

**JONATHAN
GRUBER:**

All right, so we finished the first unit of the course, or consumer theory, and we've sort of gotten to the demand curve. Now we move onto the second year of the course, which is producer theory, and talk about where the supply curve comes from. Now the good news is that a lot of the tools and skills we developed in the first few lectures will translate quite nicely to thinking about the supply curve and production.

The bad news is, supply is a lot harder. It's a lot harder, fundamentally, the big picture, because we did consumer theory. We sort of told you what your income was. We told you income and prices, and then we said, OK, here's how to optimize. With firms, they sort of get to decide what their income is. They get to decide how much they produce, and that means we're going to need an extra whole constraint we're going need to model, an extra other part of the process.

So it's going to be an extra step with producer theory, but a lot of the tools will be the same. So let's dive in. We're going to start by talking about, just as consumers have a utility function, producers have a production function. So the first parallel is that producers have production function.

Now here, their goal is not to maximize production. Their goal is to maximize profits. So as consumers want to maximize utility, producers want to maximize profits, which equals revenues minus costs. So the goal of a producer is to maximize profits, which equals revenues minus costs. And what that's going to mean, it's going to mean producing goods as efficiently as possible, maximizing your profits.

We are going to focus for the first few lectures on the cost part. In particular, we're going to focus on maximizing profits through minimizing costs. And we minimize costs by producing as efficiently as possible, OK. And that's what we'll focus on in the next few lectures. Now, what firms can produce comes from their production function.

A production function is of the general form q -- that's units of goods produced-- is a function of the amount of labor input and capital input used by the firm, so q , little q -- let me just highlight right here. I will hopefully get this right. I never in the

semester have gotten it totally right. Little q refers to a firm. Big Q refers to a market, OK. We're going to try to keep this straight.

So little q means a firm's production function. Big Q means a market production function, OK. So if I get that wrong, I'm sure you guys will tell me. OK, so basically what a production function does is it converts inputs, which are labor and capital, into output through some function, just like utility function converts goods into happiness through some function.

It's the same idea here, but here, it's more tangible. Unlike utils, output can actually be measured. So literally, it's not some preference mapping. It's literally a technological function. OK, you get your hands around it more than a utility function. It's literally a technological function by which inputs get converted to an output.

Now, we call these inputs the factors of production, OK. Labor and capital, we call the factors of production. They're the inputs they get used to produce things. Now, what are labor and capital? Labor is pretty easy. Labor is workers, OK, either number of workers or hours of work. we'll use those interchangeably, but the bottom line is, labor is workers. It's you all.

OK, that's sort of the easy part. Capital is harder. Capital is machines land, building, all the stuff that workers use to make things, OK. So capital's a vaguer concept. But for now, think of it as like machines and buildings, OK, the stuff that workers use to produce goods. And outputs are the goods and services that got produced.

Now, when we talk about inputs or factors of production, we're going to talk about them being variable or fixed. Variable means changeable, fixed means not changeable, OK. So variable inputs are inputs that can be easily changed, like hours of work. You can easily work. You guys work different amounts of hours every day.

You can easily change the hours that you work. You can pull an all nighter if something's due the next day. You can work less if there's something good on TV, OK. Fixed inputs are those which are harder to change quickly, like the size of a plant. Let's say you're siding a bigger plant. You can't just instantly do that.

It takes a lot of production process. You can see by the giant production going on as

you pass every day as you walk, if you come from east campus, you walk to this building, it's going to take years to build out those new MIT facilities, OK. So it's not simple. So fixed inputs are inputs that are hard to change quickly.

And the key distinction we draw-- and we think about variable fixed-- is the short run versus the long run. And the way we define these is that basically, in the short run, some inputs are fixed and some are variable. In particular, we're going to stay in the short run, labor is variable and capital is fixed. So in the short run, you have labor and then some fixed level of capital, k bar.

So in the short run, you've got some building. You can't change it, but you can always change how hard people work in that building. In long run everything's variable. Labor and capital are both variable, OK, so there's no k bar. Capital's variable in the long run. So the question then is, what is the short and the long run.

Well, there's no good answer to that. Intuitively, think of the short run as a matter of days or weeks or months and the long run as a matter of years or decades, for your own intuition. But technically, the definition is, the long one is the period of time over which all inputs are variable. That's the technical definition. The long run is a period of time over which all inputs are variable. That's our technical definition.

