Walrasian Markets (Lectures 4-5)

1 Where we are

- The last two classes, we've been building the case that "markets work" –

  or at least, explaining why that's the default position for most economists

- We saw two results suggesting that market outcomes will naturally be Pareto efficient,

  and therefore that we should expect markets to generally yield "good" outcomes

- I want to take one more minute to appreciate the First Welfare Theorem,

  and in some sense, why it's impressive

- a price accomplishes two things

- first, given a fixed quantity of a good, it gets the allocation right:

  with a fixed amount of supply, it ensures that the consumers with the highest value for the good get it;

  and with a fixed amount of demand, it ensures that the firms with the lowest cost to produce supply it

- and second, it gets the production level right:

  if a good is highly valued, this requires a high market-clearing price,

  which encourages more firms to produce it (even if this means redirecting resources from elsewhere)

- both these things – the efficient amount of production, and the efficient allocation of both production and consumption – have to happen for an outcome to be Pareto-efficient

- and what's cool about the first welfare theorem is that it says there are prices at which these things happen simultaneously in every market

- On the other hand, the first welfare theorem, and the Coase theorem, both depend on substantive assumptions –

  a lack of market power by firms,

  the absence of externalities or the absence of transction costs,

  and a lack of informational frictions

- This gives us a starting point to think about what types of markets we think should lead to pretty good outcomes without much intervention,

  and what types of markets will have difficulty getting good outcomes

- Where are we going next?

- This week, we're going to look at a couple of markets that we might reasonably think would fit these assumptions,

  and see how they're designed and how they function

- Specifically, we'll look at two markets that seem at first glance like Walrasian markets –

  the type of markets that fit our model of Walrasian equilibrium reasonably well

- And we'll get to see that even the "simplest" types of markets can have interesting complications

- First, three "odds and ends":

  - commodities
  - markets versus marketplaces
  - and a very brief introduction to game theory

2 Commodities

- One of the key assumptions for the First Welfare Theorem was that firms are price-takers –

  there's a market price they can't influence, and all they can do is decide how much to produce and sell at that price

- In other words, firms have no market power

- For some goods, this seems ridiculous

  - hard to imagine that Apple has no power over iPhone pricing, just sees a "market price" and decides how many to produce and sell at that price

- For some goods, though, this is reasonable –

  particularly for goods which are very standardized, so lots of firms can produce identical versions,

  and buyers don't care which producer they buy from

- Goods like these are called *commodities*

  - Roth contrasts commodity markets with *matching markets*, where buyers care who they're buying from

- Roth, in his book Who Gets What And Why that I asked you to buy,

  makes the point that goods aren't always naturally commodities,

  but are sometimes deliberately *turned into* commodities,

  and that this can have big benefits for markets

- He writes (p. 15):

  "You probably don't know where you bread was baked – but even if you do, your baker doesn't have to know who grew the wheat that went into the flour used to make the bread. That's because wheat is traded as a *commodity* – that is, it is bought and sold in batches that can all basically be considered the same. That simplifies things, although even commodities need to be designed, so that the market for wheat doesn't have to be a matching market, as it was as recently as the 1800s.

Every field of wheat can be a little different. For that reason, wheat used to be sold "by sample" – that is, buyers would take a sample of the wheat and evaluate it before making an offer to buy. It was a cumbersome process, and it often involved buyers and sellers who had successfully transacted in the past maintaining a relationship with one another. Price alone didn't clear the market, and participants cared whom they were dealing with....

Enter the Chicago Board of Trade, founded in 1848 and sitting in the terminus of all those boxcars full of grain arriving in Chicago from the farms of the Great Plains.

The Chicago Board of Trade made wheat into a commodity by classifying it on the basis of its quality (number 1 being the best) and type (winter or spring, hard or soft, red or white). This meant that the railroads could mix wheat of the same grade and type instead of keeping each farmer's crop segregated during shipping. It also meant that over time, buyers would learn to rely on the grading system and buy their wheat without having to inspect it first and to know whom they were buying from."

