[SQUEAKING]

[RUSTLING]

[CLICKING]

ESTHER DUFLO: So what is causality? So we make causal statements all the time in everyday life. And I want to put three that I'm going to keep coming back to over the lecture today. One is that, "her headache got better because she took a pill." One is, "she got a good job because she went to MIT." And the third one is, "she cannot get a job interview because she is African-American."

So these are three causal statements. But in a sense, they are very different. They are causal statements in the sense that we-- in our mind, when we are doing a causal statement, we're thinking about some counterfactual where things would have been different. So in the headache example, the counterfactual is very simple-- she doesn't take the pill. And maybe the headache gets better anyway-- maybe it doesn't.

In the second example, the counterfactual is a little bit more fuzzy. Because what's the alternative? Is it going to Harvard? Is it going not to-- is it going to UMass Amherst? Is it going to go to the community college? Is it going to go nowhere at all-- not go to college at all? So the counterfactual is not very well-specified.

And the third one, of course, is the most complicated. Because the idea of manipulating race is not really-- it's hard to think about. And the question-- the statement is not well-defined as a causal statement, in a sense, because it might mean several things.

For example, it might mean that, because she is African-American, she grew up in these places where schools were not good. And she didn't get a good education. And now she can't-- she doesn't have the qualifications to get an interview. It could mean something very different, which is-- and I think that's probably what a lot of you thought about when reading the statement. How could it be read?

So one is that there is something deep about the education of this person, the life circumstances in which she grew up, et cetera, that all culminate into that outcome. What else could this be? Yeah.

AUDIENCE: I've heard studies that if you take two resumes and change the name to make it sound like it's from a different race--

ESTHER DUFLO: Yeah, exactly. It could refer to discrimination at the interview level, which is that she is actually-- we don't know what the rest of the statement was. But maybe she's super qualified. She went to Harvard Law School. And yet, she can't get a job interview. Presumably if she went to Harvard Law School, she's only applying to a pretty good law firm. But something prevent her to get interviews that's not about her-- just about her race.

And so if we adjust the statement now, it's not a very well-defined causal question because we don't know what we are manipulating. We don't know whether we are manipulating the perception of her race by the employer or whether we are manipulating the race and everything that goes with it-- her background, et cetera, her education, and the like. OK?

So all that to say that, when we are making a causal statement, we are trying to compare what happens in state A of the world to what happens in state B of the world, which may or may not be very well defined. And answering questions that sounds trivial but is pretty important requires defining causal question one.

So in general, when we think of causality, we think of the possible effect of manipulating a cause and what would happen if we had or had not manipulated this cause in this way. And then this needs to be-- this usually needs to be specified-- like, what is the manipulation that you're really thinking about in this particular case?

One little aside, I spend my life caring about making causal statements. So I care about them a lot. Do we only care about causal statements? I would say we care about them quite a bit. Many of the questions we want to answer in economics and in social science are causal questions.

Just keeping with the voting team that has been with us for part of this class, questions that people-- questions that are relevant for today's debate-- does immigration lower the wages of the native workers? Does trade increase inequality? Would a wall between Mexico and the US stop immigration or lower the level of immigration? Those are all causal questions.

Would the Mexicans pay for it? is not a causal question. And yet, it's a pretty interesting/relevant question. So in a lot of cases, data science and social science aim to answer questions of cause and effect. So it's going to be set up-- so we are going to try and set up our research in order to answer causal questions.

But it is worth nothing that they are-- "noting," not "nothing." [LAUGHS] It is worth noting that there are some questions that are also important and that are not causal questions. For example, you might be interested in identifying early warning signs of children who are at risk of falling behind in school. So we focus our effort on them.

I just read something about young children being much more likely to be diagnosed with ADHD. There might be a causal effect between too-young-for-your-grade and ADHD, but it's likely that there actually isn't and it's just that the kids are falling behind because they are too young for what is being covered. So what they need is not the diagnosis of ADHD. What they need is some help.

On the other hand, you could say, well, age for grade is a predictor of someone having issues following, and becoming agitated, and the like, even if there is-- so it might be causal. It might not be causal. It doesn't matter. For you or for a teacher, what's going to matter is to say that, the young boys in my class who are, let's say, coming from some categories of backgrounds-- I need to pay attention to them, make sure that they are with us, pay a little bit more effort on them, et cetera. Because they might be at risk of falling behind. And it doesn't matter whether this is causal or not. Because what matters is to target them. So in some cases, we are interested in prediction for the sake of it, irrespective of the causality behind it.