So think about how long it takes to build a plant or make new machines, OK. That's the long run. So I'm never going to ask you, is the short run 8.3 days or 9.7 days, OK. There's no right answer. The right answer is, the technical answer is, the short run is a period of time over which some inputs are fixed and some are variable. The long run is the period of time in which all inputs are variable, OK.

And it's not a clean distinction. Obviously, in reality, there's a whole range of inputs ranging from workers to the gas you pipe in to use for your thing, to the raw materials you have to buy to the machines, the buildings. Obviously, in reality, there's a whole range of variability. But once again, to make life easier, we're going to shrink this down to two dimensions, labor and capital.

Labor is going to always be variable. Capital's going to be fixed in the short run, variable in the long run, OK. So that's how we'll boil this down to make life easy. Yeah.

AUDIENCE: Can you give an example of what capital is again?

**JONATHAN
GRUBER:**

Capital is the buildings, machines, the stuff that workers use to make things. Yeah, OK. Other questions? With these definitions in mind, let's talk about short run production. Let's start by talking about production in the short run. Someone needs to invent me some more indestructible chalk. OK, so let's start in the short run where labor is variable and capital's fixed.

So in the short run production function, q equals f of L and k bar, OK. That's our short run production function. OK, now that means in the short run, the firm's only decision, the firm is given a stock of capital. So think of the short run as, you are hired to manage a plant. And the plant, you don't get to decide if the plant's big or small or what machines. It's there.

Your only decision is how many workers to hire, how many hours of labor to employ. And once again, I'll go back and forth through a number of workers and hours of labor. The bottom line is the amount of labor being provided, OK. The way you're going to decide that is, you are going to look at when we're going to call the marginal product of labor, the marginal product of labor, which is simply the change in output for the next unit of labor input.

This is very much like a marginal utility of a good. Once again, going back to our powerless consumer theory, the marginal utility of pizza was the delta in utility for the next unit of pizza. The marginal product of labor is the delta in the amount produced for the next unit of labor. And we are going to assume, much as we assume diminishing marginal utility, we are going to assume diminishing marginal product.

Now, that's a little less intuitive than diminishing marginal utility was. At least, I hope you found diminishing marginal utility kind of intuitive, that the second slice of pizza would be worth less to you than the first slice of pizza. Hope you found that intuitive, OK. Now, this is less intuitive. Once again, like with utility, it's not saying the next worker doesn't help.

The next worker does help. It's just that the next worker helps less than the previous worker, OK. Now, this isn't true everywhere. Obviously, there are tasks where having

two workers together makes both better, and we'll talk about that later on. But we're going to focus on the range of production where this is true.

Think about production functions as being non-monotonic, they can go up and down. But we're going to focus on the range of production where this is true, where the next worker is not as productive, OK, because eventually that's going to be true for every firm. And why is that going to be true? Because we're holding capital fixed. The reason eventually workers will get less productive is because there are only so many machines and buildings they can work in.

The classic example we think is the example of digging a hole with one shovel. You've got one worker digging a hole. She can do a certain amount of effort. Then a second worker comes along. She can add value. The first can rest. They can trade off, and maybe she's almost as productive, or probably not quite. Then the third worker, and the fourth worker. By time you have six workers, they're mostly just standing around because there's only one shovel.

Now, each one's more productive because they can help you optimize the shift and get rest and stuff, but the truth is, clearly the sixth worker's less productive the fifth worker given there's only one shovel. So the key to understanding the intuition of diminishing Marshall practice to remember that there's a limited amount of capital. There is a fixed amount of capital.

So if there's a given building and you've got 1,000 workers and you try to shove the 1,001st in there, he's not going to do a whole lot of good, OK. So marginal product of labor comes from the notion that there's a fixed amount of capital, so each additional worker does less and less, adds less and less to the production, OK. That's the intuition for diminishing marginal product of labor.

And that's pretty much it for short run production. That's sort of what you've got to know. The more interesting action comes when we go to long run production. That gets more interesting because now, you have an optimization decision over labor versus capital. In the short run, you just decide how many workers to hire. In the long run, now you're back to the kind of utility framework we used.

We had to trade off pizza and cookies. Now, you get to trade off workers and machines. Now you own the firm. You're going to own it forever, OK, and you get to

trade off. You get to think about workers versus machines. So now, you're going to have to make a decision on that. And that decision is going to be, just as your decision of how to trade off cookies and pizza is driven by utility function, your decision about whether to employ workers or machines it's being driven by your production function.