- So a standard grading system made wheat easier to buy –

  it *lowered transaction costs* by reducing the need for each buyer to inspect a seller's wheat before offering a price,

  and it made the market thicker by allowing a buyer to buy from any seller with a particular grade of wheat, not just sellers the buyer had already learned to trust

- Roth goes on to make the point that commoditization can even improve the products that are available, not just make the already-available products easier to buy

- He writes (p. 17):

  "Coffee beans have been grown in Ethiopia for centuries, but until the twenty-first century they were traded a lot like nineteenth-century American wheat. If you wanted to buy Ethiopian coffee in bulk at the source, you had to have an agent there who could extract a sample from deep inside each sack to taste and evaluate it.

  That changed in 2008 with the creation of the Ethiopia Commodity Exchange. At its heart is a system of anonymous coffee grading, in which professional tasters sample and grade each lot put up for sale. (By the way, there was also some thoughful market design that went into the rules... involved in organizing quality grading. For example, tasting must be "blind": the tasters can't know whose beans they're tasting. Otherwise they could be bribed by the seller to inflate the grades.)

The standardization of coffee can actually improve the quality of the coffee harvest. Coffee beans grow inside a "cherry," and the best coffee is harvested when the cherry is ripe and red. But the beans are sold after being removed from the cherry and dried. So when buyers simply see coffee beans, they can't tell whether they were harvested from ripe red cherries or from unripe green ones. Before coffee was graded, coffee farmers sometimes were tempted to harvest a whole hillside at once, red and green beans, ripe and unripe. But now that tasters can tell the difference, it makes sense to have coffee pickers pluck only the red cherries and to come back later to harvest the rest of them when they are ripe. Since the graders can tell the difference, the market reliably rewards such care with a higher grade and a higher price. The ultimate result is that foreign buyers can now buy Ethiopian coffee beans in bulk from a distance, without having to taste them on the spot, and from multiple sellers, without worrying about the sellers' reputation or pedigree."

- A big part of Roth's point is that markets that work well don't happen by accident –
  they work well because they were consciously planned
- Turning idiosyncratic goods into commodities –
  allowing them to be traded in thicker markets, with lower transaction costs,
  and possibly even creating incentives for higher-quality products –
  may make a market work well,
  but it's work – someone has to actually design the process
- (Someone had to set up the Chicago Board of Trade, and the Ethiopia Commodity Exchange)
- "Commodity goods" may be the goods that best fit the simple model of Walrasian exchange,
  but these goods are commodities because someone chose to make them commodities

3 Markets vs Marketplaces

- One other distinction I want to make
- The distinction is captured well in Ed Glaeser's review of Roth's book[1]
- Glaeser writes:

  "[Roth] presents a menagerie of cases in which markets go awry, but he describes these failures as an engineer, not an economist. The engineering approach sees a failure when a market screws up by allocating goods to the wrong person. For example, if a price control means that rationed goods are allocated to people with low demand, then the engineer sees a market failure. To an economist, price controls aren't market failures. They are occasionally unwise regulations that stop markets from doing their job.

  This distinction is irrelevant for a market designer who is trying to fix a situation on the ground. The market designer doesn't really care why the marketplace is screwed up, but rather how the marketplace can be fixed. But the distinction is quite material to an economist who is trying to draw any larger lessons from the failure. If the failure reflects a deep flaw in a free market, then this points to more regulation. If the failure reflects a perverse regulation, then this points to less regulation."

- A lot of the distinction, I think, is that economists think about **markets**,

  while market design people like Roth think about **marketplaces**

- A market is the general process by which a good gets allocated;

  a marketplace is a particular institution, location, set of rules,

  by which a particular set of buyers and sellers try to allocate that good

[1]Edward Glaeser (2018), "A Review Essay on Alvin Roth's *Who Gets What – and Why*, *Journal of Economic Literature* 55.4.