Of course, in businesses, in a lot of cases, people are interested in causal questions. For example, what is going to be the reduction in demand if I increase the price by $0.10? That's a causal question. But there are also many questions that businesses are going to ask which are not causal questions.

For example, you-- Google wants to predict that someone may be interested in something specific based on their search patterns to serve the ad that's most likely to be of interest. They couldn't care less whether it's causal or not. All they need to know is what ad to give.

So there are questions that are-- so in the policy realm and in the business realm, there are some questions that are of interest which are not causal. And for this, we have tools. We use the data. We have descriptive tool to look at the data-- correlations, nonparametric regressions, and then machine learning, which Sendhil is going to talk to us about-- where that's going to give us perfectly fine predictions that may have predictive pattern, correlation pattern, association between variables and others that have nothing causal in them but might still be interesting.

AUDIENCE: I was wondering what makes a question a question in economics. So say that there are, like, computer scientists at Yale who are working on serving, like, the best restaurant recommendations or something. That's also-- I guess it's like using the same sort of techniques. I'm wondering, would that also be economics? What makes it [INAUDIBLE]?

ESTHER DUFLO: This is a philosophical question maybe. I think "economics" is what economists happen to do--

[LAUGHTER]

--right no, which some people think is imperialistic. And some people think that's just the way it is. So to me, it doesn't really matter whether something is "economics" or not as long as we are doing it-- we're doing it. I think what makes it "science," as opposed to what makes it "research," as opposed to "business," which maybe is sort of the more pertinent question, to use your example, is whether it leads to generalized insights.

Going back to the IRB discussion we had, are you doing this in order to help-- that's it? Or are you doing this because you're thinking that there might be something interesting in the way people behave that may be relevant to other settings and other-- and then this can be looked at. Depending on the tools, it can be looked at in different ways.

So I think that's maybe the most relevant distinction between-- and I'm not saying she endorses that statement. [LAUGHS]

AUDIENCE: I'll endorse it.

ESTHER DUFLO: [LAUGHS]

So now we're thinking about this causality question. We're thinking that we're maybe not always interested in causal questions, but sometimes we are. How do we-- what is kind of good ways to represent and model causal relationships that's going to help us think through what we need in order to answer causal questions as opposed to just describe the data in some interesting way?

So something that you may encounter-- you're more likely to encounter if you're coming from computer science than if you're coming from economics is this, is DAGs-- that are called directed acyclic graph. This was proposed by Judea Pearl in a book from 2000, which is called *Causality.* And I'll-- I could define them, but it's going to be easier to show you one. And then you will see what they look like.

So here is an example that Pearl gives in this book. There is some-- there is a road. There is water on the road. And there are some variables-- so one is the season, whether it's a dry season or rainy season. One is, there is a sprinkler. Then there might be rain, or there might not be rain. The sprinkler might be on, or it might be off. The pavement is wet. And the pavement might be slippery, or it might not be slippery.

And so if you had no priors about what are potential relationship between the different variables in the study and you were just collecting all of these studies and trying to see whether they associate in a way-- or on the other, you had many possible associations. So the table d if you wrote this in a matrix, you have one, two, three, four, five variables. And you put them in a 5-by-5 matrix of, this could affect that, and, this could affect that. You'd have a lot of potential relationships to look at.

But of course, we know that the slipperiness of the pavement is not causing the season. Or we are willing to make that assumption. So we are willing to make a lot of things that are really exclusion restrictions-- or we would call them in economics exclusion restrictions, which is pretty clear. We have a pretty clear a priori theory-based reason to think which relationship are related to each other and in what direction.

So this is what's represented in this DAG. So what makes it a DAG is that it's directed, which means that the arrow don't go both ways-- so the season causes the sprinkler, but the sprinkler doesn't cause the season. It's acyclical, which means it doesn't go back. So we can't have, like, a long loop that brings back to season. Otherwise, that, again, would complicate relationships of cause and effect. And it's a graph. That's for the G.

So this is just really a way to represent in a graphical form which relationship are you going to rule out between variables and which relationship you want to keep. So here, that's his representation of what's going on in this world. It's a dry season, or it's a wet season. If it's a-- that affects whether or not I turn on the sprinklers. Because in dry season, I'm more likely to put on.