So make life easy. Let's start a production function that looks just like the utility function we're using, q equals L times k . Familiar form. Before, we said how happy pizza and cookies made you. It was the square root of pizza times cookies. Now we're going to say how many goods you can produce is the square root of capital times labor.

OK, and figure 5-1 shows you what that delivers in terms of graphically. Just as to graphically represent utility, we graft indifference curves, to graphically represent production, we are going to graph isoquants. Isoquants are like firm indifference curves, but once again, they're sort of more tangible.

An indifference curve is this weird, intangible idea of points along which are indifferent. And isoquant's a tangible thing. It's the combinations of capital and labor that produce the same amount of output, OK. So for any given production function, there's different combination of capital and labor that produce the same amount of output. So for example, in our example, two units of capital and two units of labor produces two units of output.

Four units of capital and one unit labor also produced two units of output, so they would be on the same isoquant. They're this combinations of inputs that deliver the same level of output, OK. And isoquants, all the stuff we learned about indifference curves apply here. More is better, so further out is better. They can't cross, OK. And they slope downwards.

All the set of things we learned about indifference curves, that same set of intuitions applies here. The difference with production is it's more plausible to have extreme cases, OK. So let's consider two extreme cases. Let's first consider the case of inputs that are perfectly substitutable, so like a Harvard graduate and a Beanie Baby, OK, perfectly substitutable inputs.

Those are goods where the production function would be of the form, q equals L

plus k . That's perfectly substitutable because you're indifferent between the unit value and the unit value. They're perfectly substitutable. So you can see that in figure 5-2a, I do x and y instead of L and k , but it's the same idea. If there's two inputs, x and y , then with that production function, a perfectly substitutable, that would lead to linear isoquants.

Perfectly substitutable inputs will lead to linear isoquants with a slope of minus 1, you're perfectly indifferent between one or the other at all levels. At any point in time, you're indifferent between 1 more unit of x and 1 more unit of y , 1 more unit of labor and 1 unit of capital, OK. At the other extreme would be perfectly non-substitutable inputs, inputs where you can't produce one more unit without one of each input, OK.

That would look like figure 5-2b. We call this a Leontieff production function. A Leontieff production function is one where there's non-substitutable inputs, where the production function is the min of x and y , that one more unit of y does you no good unless you also get one more unit of x . So what's an example? What's a real world example of a Leontieff production function? What's a good which would have non-substitutable inputs? We'd need at least one of each. Yeah.

AUDIENCE: If you have like programmers and computers, [INAUDIBLE].

JONATHAN GRUBER: Programmers and computers, that's a good one.

GRUBER:

AUDIENCE: It's like, you need like a right shoe.

JONATHAN GRUBER: You need a right shoe and a left shoe. That's a classic example I would use. Cereal and cereal boxes stuff like that, you know, stuff where you basically need both. Shoes are the sort of classic example, OK. And that will give you sort of a Leontieff production function, OK. So basically, those extremes help you think about what isoquants are and what they mean.

Now, continuing our parallel to consumer theory, what is the slope of the isoquant? What is the slope of the isoquant? The slope of the isoquant, just as we call the slope of the indifference curve the marginal rate of substitution, since we're not

very creative in economics, we call the slope of the isoquant the marginal rate of technical substitution, because it's the same idea, but now it's technical.

It come from a technical production function, not from your preferences, OK. So marginal rate of technical substitution is the slope of the isoquant, or Δk over ΔL . And as with indifference curves, that slope varies along the isoquant. So we can see that in figure 5-3, OK. Figure 5-3 is once again drawn for our production function, q equals square root of k times L .

So let's say for example, we start with one worker and four machines at point A, OK, and now we consider adding a second worker. Well, at point A, that second worker is so productive because of diminishing marginal products, OK. You already got four machines, only one guy to run them. Like, he's not doing a lot of good. So adding a second worker and two machines helps a lot. It's not perfectly Leontieff, but you can get the intuition that two workers and two machines are the same as one worker and four machines.

You're not better off. You're the same off. You have the same isoquant. So the marginal rate of technical substitution is minus 2. That is, one worker substitutes for two machines, OK, one worker substitutes for two machines. But now, starting from point B and moving to point C, it takes two more workers to substitute for one machine because, if you go down one machine, you need a lot more workers to make up for it. So then the MRTS falls to minus 1/2.