- The market for wheat determines how much wheat is grown globally
  and how it gets turned into bread
- A particular marketplace for wheat determines how Midwestern wheat gets shipped to Chicago, graded, and then bought by supermarkets to distribute to their stores

- We can talk abstractly about the market for labor,
  and what will happen to it as more and more jobs get automated;
  but when I think about looking for a different job,
  I'm not competing with all workers in the world, or considering all jobs;
  I'm thinking about, say, globally top-100 research universities with econ departments,
  and the process by which they hire professors

- So the Coase Theorem, and the First Welfare Theorem, are statements about markets –
  big, abstract, loosely-defined allocation devices
- While the "real" markets we'll be looking at are really examples of *marketplaces* –
  specific institutions by which a particular set of goods are allocated to a particular set of interested parties

- We know to expect a free market could fail if there are large externalities,
  or widespread informational problems;
  but even without these, with products which "should" be allocated well by markets,
  we could have particular marketplaces that are poorly designed, or crippled by bad regulations, and which function quite poorly
- Roth is largely interested in fixing the latter; Glaeser is more interested in the former
- In this class, we're interested in both
  - in taking insights we have from economics about markets in general,
  - and seeing how they play out in particular marketplaces

4 The briefest possible intro to game theory

- This isn't primarily a course in game theory,

  but we're going to use some game theory from time to time,

  so I'll occasionally be teaching you some

- Today, a very brief intro to game theory,

  with a focus on static games, which I'll define in a minute

- Game theory is basically a language to formalize how we're thinking about a strategic interaction,

  allowing us to take a complex interaction and analyze it mathematically

- It's useful to think of it as both art and science –

  the science coming once the game has been specified and you're analyzing it,

  the art coming in the first stage, when we figure out how to take a real-world interaction and write it down as a game

- Formulating something as a game requires us to specify three things:

  1. who the **players** are
     (could be two, could be a lot, could even be an infinite number,
     but most typically $i \in \{1, 2, \ldots, N\}$)
  2. what **actions** they each are choosing between
     (a set A_i of possible actions for each player – again, could be binary, finite, infinite...
     and need not be the same for each player)
  3. what **payoffs** they each get, given the actions they choose
     (for each player i, a function $u_i : A_1 \times A_2 \times \cdots \times A_N \rightarrow \mathbb{R}$,
     giving player i's payoff for each combination of everybody's actions)

- In a dynamic game – a game that unfolds over time – we would also need to specify timing, or more precisely, information – who knows what when they make their choices

- But for now, we'll focus on static games – games where all the players act at the same time, so there's no question of who knows what when

- We assume everyone knows everything about the game –

  the players know who the other players are, and what their possible actions and payoffs are –

  but nobody knows what the other players chose to do until after they've acted

- And finally, we assume the players are all rational –

  each player will do whatever is best for him or herself –

  and everyone knows everyone is rational, and everyone knows that, and so on

- In a sense, what game theory lets us do is to specify exactly what's "in" our model of reality and what's outside it

- Any choice I make could potentially affect everyone in the world – 7 billion other people

- And I make lots of choices every day, and so do each of them

- When we write down a model of a strategic situation, we're deciding **which** interactions matter

  we're explicitly saying, "these five people are the ones we're paying attention to,

  this is the decision we're considering,

  and everything else is outside the model"

- And we're being clear about what motivations we're ascribing to each player

- Of course, game theory gets it wrong when the game is specified wrong

- My PhD advisor, who studies markets, likes to say,

  "the game is always larger than you think" –

  sometimes we miss things because there are other decision-makers we forgot to consider,

  or because we oversimplified peoples' motives – their payoffs – and left something out

- But that's a limitation of any model;

  at least we can be clear about where we're drawing the line of what's in and what's out,

  and if we discover later we drew it in the wrong place,

  we can change the way we're modeling the game

- Now, the "classic" example of a game is the Prisoner's Dilemma
- The story, in brief:
  - two friends commit a crime, they get caught,
    and they're being interrogated by separate police in separate interrogation rooms
  - Each one has to decide whether to keep his mouth shut,
    or confess to the crime, offer to testify against his friend, and hope to make a deal with the DA
  - If they both keep quiet, there's not enough evidence to convict them of a felony,
    but there's enough to get them on a lesser crime, with some light punishment
  - if one confesses, he can make a deal, and walk free,
    while his friend gets punished severly;
  - and if both confess, they'll both be punished pretty severely,
    but with some credit for having cooperated
- if we model this as a game, we have:
- the **players** are player 1 and player 2
- the **strategies** for each are just $A_i = \{\text{stay quiet}, \text{confess}\}$
- the **payoffs** can be represented in lots of ways, but most typical is a *payoff matrix*:

		Player 2	
		Quiet	Confess
Player 1	Quiet	-1, -1	-10, 0
	Confess	0, -10	-5, -5

- (always player 1 along the side, 2 across the top,
  and always player 1's payoff first in each box)