This basically, by the way, could be a probabilistic statement or a certain statement. This is a probabilistic statement that it's more likely that it happens. If it's a wet season, it's more likely to rain to the seasons or to affect the season. The sprinkler and the rain both will cause the pavement to be wet. And it's the wetness of the pavement that will cause it to be slippery.

So now we have a-- say, for x1, we have two possibility. Once we know x1, we have four possibility for x3. Once we know x1, we have two possibility for x-- four possibilities for x2. Then we have eight possibilities for x4. And that brings to one for x5. So we have many less table entries. If we were trying to describe this at a table, basically we have pruned a lot of the possible relationships between things. And it's much more manageable.

So it's kind of just a graphical way-- economists like to write things in equations. And this is another way to write it, in graphs instead, to say what can happen. Then there is a lot of language that goes with it.

So each little guy in this graph axis is called a node. The one at the top is the roof. The path is a relationship. It is a way to go, say, from x1 to x5 in the path-- 2x, 2x, x4, 5. Then he also talks about parents, and children, and grandparents, which is this kinship methodology. So for example, the season is the parent of rain, et cetera.

Then you have these different ways of writing the graph. So there are some chains. That x1, x2, x4, x5 is a chain, for example. We have some forks. So from x1, you can go to x3, or you can go to x2. You have "colliders" when two things springs to the same thing.

And the last thing that you hear about is these Markovian parents. It's just a way of saying, is it the case that conditional or knowing-- so it should be an underscore-j. What it's saying is that a node is another node's Markovian parents if a conditional or knowing that node, the probability the probability of xj, knowing the Markovian parents, is the probability of x node, knowing everything else.

So you have here what the Markovian parents are. So x1 doesn't have parents. The parents of x2 is x1. The parents of x3 is x1. The parents of x4 is x2 and x3. We need them-- it's a set of parents. Because if it was just x2, then it would be insufficient to describe the event that could lead to x2, et cetera.

Then he is interested in how do you represent causal effect in this framework. So here's this kind of funky little notation, which is, what would happen to x5 if I manipulated x3? So this is a do x3, which is different from just a conditional probability. Because a conditional probability could or could not be related to manipulating things.

So this kind of little notation here is trying to represent the fact-- it's trying to represent this idea of causality, which is, I'm asking what would happen if I move x3, as opposed to, what does x3 tells me about the probability of observing a particular value for x5? What does the particular value of x3 tells me about x5?

It could be that, in general, the sprinklers are off when the rain is-- when it's raining. But it doesn't mean that it's the rain that stops the sprinkler. It is that the sprinkler-- we put them on only in dry season, and it doesn't rain in dry season.

So in this graph, there is no causal effect of rain to sprinkler-- there could be, by the way, because there could be a gardener that is actually moving the things. But in fact, there is not-- by assumption, there is not here. And yet, there could be a correlation. So px-- p or x2-- do x-- there would be no causal effect on the rain on sprinkler. And yet, they could be correlated.

So that's kind of what this notation is trying to get at. And the way you represent it is by pruning many of the other-- you're thinking of an intervention. If I'm interested in the effect of sprinkler, I'm interested in just, what would happen if I modify the sprinkler and nothing else happened? So we are pruning the other part of the graph that are not related to the intervention.

And I could go on and on. In a sense, there is an entire book of going at that, which people-- I think people start using it quite a bit in other fields. We don't use them in social science yet-- or maybe ever. I don't know-- certainly not in economics. I think there could be a useful way for you to think about and succinctly express what's the causal model in your data. But they are not ever going to be a substitute for carefully thinking about causality, and manipulation, and the like.

So what I want to warn you about as you will encounter these DAGs in your life is that, at the beginning of the book, it's pretty clear that, in the end, if you want to estimate causal effects, you're going to need to make, a, some assumptions-- they are described in the graph-- and b, some statement about identification, which is, where is the source of variation that is giving me identification? That's coming from-- what is the reason why this variable has moved in this way? What does it tell me about potential causal effect of that variable?

As you move into this book, sometimes that distinction is a little lost. And a miracle happens. And by writing DAGs, you can get causal statement. Maybe I misunderstood, like, the depth of it. That's quite possible. But I also think that it's-- eventually, you can be a little bit misled by the-- let's say, by the technique of it.