So, going from A to B, one worker makes up for two machines. Going from B to C, it takes two workers to make up for one machine. And that's because of diminishing marginal products, OK? That's because of diminishing marginal products.

OK, indeed, there's a convenient way mathematically to relate the marginal rate of technical substitution to marginal products. Think of what the MRTS is asking. Think of what we're asking along the isoquant. We're saying what combinations of capital and labor yield the same output. That's what we're asking, OK?

So another way to think about it is what change in capital plus an equivalent change in labor leads to the same level of output. So you can ask, well, the change in labor times the marginal product of a unit of labor, times the marginal product of a unit of labor, plus the change in capital times the marginal product of a unit of

capital, which is the same. I didn't define marginal product of capital. It's the same idea as marginal product of labor. It's dq/dk is the marginal product of capital. That equals 0 along an isoquant.

Think about it. Along an isoquant, the next unit of labor times how productive that labor is plus the next unit of capital times how productive that capital is equals 0 because you're staying along an isoquant. So, if you're taking away one unit of labor, if this is minus 1, and this is plus 1, then, based along the isoquant, you're choosing the point where the MPL equals the MPK.

Or, more generally, if you reorganize this, you get that $\Delta k / \Delta L$, which is the slope, equals minus MPL over MPK. And that is the MRTS, OK? The marginal rate of technical substitution is the negative of the ratio of the marginal product of labor and the marginal product of capital.

Once again, should look familiar. It's just like the marginal rate of substitution. It's the negative of the marginal utility of the good on the x-axis for the marginal utility of the good on the y-axis, same idea. I derived it in a slightly different way here, but it's the same idea. And it comes from the notion that you wanted to stay-- that you're staying constant production, as you change labor and capital along this curve. Yeah?

AUDIENCE: [INAUDIBLE] MPK over MPL [INAUDIBLE] give you the same ratio or no?

JONATHAN GRUBER: Well, MPK over MPL would give you-- I mean, basically, we're defining the marginal rate of technical substitution the way we define-- the way we define the marginal rate of substitution. You basically want-- because you want it to be downward sloping. If you define that, the inverse would be upward sloping. So what we're defining is the downward-sloping concept, which is the marginal product of the good on the x-axis and marginal product of the good on the y-axis. So it's not invertible. It's not freely invertible. Yeah?

AUDIENCE: What's the marginal rate of technical substitution for a Leontief production function?

JONATHAN GRUBER: Ah, great question, great question. So, the marginal rate of technical substitution, so let's go back to Leontief. OK, the marginal rate of technical substitution actually

sort of depends on-- it sort of depends on where you are. It's sort of a nonlinear marginal rate of technical substitution, right? So, basically, it's going to very much depend-- depend on where you are. So, basically, it can be negative infinity or positive infinity or 0, depending on where you are on the curve.

So we'll actually-- I don't want to give you more on that because this problem may-- I'm not giving anything away-- could obviously be a problem set problem. So I don't want to give more answers than that away, but, certainly, it's going to be-- it's not going to be constant, OK? Other questions? Yeah?

AUDIENCE: It's just like a line, right?

JONATHAN GRUBER: The marginal rate-- if the curve is just a line, the marginal rate of substitution would be constant. For perfectly substitutable inputs, it would be constant. That's right, just like the marginal rate of substitution would be constant if your indifference curves were linear, OK? Good questions.

OK, so that's production, OK? Other questions about production? OK, that's the basics, and we went fast because, basically, a lot of it's just parallel to what we did with consumer theory, OK? Now but I want to talk about two other aspects of production that we need to keep in mind as we move forward.

The first and the fourth topic for today is returns to scale, returns to scale, OK? This is what returns to scale are asking is what happens to production when you increase all inputs proportionally. So, if you double all inputs or triple all inputs or whatever, cut all inputs by 73%, what happens to production? So it's not about K versus L . It's about a scale, a scaling up or down of the operation, OK?

Now we know, obviously, if you double inputs, production will go up. More is better. The question is by how much. So our baseline we can think about as what we call a constant returns to scale production function. That would be one where f of $2L, 2K$ equals 2 times f of L, K .

So a constant returns to scale function means, if you double inputs, you double output. If you double inputs, you double output. That's a constant returns to scale production function.

But you could also define increasing returns to scale, where doubling inputs leads to

more than double the output, or decreasing returns to scale where doubling the inputs leads to less than double the output. So constant returns to scale means doubling the inputs leads to double the output. Increasing returns to scale means more than doubling the inputs more than doubles the output-- I'm sorry, means double the inputs more than doubles the output. Decreasing returns the scale means doubling the inputs less than doubles the output, OK? And that gives you-- that's your definition of returns to scale.

