i(s_i)ds_i$$

  or

$$\int_{a_i}^{t_i} Q_i(s_i)ds_i = t_i Q_i(t_i) - E_{t_{-i}} x_i(t_i, t_{-i}) = U_i(t_i)$$

  which is exactly the envelope theorem combined with $U_i(a_i) = 0$

- (The exact transfers $x_i(t)$ are not uniquely determined by incentive compatibility and the allocation rule p; what is uniquely pinned down is $E_{t_{-i}} x_i(t_i, t_{-i})$, because this is what's payoff-relevant to bidder i. The transfers above are just one rule that works.)
- Next, we consider what the solution looks like for various cases.

3 The Regular Case

- With one additional assumption, things fall into place very nicely.
- Define a distribution F_i to be *regular* if

$$t_i - \frac{1 - F_i(t_i)}{f_i(t_i)}$$

  is strictly increasing in t_i

  - This is not that crazy an assumption
  - Most familiar distributions have increasing hazard rates – that is, $\frac{f}{1-F}$ is increasing, which would imply $\frac{1-F}{f}$ decreasing
  - This is a weaker condition, since $\frac{f}{1-F}$ is allowed to decrease, just not too quickly.
- When the bid distributions are all regular, the optimal auction becomes this:
  - Calculate $c_i(t_i) = t_i - \frac{1-F_i(t_i)}{f_i(t_i)}$ for each player
  - If $\max_i c_i(t_i) < t_0$, keep the good; if $\max_i c_i(t_i) \geq t_0$, award the good to the bidder with the highest value of $c_i(t_i)$
  - Charge the transfers determined by incentive compatibility and this allocation rule
- Note that this rule makes Q_i monotonic – $Q_i(t_i) = 0$ for $t_i < c_i^{-1}(t_0)$, and $\prod_{j \neq i} F_j \left(c_j^{-1}(c_i(t_i)) \right)$ above it
- So the rule satisfies our two constraints, and it's obvious that it maximizes the seller's objective function
- There's even a nice interpretation of the x_i we defined above. Fixing everyone else's type, p_i is 0 when $c_i(t_i) < \max\{t_0, \max\{c_j(t_j)\}\}$ and 1 when $c_i(t_i) > \max\{t_0, \max\{c_j(t_j)\}\}$, so this is just

$$x_i(t) = t_i - \int_{t_i^*}^{t_i} ds_i = t_i - (t_i - t_i^*) = t_i^*$$

  where t_i^* is the lowest type that i could have reported (given everyone elses' reports) and still won the object.
- This payment rule makes incentive-compatibility obvious: for each combination of my opponents' bids, I face some cutoff t_i^* such that if I report $t_i > t_i^*$, I win and pay t_i^*; and if I report less than that, I lose and pay nothing. Since I want to win whenever $t_i > t_i^*$, just like in a second-price auction, my best-response is to bid my type.

3.1 Symmetric IPV

In the case of symmetric IPV, each bidder's c function is the same as a function of his type, that is,

$$c_i(t_i) = c(t_i) = t_i - \frac{1 - F(t_i)}{f(t_i)}$$

which is strictly increasing in t_i. This means the bidder with the highest type has the highest c_i, and therefore gets the object; and so his payment is the reported type of the next-highest bidder, since this is the lowest type at which he would have won the object. Which brings us to our first claim:

Theorem 1. *With symmetric independent private values, the optimal auction is a second-price auction with a reserve price of* $c^{-1}(t_0)$.

Note, though, that even when $t_0 = 0$, this reserve price will be positive. The optimal auction is not efficient – since $c(t_i) < t_i$, the seller will sometimes keep the object even though the highest bidder values it more than him – but he never allocates it to the "wrong" bidder.

Also interesting is that the optimal reserve price under symmetric IPV does not depend on the number of bidders – it's just $c^{-1}(t_0)$, regardless of N.