- How do we analyze games like this?
- We look for Nash equilibria
- A Nash equilibrium is a plan for each player, $a^* = (a_1^*, a_2^*, \ldots, a_N^*)$,

  such that given what everyone else is doing, player i can't do any better than playing a_i^*,

  that is, such that for every player i,

$$u_i(a_i^*, a_{-i}^*) \quad \geq \quad u_i(a_i', a_{-i}^*)$$

  where a_{-i}^* is the non-i elements of a^* and a_i' is any strategy in A_i

- or put another way, a Nash equilibrium is a plan for everyone

  such that no player has a *profitable deviation* –

  no player can improve their own payoff by switching to a different strategy,

  given what everyone else is doing

- for a game in matrix form, we can look for equilibria by contingent reasoning:

  each player considers his best response to each possible action by the other player, and we circle payoffs corresponding to best responses;

  then any cell with both payoffs circled is a Nash equilibrium

- hence the name "prisoner's dilemma" –

  game theory predicts both confess,

  even though that's Pareto-dominated by both keeping quiet

- (lots of strategic interactions have this flavor – we could potentially gain by cooperating, but each of us would have a private incentive to "defect")

5 Continuous limit order books

- With all that out of the way, let's talk about an actual market
- Specifically, we'll talk about how stocks are traded in the U.S.

- This is the New York Stock Exchange
- Stocks are shares of actual companies –

  if I own them, I'm entitled to a share of the company's profits, which I receive as dividends,

  and I also have a share of the control rights – I can vote on decisions the firm has to make
- U.S. stocks trade in multiple locations, but the two largest are the NYSE and NASDAQ;

  the firms whose shares trade at the NYSE are worth a combined $20 trillion,

  and the daily volume of trades is in the hundreds of billions of dollars
- So, this is a big market

- Trade on these markets work via what's called a Limit Order Book
- Traders submit limit orders, which consist of three pieces of information (beyond which stock they want to trade):

  whether they're buying or selling,

  how many shares,

  and the price at which they're willing to trade
- So if we're talking about Facebook stock, I might place an order to buy 10 shares at a price of $165 or less per share
- Now, if someone had already placed an order to sell at that price (or less), I would get to buy immediately;

  and if not, my order would be stored in the order book, awaiting someone else showing up with an order to sell at a compatible price

- EXAMPLE of what the order book would look like
  - start with an empty book
  - I place a buy order (a "bid") for 10 shares at $165
  - Since there are no standing sell orders ("asks"), my order goes into the book
  - Someone else comes along and wants to buy 30 at $163
  - Someone comes along and wants to sell 50 at $170
  - Someone else wants to buy 30 also at $163

  - Now someone comes along and places an order to sell 50 at $162
  - They get to trade immediately:
    they sell me my 10 shares at $165,
    and then they sell 40 more shares at $163
    (30 to the person who placed their order first, and the remaining 10 to the person who placed their order second)

  - so now there's a standing buy order for 20 more shares at $163,
    and a standing sell order for 50 at $170
  - someone comes along and wants to sell 50 at $160
    they sell 20 to the standing buy order at $163,
    and the rest of their order goes into the book – 30 more at $160
- and all of this is happening in continuous time – orders are processed immediately as they come in
- orders can be cancelled at any time, but they stay active until they either trade or get cancelled by the trader who submitted it

- now in practice, for frequently-traded stocks, there are lots of buyers and lots of sellers, and lots of existing buy and sell orders in the book at a time
- here's an example of what a limit order book might look like, with price on the x axis
- the price in the middle between the lowest ask and the highest bid is called the "mid-price"
- all the bids are below that, and all the asks are above that – or else they would have already traded
- and the distance between the highest bid and the lowest ask is called the "bid-ask spread",

  and is one measure of how "liquid" a market is

- Now, the market could work if each time someone wanted to buy, they had to wait to find someone else who specifically wanted to sell
- But in fact, there are lots of big traders – investment banks – who are willing to both buy and sell, as long as they can get good prices in both directions
- So they might place both buy and sell orders – offering to buy at $164, or to sell at $166
- (Actually, spreads are much smaller than that...)
- They do this because lots of traders show up at the market with a pretty urgent need to trade,

  and will just take whatever is the best deal waiting for them
- So if the asset is really worth $165, the bank feels they can make a profit either buying a little below that, or selling a little above that, and is happy to do either