Now where could these come from? So increasing returns to scale, for example, where could increasing-- that's the world's worst S. Where could increasing returns to scale come from, OK? So, for example, one reason for increasing returns to scale is that, basically, as a firm gets bigger, it might learn to specialize.

So maybe a firm with two workers and two computers, and you double, and you get four workers and four computers, and then you could specialize the tasks more. And each worker is more efficient in their specialized task. That could lead to increasing returns to scale. That's an example of something that could lead to increasing returns to scale.

Decreasing returns to scale could come through something like difficulty of coordination. Maybe, when I've got two workers and two computers, I can keep an eye on them and make sure they don't slack off. But, with four workers and four computers, there's more slacking off because I can't keep an eye on them all the time and more so with 8 and 16, et cetera. Yeah?

AUDIENCE: So, when-- I have to ask, why is doubling the inputs greater than two times the outputs?

JONATHAN GRUBER: Yeah, so, basically, doubling the input-- so, when I move to two workers and two computers to four workers and four computers, I more than double my output. And that's because maybe they specialize and get more productive.

AUDIENCE: Oh, so f of L, K equals the original output.

JONATHAN GRUBER: Yeah, so, well, it's one function. f of L , it's literally one function. It's literally-- so I'll write it out. It's literally saying doubling my inputs leads to more than twice what I get with just-- without doubling the outputs, OK? Is that another way to think about

it? Yeah?

AUDIENCE: So, when we're talking about the returns of our-- our return to scale, is that how much product is being produced or how much profit is being made?

JONATHAN GRUBER: How much product. We're only-- we haven't gotten to profit yet. We're only talking about quantity. We're only talking about quantity. f is-- remember, f is the function that translates inputs to q . Yeah?

AUDIENCE: Are things intrinsically like increasing return to scale functions or decreasing return to scale functions?

JONATHAN GRUBER: Well, great question. So what do you think? What's the right answer? What do we think in reality?

AUDIENCE: I mean, like, perhaps, maybe there could be ways to shift it or not. Like, going back to the whole example you gave about decreasing-- like, if there's more computers, people can start slacking off-- if he set like-- if you set some parameters or like you deactivated social media on those computers that they couldn't go on Facebook when like you weren't watching them, you could make them be more productive per se.

JONATHAN GRUBER: Well, it's a great question. Let's start by looking at figure 5-4 and show some examples of what we think about like decreasing, increasing returns to scale. So figure 5-4 has some examples of the kind of industries people think are potentially decreasing, increasing returns to scale, OK?

So, for example, we think the production of tobacco is a decreasing returns to scale activity. That is you're kind of farming tobacco. You're growing it. You're producing it.

Then, if you kind of double it up, there's still a certain amount of land you're working on. You can't-- there's still sort of a certain amount of crop. You're not going to produce twice as much by having twice as many threshers and workers, whereas, maybe something like producing primary metal, OK, you could basically maybe work a lot more efficiently by having more machines and more workers together producing that metal.

So what is the right answer? The right answer is we don't know, but the one thing we do know is there can never be forever increasing returns to scale. And why is that? Well, at least, we used to think this maybe 15 years ago. Why is that? Why can there-- what would happen in an economy if a firm had forever increasing returns to scale? Yeah?

AUDIENCE: You'd get a monopoly.

JONATHAN GRUBER: It would own the economy because, the bigger I got, the more productive I'd get. So I would just eventually grow and own the whole economy. Now, actually, that may be happening. So maybe it's not as weird as we thought it was 15-- maybe Google and the big five have increasing returns to scale.

But, eventually, we think returns to scale must decrease. We think your scale of production must get so unwieldy that doubling it means you just can't manage it as effectively. Eventually, we think returns to scale must decrease. That's sort of the one sort of principle we have that we don't know-- we think, generally, probably, in the life cycle of firms, returns to scale are probably increasing and then decreasing.

But we don't know where it happens. And, certainly, companies like Google and Amazon are showing us that point of decreasing may happen a lot later than we thought, OK? And that's because I think what we didn't account for in our traditional producer theory is networks, the fact that networks get ever more productive.