3.2 Asymmetric IPV

When the bidders are not symmetric, things are different. With different F_i, it will not always be true that the bidder with the highest c_i also has the highest type; so sometimes the winning bidder will not be the bidder with the highest value. (As we'll see later, efficiency is not standard in auctions with asymmetric bidders: even a standard first-price auction is sometimes not be won by the bidder with the highest value.)

One special case that's easy to analyze: suppose every bidder's bid is drawn from a uniform distribution, but uniform over different intervals. That is, suppose each F_i is the uniform distribution over a (potentially) different interval $[a_i, b_i]$. Then

$$c_i(t_i) = t_i - \frac{1 - F_i(t_i)}{f_i(t_i)} = t_i - \frac{(b_i - t_i)/(b_i - a_i)}{1/(b_i - a_i)} = t_i - (b_i - t_i) = 2t_i - b_i$$

So the optimal auction, in a sense, penalizes bidders who have high maximum valuations. This is to force them to bid more aggressively when they have high values, in order to extract more revenue; but the price of this is that sometimes the object goes to the wrong bidder.

3.3 What About The Not-Regular Case?

Myerson does solve for the optimal auction in the case where c_i is not increasing in t_i, that is, where the auction above would not be feasible. Read the paper if you're interested.

4 Bulow and Klemperer, "Auctions versus Negotiations"

We just learned that with symmetric IPV and risk-neutral bidders, the best you can possibly do is to choose the perfect reserve price and run a second-price auction. This might suggest that choosing the perfect reserve price is important. There's a paper by Bulow and Klemperer, "Auctions versus Negotiations," that basically says: nah, it's not that important. Actually, what they say is, it's better to attract one more bidder than to run the perfect auction.

Suppose we're in a symmetric IPV world where bidders' values are drawn from some distribution F on $[a, b]$, and the seller values the object at t_0. Bulow and Klemperer show the following: as long as $a \geq t_0$ (all bidders are "serious"), the optimal auction with N bidders gives lower revenue than a second-price auction with no reserve price and $N + 1$ bidders.

To see this, recall that we wrote the auctioneer's expected revenue as

$$t_0 + E_t \left\{ \sum_{i \in N} p_i(t) \left(t_i - \frac{1 - F_i(t_i)}{f_i(t_i)} - t_0 \right) \right\} - \sum_{i \in N} U_i(a_i)$$

Consider mechanisms where $U_i(a_i) = 0$, and define the **marginal revenue** of bidder i as

$$MR_i = t_i - \frac{1 - F_i(t_i)}{f_i(t_i)}$$

so expected revenue is

$$E_t \left\{ \sum_{i \in N} p_i(t) MR_i(t) + \left(1 - \sum_{i \in N} p_i(t) \right) t_0 \right\}$$

So if we think of the seller as being another possible buyer, with marginal revenue of t_0, then the expected revenue is simply the expected value of the marginal revenue of the winner.

Jump back to the symmetric case, so $F_i = F$. Continue to assume regularity. In an ordinary second-price or ascending auction, with no reserve price, the object sells to the bidder with the highest type, which is also the bidder with the highest marginal revenue; so the expected revenue in this type of auction (what Bulow and Klemperer call an "absolute English auction") is

$$Expected \;\; Revenue = E_t \max\{MR(t_1), MR(t_2), \ldots, MR(t_N)\}$$

(This is Bulow and Klemperer Lemma 1.)