- Now, whichever bank is offering to both buy and sell is acting as a "market maker", or "providing liquidity"
- They're doing this to make a profit – by buying lower, and selling higher
- But they're also taking on some risk to do it
- Suppose some bank has placed orders to buy at $164 and to sell at $166,

  and bad news comes out about Facebook –

  user growth was slower than expected, or they did something really slimy with peoples' private information

  and everyone wants to sell their Facebook stock
- So a bunch of sell orders come in, the bank's $164 buy order gets hit, and the price drops to $158

- By providing liquidity, I expect to earn money when a stock keeps trading in a narrow range, because I'm buying cheaper than I'm selling,

  but I expect to lose money when the stock moves significantly and I end up on the wrong end

- That's why I can get away with charging a positive spread

- (If there was no risk, someone else would be willing to provide liquidity at a narrower spread, and the spread would get competed down to basically zero)

- so in a limit order book, there are standing offers to buy and sell at different prices

- and if news comes out that makes a stock worth more or less money, whoever has the news first – and anticipates the stock price is going to go down, or go up – can make money by trading first, hitting those existing "stale" quotes

- a recent paper talks about a particular way that this can happen:

  when the "news" about a stock's value is actually the price of the stock somewhere else

- lots of stocks trade on multiple exchanges

- for example, on the New York Stock Exchange and the Chicago Mercantile Exchange

- of course, the prices are typically the same – if a stock were cheap on one and expensive on the other, there would be free money to buy on one and sell on the other

- but suppose a couple of eager buyers show up in Chicago, pushing the price up, and the order book in New York hasn't yet adjusted

- if I'm the first person to notice, I can buy at the old price in New York, and sell at the new higher price in Chicago, and get free money

- this is called an *arbitrage* – basically a risk-free way to make money trading

- but in order to do it, I need to do it before someone else does –

  before someone else snaps up all the low sell orders in New York and the price adjusts

- so how fast do I need to be?
- basically, I need to be fastest –

  I need to see the price change in Chicago,

  and get my buy order to the New York exchange before anyone else does, and before the person who submitted the old sell order cancels it
- so there's profits to be made if I'm the fastest

- The paper – Budish, Cramton and Shin, "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response" – focuses on one particular security, that's not even a single stock, but an exchange-traded security tied to the value of the S&P 500 index
- the S&P 500 is a market index – basically, the average price of 500 large U.S. companies
- it's a common measure of the U.S. market as a whole, and there are lots of mutual funds that try to match its performance
- and for the last several years, there are individual securities – which trade like stocks – whose value is tied to the level of the index
- one such security trades in New York, and one in Chicago
- the paper uses detailed data from the New York and Chicago exchanges,

  and finds that on an average day, there are 800 arbitrage opportunities in this one security –

  800 times when at least for a moment, the highest bid in one market is above the lowest ask in the other,

  so I could make free money buying low and selling high at the same time
- the average arbitrage would earn profits of about $98
- so I could earn about $80,000 a day in free money,

  if I could get a message between Chicago and New York a little faster than everyone else
- and that's just in this one security – obviously, there are similar opportunities in any stock that's traded in both markets

- so how fast do I need to be?
- well, traders in each market get a market feed that keeps them up-to-date in real time on the orders in the book on each exchange
- and you obviously wouldn't have to call up your friend in New York and tell him to make a trade, you would have software that would submit the order automatically
- so the limiting factor is literally the time it takes to get a message from Chicago to New York
- if I can get a message from Chicago to New York before everyone else, I'll be able to trade at the old prices, and make money

- Chicago to New York is a little under 800 miles
- Prior to 2010, the fastest way to get a message from Chicago to New York was on dedicated fiber optic lines, which ran along railroad tracks, around mountains, and so on, and got a message through in 16 milliseconds
- In 2010, Spread Networks finished construcction of a more direct line – still fiber optic, but straighter,

  which cost $300,000,000 to build,

  and lowered transmission time from 16 to 13 milliseconds
- And every major trading firm knew it had to pay them to transmit their signals: quoting the article,