We always thought about buildings and workers, and there's a limit to how productive they can get. But networks, by bringing in more and more people, can get ever more productive. But, at some point, we think these things have to decrease. At least, we traditionally thought so, but maybe, in 10 years when Google owns everything, I'll change my tune, OK? But that's sort of the one sort of rule of thumb we have in thinking about this, OK? Other questions about that?

OK, let's talk about the last topic then, which is productivity, how this stuff all matters in the real world. So we're going to come back next lecture and come back to maximizing profits and all that stuff, but I want to sort of step aside now and ask why does this all matter.

And, to do so, let's step way back to the original dismal scientist, Thomas Malthus.

Thomas Malthus in 1798 wrote a book, which said-- which was really pretty depressing. He said, look, let's think about how basic economics works. Now he didn't do the math. This is pre-math. So let's get the basic intuition.

Think about the production of food. OK, the production of food has two inputs, labor and land. There's workers, and then-- you know, there's machines, but they're pretty simple machines. OK, you sort of till the land, and there's land.

Well, in the long run-- in the short run, labor is variable. In the long run, labor is variable. But land is never variable. Land is a forever fixed input. There's no long run. Unless we discover a new planet, there's no long run over which land is variable.

What that means is that there will be ever diminishing marginal product to farming. He didn't say it this way, but this was sort of his intuition that more and more workers will try to cram on a given acre of land. Each additional worker can only do so much. And, eventually, the marginal product will be diminishing, OK?

The result is that productivity will fall, the marginal product of labor, when each additional worker will be less and less. And, as a result, we'll starve because, basically, we have all these people looking for work. There's nothing to do because only a certain amount of land. They won't have anything to do, and, eventually, they'll starve.

So Malthus actually predicted we would see cycles of mass starvation through history, fun guy to have at a party. OK, he'd basically say we're going to get overpopulated. These guys will have nothing to do because there's only so much land they can work on. They'll die off. We'll eventually grow overpopulated again. We'll get these cycles of mass starvation. That was his prediction.

Now, since he wrote that book, world population has increased about 1,000%. And, yet, we're fatter than ever. I'm not saying food deprivation isn't a problem around the world, but, certainly, the world is much better fed that it was in 1798, OK? What did Malthus miss? What did Malthus miss? What did we get wrong? Yeah?

AUDIENCE: The classic example against it is like he didn't account for innovations.

JONATHAN He didn't account for innovation or what we call productivity-- productivity, or you

GRUBER:

can also call it innovation-- and neither have we so far. We have written production functions of the form q equals f of L and K , but, in reality, the production function is actually q equals A times f of L and K , maybe A of t , A sub t .

And that's a production factor that, basically, for a given amount of labor and capital, as you get more productive, you can produce more things. The production function itself changes. You get more productive over time, OK?

In agriculture, how did we do this? Well, we did it in lots of ways. We invented cool new ways to harvest the crop, tractors. We invented fertilizer, chemical fertilizer. We invented seed-resistant crops. We invented lots of things that Malthus didn't see coming.

So, as a result, even though the land is just as fixed as it was in Malthus' time-- we still haven't discovered a new planet we can farm on-- we produce a lot more food because of the factor A . The production function itself has changed. We've become more productive.

So productivity is the factor that allows us-- or innovation is the factor that allows us to produce more and more with a given amount of inputs. So, actually, food consumption per capita is rising. Since 1950, food consumption per person in the world is up 40%, OK? So, while we have starvation, and it's terrible, it's up.

One side note, some of you may have heard of a very famous economist named Amartya Sen. He's a Nobel Prize winning economist. His biggest-- one of his main contributions was he studied famines, and he said famines are not a technological problem. He said there's never been in history a famine in a democracy. No democratic nation has ever had a famine.

Famines are not about technology. Famines are about politics and corruption and the things that get in the way of proper food distribution. So really we have enough food, OK? The food is there. Malthus was wrong.

This is not just true in agriculture. It's true all over the world. Let's look at car production, one of most famous examples, OK? Cars have been around since the late 1800s, OK?

And they're basically-- when cars were first invented, they were essentially craftsmanship. Someone would sit down and make a car if you can believe it. They'd literally make all the parts. They'd make a car or a couple people together.

In the early 1900s, Henry Ford introduced the idea of mass production-- it seems sensible now, but it's not the way they used to do it-- a series of workers who each did a discrete task along the way, constructing a car. So no worker did a whole bunch of the car. Each worker did a little piece, which massively led to increasing returns to scale by specialization.