The fact that expected revenue = expected marginal revenue of winner also makes it clear why the optimal reserve price is $MR^{-1}(t_0)$ – this replaces bidders with marginal revenue less than t_0 with t_0. So (counting the seller's value from keeping the unsold object) an English auction with an optimal reserve price has expected revenue

$$Expected \;\; Revenue = E_t \max\{MR(t_1), MR(t_2), \ldots, MR(t_N), t_0\}$$

So here's the gist of Bulow and Klemperer, "Auctions Versus Negotiations." They compare the simple ascending auction with $N + 1$ bidders, to the optimal auction with N bidders. (We

discovered last week that with symmetric independent private values, the optimal auction is an ascending auction with a reserve price of $MR^{-1}(t_0)$.) The gist of Bulow and Klemperer is that the former is higher, that is, that

$$E\max\{MR(t_1), MR(t_2), \ldots, MR(t_N), MR(t_{N+1})\} \geq E\max\{MR(t_1), MR(t_2), \ldots, MR(t_N), t_0\}$$

so the seller gains more by attracting one more bidder than by holding the "perfect" auction. (They normalize t_0 to 0, but this doesn't change anything.)

Let's prove this. The proof has a few steps.

First of all, note that the expected value of $MR(t_i)$ is a, the lower bound of the support. This is because

$$\begin{aligned} E_{t_i} MR(t_i) &= E_{t_i}\left(t_i - \tfrac{1-F(t_i)}{f(t_i)}\right) \\ &= \int_a^b \left(t_i - \tfrac{1-F(t_i)}{f(t_i)}\right) f(t_i) dt_i \\ &= \int_a^b \left(t_i f(t_i) - 1 + F(t_i)\right) dt_i \end{aligned}$$

Now, $tf(t) + F(t)$ has integral $tF(t)$, so this integrates to

$$t_i F(t_i)|_{t_i=a}^{b} - (b-a) = b - 0 - (b-a) = a$$

which by assumption is at least t_0. So $E(MR(t_i)) \geq t_0$.

Next, note that for fixed x, the function $g(y) = \max\{x, y\}$ is convex, so by Jensen's inequality,

$$E_y \max\{x, y\} \geq \max\{x, E(y)\}$$

If we take an expectation over x, this gives us

$$E_x \left\{E_y \max\{x, y\}\right\} \geq E_x \max\{x, E(y)\}$$

or

$$E\max\{x, y\} \geq E\max\{x, E(y)\}$$

Now let $x = \max\{MR(t_1), MR(t_2), \ldots, MR(t_N)\}$ and $y = MR(t_{N+1})$;

$$E\max\{MR(t_1), MR(t_2), \ldots, MR(t_N), MR(t_{N+1})\} \geq$$

$$E\max\{MR(t_1), MR(t_2), \ldots, MR(t_N), E(MR(t_{N+1}))\} \geq$$

$$E\max\{MR(t_1), MR(t_2), \ldots, MR(t_N), t_0\}$$

and that's the proof.

Finally (and leading to the title of the paper), Bulow and Klemperer point out that "negotiations" – really, any process for allocating the object and determining the price – cannot outperform the optimal mechanism, and therefore leads to lower expected revenue than a simple ascending auction with one more bidder. They therefore argue that a seller should never agree to an early "take-it-or-leave-it" offer from one buyer when the alternative is an ascending auction with at least one more buyer, etc.

References

- Roger Myerson (1981), "Optimal Auctions," *Mathematics of Operations Research* 6
- Jeremy Bulow and Paul Klemperer (1996), "Auctions Versus Negotiations," *American Economic Review* 86

Lecture 7

Bilateral Private Information

If buyer and seller both have private information,
full efficiency is impossible.

- In the last two classes, we've solved the seller's optimization problem when facing several buyers with private information...
- Under the assumption that the bidders are risk-neutral and have *independent* private values
- We found that...
  - In the symmetric case, the best the seller can do is a "standard" auction with an optimally-set reserve price
  - In the asymmetric case, we can derive a rule for the optimal auction, but it doesn't correspond to anything "obvious"
- Today, we'll do two extensions
- First: what happens when bidders' values are correlated, rather than independent?
- And second: what happens when both buyer and seller have private information?