And at the end, we always have to go back to thinking, well, what are we really saying? What are we manipulating in our mind? What's the ideal experiment we are considering? And what does this mean in this context?

So that's for DAGs. Any questions on DAGs? The book is pretty well written, by the way-- very informative. So you can learn more about them.

What do we do in economics? In general, we try to represent things with simple equations. And when-- I don't want to do much about regressions today. But I just want to think about them. In what way do they relate to these questions of causality?

So a very simple regression is a bivariate regression model, where we have some variable y that we assume is related to some other variable x via a coefficient, beta. And then there is an error term. The error term is just a prediction error between-- which is the difference between y and beta x. You can just write it y-- epsilon is literally y minus beta x. Epsilon i for an individual i is yi minus beta xi.

Typically, although not always-- I meant to say. Typically, although not always, we think of this model as a causal-- we think that this model has a causal interpretation, but it's not always the case. Sometimes we are perfectly fine to think of it in a non-causal way but just as a conditional expectation function.

But typically, we are thinking of this as-- we often think of this regression to have causal interpretations. And in that case, what are we thinking about? Well, we're thinking an ideal experiment, where we control a random variable X to the value little x. And we leave the rest of the world unchanged at some other value. And then that regression framework tells us that the value y of the random variable big Y is given by beta x plus epsilon.

So typically, we're thinking-- not always-- sometimes it's just purely descriptive. And that is fine. But often, we are thinking of this as causal statement. And in that case, that's the causal statement-- the causal statement of, if I were to manipulate some random variable X to the level little x, it would give me-- the value of the random variable big Y will be little y as beta x plus epsilon.

Sometimes we will include control variables. So for example, here, we can say, well, the outcome would be the wet floor. s is the sprinkler treatment. And X are the control variable, seasonal rain. Then there is no formal distinction in econometrics between the treatment and the controls, but it's a conceptual one.

We're interested, let's say, as the rain, the season, and the like as-- they are not our main variable of interest. But we are interested in the treatment effect of putting the sprinkler on whether the pavement is going to be wet, for example. Because we want to be able to move the sprinkler so that the pavement is not wet. Or we want to-- we are judging a law case to say that-- maybe someone fell off a laugh. We say that maybe it's the fault of the sprinkler.

So here, we are saying, well, it is important to control for the seasons. Because in wet season, sprinklers are typically off. And the pavement might be wet. So the two things, the sprinkler and the seasons, are correlated. And the status of the road and the season will also affect directly the status of the road.

So there, again, epsilon is the residual or the error term, the difference between the true value of Y that we observe and the fitted value from our best prediction. More on this later. But again, you can think of this as-- there is-- for any DAG, you can write them as regression. Probably for any regression, you can write them as DAGs, given some assumptions.

The idea is always there, which is, I'm modifying some treatment. And how much does it-- or conceptually, I'm thinking I'm modifying some treatment. How-- if it's a causal relationship that we're trying to interpret, the causal-- if we want to give a causal interpretation to that regression, we're always thinking of this in this way. I'm modifying some variable. How does it affect the prediction of my outcome variable? So that's maybe the absolute workhorse model that we work with in economics to represent causal relationships. Yep?

AUDIENCE: [INAUDIBLE]

ESTHER DUFLO: Well, here is that-- I didn't mention control. I mentioned treatment as an intervention, a thing you move. So here, your treatment is your sprinkler. Your sprinkler is on or off.

AUDIENCE: We were controlling x.

ESTHER DUFLO: Oh, treatment and control variables. Yes. So the control variables are stuff that maybe you're not directly interested in. As I said, there is no formal distinction between treatment and control variables. They are all random variables on here. But the control variables are things that may affect both the sprinkler being turned out and the wetness of the pavement. And therefore, we want to account for-- we think it's going to be important to account for in order for this relationship to be interpreted as causal.

For example, you might find that-- suppose you didn't include-- suppose you were just looking at the correlation between the wetness of the road and whether the sprinkler on or off. You might find that the road is wet when the sprinkler are off. Much more likely-- the road is much more likely to be wet when the sprinklers are off.

And you might say, well, does it mean that the causal effect of the sprinkler is to make the road dry? And clearly, that doesn't sound very intuitive. And what did we forget? Well, we forget the fact that the sprinkler are off in the rainy season, where it's raining. And hence, the road is wet.