  "Industry observers remarket that 3 milliseconds is an "eternity" to high-frequency trading firms, and that "anybody pinging both markets has to be on this line, or they're dead"
- In fact, Spread Networks' new line would not be state of the art for long:

  since light moves faster through air than through fiber-optic cables,

  new microwave networks that transmit signals through air are now faster, cutting transmission time first to 10 ms, then 9, and currently about 8

- The paper shows that even if you're just as fast as everyone else,

  your stale orders are likely to get "picked off" when the market shifts –

  because your message to cancel your stale order is competing with other orders to trade with it,

  so if there are N traders who are equally fast and you were the one providing liquidity,

  with probability $\frac{N-1}{N}$, since orders are processed in the order received, someone gets you

- to compensate for this risk, you have to charge a bid-ask spread – you need to make a profit the rest of the time, to compensate for the money you lose to these arbitrages

- they point out that investing in speed is a prisoner's dilemma

- suppose there were just two trading firms, and if they were equally fast – both could transmit at 16 ms, or both could transmit at 13 ms – they would both earn profits of \$200 MM

- investing in the speed upgrade costs \$50 MM

- but if one invests and one doesn't,

  the faster firm gets \$300 MM in profits to the slower firm's \$100 MM

		Trading Firm 2	
		16 ms	13 ms
Trading Firm 1	16 ms	\$200,000,000, \$200,000,000	\$100,000,000, \$250,000,000
	13 ms	\$250,000,000, \$100,000,000	\$150,000,000, \$150,000,000

- it's better for the two firms if they don't pay the extra to invest in speed –

  but the equilibrium is for both of them to invest

- hence, a socially wasteful arms race

- (of course, it could be good for others – for market liquidity, or whatever –

  but 16 to 13 ms is such a small improvement, hard to see how it significantly benefits the real economy...)

How to fix it...

- in the paper, the authors propose a fix
- they point out that these arbitrage opportunities are an inevitable cost of trading in continuous time – that is, processing each order immediately as it arrives
- (They show that as speed has increased, these opportunities haven't gone away –

  they've gotten shorter and shorter, as they get snapped up in less and less time,

  but how often they're available, and their size, has stayed fairly constant)

- they propose instead that the exchanges aggregate orders that come in in a brief interval, and then clear the market with a "batch auction" using all those orders
- that is, for some period of time, don't make any trades, just accumulate all the orders that come in;

  then after a little while, take all those orders, create a supply and demand curve,

  and do the trades that clear that market
- the orders that don't get filled then stay in the book to start the next period
- they point out that trade could still be incredibly fast – faster than humans could process anyway
- but if we slowed it down to trade in, say, 100 millisecond intervals –

  trade happens once every tenth of a second, batch processing all the orders that arrived in the previous tenth of a second –

  the prisoner's dilemma problem would go away,

  because there would be less gain to being a little bit faster than everyone else,

  for two reasons

- first, if we're all pretty fast, then most of the time, I can cancel my standing order before someone gets to trade against it
- that is, suppose we all take 10 ms to get a message through, and trade happens every 100
- if news comes in (the price changes in Chicago), then nine times out of ten, I'll still be able to cancel my order before trade happens;

  and if I don't, then none of the other firms have time to get a new order in before trade happens anyway
- this is good because it reduces my risk from providing liquidity,

  which means I can charge a lower bid-ask spread, or offer more shares for sale –

  I can make the market more liquid, since I'm at less risk of getting "sniped"

- but they go on to point out, I'm also protected even if I'm much slower than anything else,

  because the batch processing would prioritize trade by price, not time
- suppose I'm very slow – I take a whole second to make a trade,

  and I've placed a limit order to buy at $160 in New York, and the price drops to $150 in Chicago
- with continuous time trading, whoever is fastest can sell to me at $160, and I get burned
- with batch processing, I'm still exposed at $160, but all the fast firms compete with each other to be the one to trade with me
- if they all submit orders to sell at $160, only one of them will get chosen,

  so one of them instead submits an order at $159,

  so another one of them submits an order to sell at $158, and so on,

  so competition between the fast traders leads to me actually trading closer to $150
- (SHOW THE MECHANICS)

- so, the paper proposes a switch, from continuous-time trading, to very frequent batch auctions