1 Mechanism Design with Correlated Values

- If bidders' values are correlated, things change significantly
- In fact, if bidders have correlated values, this is good for the seller
- Under the assumptions we've been making – the entire environment (including the joint distribution of bidders' values) is known to both the seller and all the bidders – the seller can use this correlation against the bidders, to extract more of the surplus
- Myerson gives an example to show that the seller may be able to do the best he could *possibly* do – achieve the efficient outcome (maximize total combined payoffs), *and* give every bidder expected payoff 0 regardless of type, so he's maximizing total surplus and capturing all of it

  Given individual rationality, that's the best the seller could possibly hope to do
- The example is discrete, but the intuition extends to continuous cases as well

  In a later paper, Cremer and McLean give a sufficient condition for this type of trick to work in general.
- So, here's Myerson's example.
  - Two bidders
  - The joint distribution of their types (t_1, t_2) is $\Pr(100, 100) = \Pr(10, 10) = \frac{1}{3}$, $\Pr(10, 100) = \Pr(100, 10) = \frac{1}{6}$
  - For simplicity, suppose $t_0 = 0$
- Consider the following direct mechanism:
  - If both bidders report high types, flip a coin, and give one of them the object for 100
  - If one reports high and one reports low, sell the high guy the object for 100, charge the low guy 30 and give him nothing
  - If both report low, flip a coin, give one 15 and give the other 5 plus the object
- First, we'll verify that this is feasible, that is, that truthful revelation is an equilibrium; then we'll consider the logic behind it
  - Suppose your opponent will be reporting his true type
  - If you have a low value, you know, conditional on that, the the other guy is low with probability $\frac{2}{3}$. So if you declare low, you get an expected payoff of

$$\frac{2}{3}(15) + \frac{1}{3}(-30) = 0$$

- On the other hand, if you misreport your type as high, then with probability $\frac{2}{3}$, your opponent reports low and you pay 100 for an object you value at 10; and with probability $\frac{1}{3}$, you both report high, and with probability $\frac{1}{2}$ you pay 100 for the object; so your payoff is
$$\frac{2}{3}(10-100)+\frac{1}{3}\cdot\frac{1}{2}(10-100)=\frac{5}{6}(-90)<0$$
so your best response is clearly to report truthfully
- Now suppose you have a high value
- If you report high, then either you win the thing and pay 100, or you get nothing, so your payoff is 0
- If you report low, then with probability $\frac{2}{3}$, your opponent reports high and you lose 30. With probability $\frac{1}{3}$, your opponent reports low, and you either get the object plus 5, or you get 15; so your expected payoff is
$$\frac{2}{3}(-30)+\frac{1}{3}\left(\frac{1}{2}(15)+\frac{1}{2}(105)\right)=-20+\frac{1}{6}(120)=0$$
- So it's a weak best-response to report truthfully

- So this mechanism is incentive compatible
- Since both types get expected payoff 0 in equilibrium, it's also individually rational
- So it's feasible
- Also note that under truthful revelation, it's efficient – the object is always allocated to someone, and a low type never gets it if a high type is available
- So the seller is achieving both efficiency and full surplus extraction – his best-case scenario

Now, where the hell did this auction come from?

- In a "normal" setting, bidders with high types get positive expected payoffs, because you have to "leave them" some surplus to prevent them from lying and saying they have a low value
- But in this case, a bidder with a high type has information about the other type's likely value
- So what you do here is this: when a bidder claims to have a low type, you also force him to accept a "side-bet" with you about the other bidder's type
- And this bet is rigged to have 0 expected value when his type is actually low, but negative expected value when his type is actually high