He did that, and, basically, this was radical at the time. It seems obvious to us now, but he cut the price of building a car more than in half almost overnight and basically wiped out all his rivals through the introduction of mass production. Now you might think that's pretty cool, but, you know, that's an old-time story. But it's not over. Innovation in car production continues.

The Indian company Tata, you may have heard of them. They do a lot of MIT-- they finance a lot of stuff at MIT. They have a car called the Nano that they produce for \$2,500, OK?

It's a tiny car. It's lighter. They use extra light materials. It's smaller because they do things like putting wheels on the extreme outside of the car, rather than sort of underneath the car, OK? And they minimize the parts that are used to make it easily fixed and interchangeable with other cars.

So innovation is going on all the time. Look at hybrid. Look at the innovation in the fuel space with hybrid cars and electric cars and Tesla. OK, innovation is happening all the time.

Now what's key about this, besides the fact, technically, what this means is that, when you write-- when we think about production-- now we're not going to talk about this a lot. We'll assume that there's constant production functions. But what that means technically is, when we think about over time production, innovation is a key factor.

But what this means, in terms of all of us sitting in this room, is that productivity innovation is fundamentally what determines the standard of living in a country. Our

standard of living is determined by productivity, OK? So, basically, if you think about us as workers, if we're going to get richer, we're going to have to make more stuff, OK? We're going to have to make more q or more valuable q , OK?

Now, given our amount of labor, that's either going to happen through more K , through more capital, or through a faster A , through faster innovation. So, ultimately, what determines our standard of living-- that is what determines how much shit we have for a given amount of work-- is going to be how much we save and how innovative we are. I'm sorry, back, how much capital we have and how innovative we are.

Capital it turns out is going to come from savings. I sort of cheated there. We're going to talk about that in about-- about maybe 12 lectures from now. We'll talk about where capital comes from. The hint is capital comes from how much we save, and I'll explain why that is.

But our standard of living is determined of how much capital we have, which is a function of how much we save, but it's mostly determined by how innovative we are, how productive we are, how much more we can produce for a given level of inputs, OK? And, if you look at-- if you ask how does production go up given an amount of capital and labor, we call that total factor productivity. That is, conditioning on all the factors, how much does productivity go up?

Now it turns out we have seen a massive shift in productivity in the US. From 1947, after World War II, to 1973, productivity growth in the US was very rapid, about 2 and 1/2% a year. What that meant-- let's think about what that meant.

That meant, not doing anything else, working just as hard as we were working, we could get 2 and 1/2% more stuff every year, OK? That's what I mean by our standard of living. Literally, it's saying, working the same 40-hour week, every year, we got 2 and 1/2% more stuff, OK?

However, from 1973 until the early 1990s, productivity growth slowed down massively down to about 1% a year. It dropped massively, OK? Now what happened? Well, one thing that happened is we started saving less. K went down. K is driven by savings, and we started saving less. We save a lot less than other nations.

But, in fact, that's not much of it because, even though we don't save much, productivity jumped again in the-- about from 1995 to 2005, productivity jumped again and went up again to about 2 and 1/2% a year. And, essentially, we felt, ah, this is the IT boom.

OK, computers were around since the 1970s. And, throughout the late 1980s and early 1990s, people kept saying where's the productivity gain from computers. And it appeared to show up in the mid 1990s. Suddenly, things got more productive in the mid 1990s to the mid 2000s, OK? Productivity rose to about 2.3% a year.

But, much to our chagrin, productivity has stopped growing rapidly, and it's back to about 1 and 1/2% a year. So we're not as slow as we were. So we were-- so, from 1947, '47 to '73, we grew at about 2 and 1/2% a year. '73 to '95, it was about 1% a year. So that meant, with the same amount of work, we only got 1% more stuff, OK?

'95 to '05, we went up to about 2.3%. We jumped back up. But, since '05, we're down at about 1.5%, so better than we were at our minimum, but not nearly as high as we were at our peak, OK? So, basically, this raises three key questions. Yeah?

AUDIENCE: How is that productivity measured?

JONATHAN GRUBER: Oh, great question. So, basically, we look at, essentially, a way-- roughly speaking, we look at how much stuff gets produced given how many hours of labor there are submitted to the economy. Roughly speaking, we say how much do people work. How much stuff gets made? Boom, that's productivity, nothing super fancy. Yeah?

AUDIENCE: I might have missed it, but what does TFP stand for?