- Since bidders are risk-neutral, this doesn't hurt low types, but it lowers the payoff the high type gets from misreporting his type; so it lowers the payoff you have to give him when he reports truthfully
- To see how this mechanism works, think of it this way:
  - Hold a first-price auction, where the only bids allowed are 10 and 100
  - But a bidder who bids 10 also accepts the following bid: "if the other bidder bids 10, I win 15; if the other bidder bids 100, I lose 30"
  - A bidder with low type expects his opponent to have a low type two-thirds of the time, so this bet has 0 expected payoff
  - A bidder with high type expects his opponet to have a high type two-thirds of the time, so this bet has expected value $\frac{1}{3}(15) + \frac{2}{3}(-30) = -15$
  - Since a high bidder who bids low only wins with probability $\frac{1}{6}$, this wipes out the surplus of 90 he would get the times he did win.
- Cremer and McLean (1988, "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions", Econometrica 56.6) generalize this, showing when full surplus extraction is possible with correlated types
  - They assume finite, discrete types
  - Full surplus extraction is possible if for each bidder, the matrix of probabilities of his opponents' types given his own types has full rank – so that his beliefs about his opponents are different enough for the right side bets to be created
- However, this sort of mechanism does not seem very robust
  - more than most auctions, it is extremely sensitive to the seller being right about the true distribution of types
  - it also clearly requires risk-neutrality, since you need the low types to be willing to accept a large zero-expected-value bet
  - finally, it's very sensitive to collusion – both bidders bidding low is profitable for them (it's even an equilibrium!)
- I don't think Myerson is suggesting you would actually run this auction in this way – just making the point that when bidder values are correlated, the optimal auction may be complicated but may outperform anything you would think of as a "regular-looking" auction.

2 Bilateral Private Information

- So far, we've focused on the case of "one-sided" private information
  - buyers have private information, but the seller doesn't
  - or if he does, it's doesn't matter, because he's the one designing the mechanism
- Now we shift to the case where there is a single buyer and a single seller, but both have private information about his own valuation of the good
- Environment: $v_b \sim F_b$, $v_s \sim F_s$
- For simplicity, assume $\text{supp}(F_b) = \text{supp}(F_s) = [0, 1]$
- Think of v_s as the cost to the seller from giving up the good
  - just like the buyer gets payoff $v_b - p$ from buying, the seller gets $p - v_s$ from selling
  - and both get payoff 0 if there is no trade
- The general problem is to maximize ex ante surplus
  - imagine buyer and seller choose a mechanism *before* either one knows their type, to maximize expected payoff once they learn types and trade
  - but individual rationality still has to hold for each type – after they learn their types, either one could still refuse to play
- The revelation principle still holds – we can imagine both buyer and seller report their types to a neutral third party ("mediator"), who then tells them whether to trade and how much to pay each other
- Let $p(v_b, v_s)$ be the probability they trade, given types v_b and v_s
- The question: is there a feasible mechanism where trade happens whenever $v_b > v_s$, i.e.,

$$p(v_b, v_s) \quad = \quad \begin{cases} 1 & if \quad v_b > v_s \\ 0 & if \quad v_b < v_s \end{cases}$$

- The answer turns out to be no. From Myerson and Satterthwaite (1983):

Theorem. *If buyer and seller each have private information about their own private value, and the support of their valuations overlap, there is no feasible mechanism that yields fully efficient trade.*

- To prove it, we will suppose such a mechanism did exist, and then show it would have to violate individual rationality for either low-value buyers or high-cost sellers.
- Remember, anything we can accomplish with any mechanism, we can do with a direct mechanism, so we'll do that
- Let $p_b(v_b) = E_{v_s} p(v_b, v_s)$ be a buyer's expected probability of trade given reported type v_b, and let $x_b(v_b)$ be his expected payment
- Given the seller is reporting truthfully, a buyer with type v_b gets expected payoff

$$U_b(v_b) \quad = \quad \max_{v_b'} v_b p_b(v_b') - x_b(v_b')$$

- The envelope theorem says that $U_b'(v_b) = p_b(v_b)$, so

$$U_b(v_b) \quad = \quad U_b(0) + \int_0^{v_b} p_b(v) dv$$

- Letting $U_b = E_{v_b} U_b(v_b)$ be the bidder's ex-ante expected payoff,

$$\begin{aligned} U_b &= \int_0^1 \left[U_b(0) + \int_0^{v_b} p_b(v) dv \right] f_b(v_b) dv_b \\ &= U_b(0) + \int_0^1 \left[\int_v^1 f_b(v_b) dv_b \right] p_b(v) dv \\ &= U_b(0) + \int_0^1 (1 - F_b(v)) p_b(v) dv \end{aligned}$$

- Similarly, if we let $p_s(v_s)$ and $x_s(v_s)$ be the seller's probability of trade and expected payment received,

$$U_s(v_s) \quad = \quad \max_{v_s'} x_s(v_s') - v_s p_s(v_s')$$

- So by the envelope theorem, $U_s'(v_s) = -p_s(v_s)$, and therefore

$$U_s(v_s) \quad = \quad U_s(1) + \int_{v_s}^1 p_s(v) dv$$

- Taking the expected value over v_s, then,

$$\begin{aligned} U_s &= \int_0^1 \left[U_s(1) + \int_{v_s}^1 p_s(v) dv \right] f_s(v_s) dv_s \\ &= U_s(1) + \int_0^1 \left[\int_0^v f_s(v_s) dv_s \right] p_s(v) dv \\ &= U_s(1) + \int_0^1 F_s(v) p_s(v) dv \end{aligned}$$

- But we can also calculate combined payoffs as expected gains from trade, which are

$$W = \int_0^1 \int_0^1 (v_b - v_s) p(v_b, v_s) f_b(v_b) f_s(v_s) dv_s dv_b$$

- Since $U_b + U_s = W$, this means

$$U_b(0) + \int_0^1 (1 - F_b(v)) p_b(v) dv + U_s(1) + \int_0^1 F_s(v) p_s(v) dv$$

$$= \int_0^1 \int_0^1 (v_b - v_s) p(v_b, v_s) f_b(v_b) f_s(v_s) dv_s dv_b$$

- Now, $p_b(v_b) = \int_0^1 p(v_b, v_s) f_s(v_s) dv_s$, and $p_s(v) = \int_0^1 p(v_b, v_s) f_b(v_b) dv_b$, so we can write these as

$$U_b(0) + \int_0^1 \int_0^1 \frac{1 - F_b(v_b)}{f_b(v_b)} p(v_b, v_s) f_b(v_b) dv_b f_s(v_s) dv_s$$

$$+ \quad U_s(1) + \int_0^1 \int_0^1 \frac{F_s(v_s)}{f_s(v_s)} p(v_b, v_s) f_b(v_b) f_s(v_s) dv_b dv_s$$

$$= \int_0^1 \int_0^1 (v_b - v_s) p(v_b, v_s) f_b(v_b) f_s(v_s) dv_b dv_s$$

- If we move all the integrals to the right hand side, we get

$$U_b(0) + U_s(1) = \int_0^1 \int_0^1 \left[\left(v_b - \frac{1 - F_b(v_b)}{f_b(v_b)} \right) - \left(v_s + \frac{F_s(v_s)}{f_s(v_s)} \right) \right] p(v_b, v_s) f_b(v_b) f_s(v_s) dv_b dv_s$$

for any feasible mechanism, which is basically Theorem 1 of Myerson and Satterthwaite.

  - How to interpret this? Think of a mechanism designer setting up a mechanism to simuiltaneously sell to the buyer and buy from the seller
  - $v_b - \frac{1-F_b(v_b)}{f_b(v_b)}$ is the old Myerson term – the "marginal revenue" from selling to a buyer of type v_b, given incentive compatibility
  - $v_s + \frac{F_s(v_s)}{f_s(v_s)}$ is the analogous term for the seller – the "marginal cost" of buying the good from a seller of type v_s
  - So this is the expected value of, whenever we want the two parties to trade, simultaneously buying from the seller and reselling to the buyer
  - If we were a middle man setting up a mechanism to maximize our own profit, the RHS is the thing that we would maximize
  - Here, there's no middleman, just the actual buyer and seller, so that's the "excess" surplus generated by the mechanism (beyond the payoffs needed to satisfy incentive compatibility) – which is the extra payoff one or both of them get