JONATHAN GRUBER: Total factor productivity. That's productivity controlling for capital. But the productivity numbers here are not total factor productivity. They're just labor productivity, allowing capital to change. Yeah?

AUDIENCE: You said that K would decrease when people save less. But, if you save less, isn't your spending someone else's likely [INAUDIBLE]?

JONATHAN GRUBER: You know what? I don't want to go there. We're going to spend a whole lecture on that. So I don't want to go there. K depends on savings. Just take that as a given for now, and we'll come back to that. We'll spend two lectures on it actually, OK?

Now I want to raise three questions, before we go, about these facts, OK? The first question is why didn't the IT revolution and the computer revolution lead to longer lasting productivity gains? Why did productivity slow back down after 2005, OK?

We don't really know. Folks thought that computers would be the next Industrial Revolution. This was going to be a-- this was going to transform our lives, OK? It looks like what it mostly did is transform how we watch porn, OK?

And, basically, it looks like, in terms of productivity, it did not actually change things that much. And we don't quite know why, but it is still a bit worrisome that, in terms of the long run, that, in some sense, there wasn't longer lasting gains from innovation. If it's a question about porn, I'm not answering it.

[LAUGHTER]

AUDIENCE:

Maybe with like how, if you pull more people into like a team, working on team projects, the rate at which the project is worked on tends to decrease.

JONATHAN GRUBER:

You know, there's lots of theories. We could hypothesize all day about why it is. I'm just going to state the facts and say that it's disappointing. And we need to figure out what to do about it.

The second question this raises is how do we spend increases in productivity. What do I mean by that? What I mean by that is, if there's an increase in productivity of 2 and 1/2%, that means we have 2 and 1/2% more stuff for the same amount of work. But why do we have the same amount of work?

Another way to say that is we can work 2 and 1/2% less and have the same amount of stuff, roughly speaking, OK? So I have assumed we work the same, and we get 2 and 1/2% more stuff, but why is that the right answer, OK?

And, in fact, the US and Europe, since World War II, have taken very different paths in this dimension. In the US, we've taken all our productivity and put it into cooler stuff, and we work harder than ever. In Europe, they work less hard. I mean, starting jobs in Europe have six weeks vacation, OK? Nothing gets done in August in Europe, OK?

They've said-- and, you know, if you go to Europe, it's a little bit more rundown, OK?

It's not quite as gleaming and cutting edge as the US in many places, OK? Basically, Europe has decided to take some of that productivity increase and put it into more leisure time. We've decided take all that productivity increase and put it into better phones and gadgets, OK?

So the question is who's right. We don't know. But the important point is that's an open question. Just because we're more productive doesn't mean we should just consume more stuff. There's an open question of how you spend your productivity gains.

And then there's the final question and maybe the most important, which is who actually gains from productivity increases. So, from 1947 to 1973, productivity went up 2.5%, and virtually every group in society saw their incomes go up 2 and 1/2% a year. Since 1973, on average, productivity growth has been about 1.5%, 1.6% on average. You average these three series, about 1.6%. And average incomes have only gone up 0.4%.

So productivity has gone up 1.6%, but average income has only gone up 0.4%. The difference is the gains have all gone to the top of the income distribution. So, basically, virtually all of the gains from 1973 until a couple of years ago-- it's started to get better-- essentially, the bottom 80% of people saw no improvement of their standard of living over a 45-year period, whereas the top 20% saw a massive improvement. And, even within that, the top 1% saw a really massive improvement. And, even within that, the top 0.1% and 0.01%, et cetera, saw massive improvements.

So, as a result, in 1995, the richest 10%-- or the richest 10% of the population earned 15% of the income. Today, it's close to 25% of the income. It's getting even worse. Since 2009, if you look from 2009 to 2016-- I don't have it updated-- and you look at all the money that was made in society, on net, all of it went to the top 1%.

What do I mean by that? The top 99% were, in 2016, in the same place they were in 2009, even though the economy had grown. And all the growth went to the top 1%. So we're actually in an interesting world here where productivity gains by itself may not be enough if we care about what it does to the average standard of living.

And that leads to a very interesting set of issues around equity and fairness that we'll spend time on later in the semester. But I want to raise that issue, both that productivity gains can be spent in different ways on goods and leisure, and they can be distributed in different ways. And those are the sorts of things we need to be thinking about as we think about economic policy, OK?

So let me stop there. We'll come back-- I guess no section on Friday, right? It's an MIT holiday. And we'll come back on Monday and talk more on producer theory.