- For our purposes, we want to show there is no way to get trade whenever it's efficient – that is, there's no feasible mechanism such that

$$p(v_b, v_s) \quad = \quad \begin{cases} 1 & if \quad v_b > v_s \\ 0 & if \quad v_b < v_s \end{cases}$$

- To prove this, suppose there was such a mechanism

- In that case, we would get

$$U_b(0) + U_s(1) \quad = \quad \int_0^1 \int_0^{v_b} \left[\left(v_b - \frac{1 - F_b(v_b)}{f_b(v_b)} \right) - \left(v_s + \frac{F_s(v_s)}{f_s(v_s)} \right) \right] f_b(v_b) f_s(v_s) dv_s dv_b$$

- Split that into two integrals,

$$\begin{aligned} U_b(0) + U_s(1) \quad &= \quad \int_0^1 \int_0^{v_b} \left(v_b - \frac{1 - F_b(v_b)}{f_b(v_b)} \right) f_b(v_b) f_s(v_s) dv_s dv_b \\ &\quad - \int_0^1 \int_0^{v_b} \left(v_s + \frac{F_s(v_s)}{f_s(v_s)} \right) f_b(v_b) f_s(v_s) dv_s dv_b \end{aligned}$$

- Simplifying the first one gives

$$\begin{aligned} INT1 \quad &= \quad \int_0^1 \int_0^{v_b} \left(v_b - \frac{1 - F_b(v_b)}{f_b(v_b)} \right) f_b(v_b) f_s(v_s) dv_s dv_b \\ &= \quad \int_0^1 \left(\int_0^{v_b} f_s(v_s) dv_s \right) \left(v_b - \frac{1 - F_b(v_b)}{f_b(v_b)} \right) f_b(v_b) dv_b \\ &= \quad \int_0^1 F_s(v_b) \left(v_b f_b(v_b) - (1 - F_b(v_b)) \right) dv_b \end{aligned}$$

- Simplifying the second one gives

$$\begin{aligned} INT2 \quad &= \quad \int_0^1 \int_0^{v_b} \left(v_s + \frac{F_s(v_s)}{f_s(v_s)} \right) f_b(v_b) f_s(v_s) dv_s dv_b \\ &= \quad \int_0^1 \left(\int_0^{v_b} \left(v_s f_s(v_s) + F_s(v_s) \right) dv_s \right) f_b(v_b) dv_b \\ &= \quad \int_0^1 \left(v_s F_s(v_s) |_{v_s=0}^{v_b} \right) f_b(v_b) dv_b \\ &= \quad \int_0^1 v_b F_s(v_b) f_b(v_b) dv_b \end{aligned}$$

- Recombining the two integrals, then, gives

$$\begin{aligned} U_b(0) + U_s(1) &= INT1 - INT2 \\ &= \int_0^1 F_s(v_b)\left(v_b f_b(v_b) - 1 + F_b(v_b)\right) dv_b - \int_0^1 v_b F_s(v_b) f_b(v_b) dv_b \\ &= -\int_0^1 F_s(v_b)\left(1 - F_b(v_b)\right) dv_b \end{aligned}$$

- As long as the supports of the two distributions overlap, the right side is strictly negative
- Which means any mechanism that gave efficient trade would have to give $U_b(0) + U_s(1) < 0$
- Which would violate individual rationality for either the highest seller type or the lowest buyer type, which proves the theorem
- (Myerson and Satterthwaite then go on to solve the optimization problem

$$\max_p \{U_b + U_s\}$$

to find the mechanism that maximizes combined payoffs, knowing it won't be efficient; they solve the problem, but the solution is pretty complicated – read the paper if you're interested!)

References

- Jacques Cremer and Richard McLean (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica* 56
- Roger Myerson and Mark Satterthwaite (1983), "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory* 29