AUDIENCE: So the DAG would be control treatment and--

ESTHER DUFLO: The DAG is this DAG.

AUDIENCE: No. For this case, it would be control affects treatment and yi, and treatment also affects yi.

ESTHER DUFLO: No, the DAG is this one. I was trying to represent an equation that represents this DAG, which is, the season affects whether it rains and affects the sprinkler.

AUDIENCE: If you were to--

ESTHER DUFLO: So when the season is wet-- so there is no-- the treatment variable that I'm interested in is the one that I'm willing to manipulate, that there is something that I can manipulate. So the treatment variable-- I'm calling the "treatment variable"-- is the sprinkler. The control is-- it could be the season. It could be the rain. So that's what I'm going to control for in my regression in order to be able to interpret the beta as a potential causal effect of the sprinkler.

If I didn't put it in-- and we'll put this-- when we go to regression, we'll go through this much more formally. But if I didn't put it in, in this case, I would probably get the results exactly backwards. Because in my data, I'm going to see that whenever the sprinkler off, the road is much more likely to be wet than when the sprinklers are on. That make sense?

So we'll go back to regression. That's the most common way to represent causality. But in a way, I think it will help us make a little detour through another way to think about causality that will also help us think about regression for what I mean. And then sort of the technique of actually estimating the coefficients, et cetera-- we'll go back in a couple of weeks.

So here is a third way. By the way, I think all of these things kind of map into each other very neatly. But here's a third way to think about it. And I'm sorry for the confusion. But we are now going to represent treatment and control different. What I call before "control variables" we are going to-- will be called "covariates" from now on. And in fact, I should have called them "covariates" from now on. And here, a "control" is going to be a unit that is not treated.

So the potential outcome framework is due to a statistician at Harvard called Donald Rubin. I do find it very useful to think about randomized control trial, which as you know will occupy most of my life and will occupy us for a couple of lectures, and about causality more generally. That's not the only way to think about things. I think equations are writing equations much more common. But it is spreading. And I think it's very useful to be conversant with it and to be able to toggle from regression interpretation and a Rubin causal model to be able to understand what you're doing in both kinds of the world.

So what's the Rubin causal model? You take a unit-- for example, a person-- and a set of actions. For example, that person could have taken a pill or could have not taken a pill. And we associate each action unit pair with a potential outcome. So let's take the examples that we started from. "Her headache got better because she took a pill." The first example-- we could say, well, the headache-- there are only two possible outcomes. You have a headache or you don't. Then there are only two possible value for pill, yes or no. So sometimes we will be referring to the pill as treatment and the no-pill as control.

So here, for each person, there are two potential outcomes if they take the pill. There is-- so each person realize one potential outcome if they take the pill-- either headache or no headache-- and one potential outcome if they don't take the pill-- again, either headache or no headache. So that's a very simple example, where there are four possible pairs.

The second example, as we discussed, is a bit less clear because we need to define the alternative if she didn't go to college. So MIT-- are we interested in MIT versus any other possibility? Are we interested in any college versus any non-college?

Are we interested in any research university versus another type of university? et cetera. But once we have defined what constitutes the treatment and what constitutes the different scenarios of the world that we are interested in, we'll be able to say, in principle, everybody is on loud for potential outcome for each of these state of the world.

And by the way, there don't need to be two. There could be any number of distinct values. So you could have gone to MIT. You could have gone to Harvard. You could have gone to-- let's say we are staying to Boston. You could have gone to UMass, or you could have gone to Bunker Hill Community College, or nowhere, and anything else. So this gives us several state of the world. And for each of those, we could say a particular person would have had some income at the end of her studies had she gone to any of these different options. And those are the potential outcomes.

Our third example, as we discussed, is even less clear. What do we mean by, what would happen if she was from another race? There are various ways to think about it. For example, we could say, I'm only going to be interested in the resume. And what will happen if there was something in her resume indicating that she is African-American versus that she is not African-American? And we'll get back to the specific example of that that you mentioned earlier.

So now that we have-- once we have potential outcome, and we know that, say, there is my potential outcome, and we are interested in the two-- we have defined them. We have defined the set of possible state of the universe and the potential outcome for those that we are interested in-- for example, pill, no-pill.

Then the causal effect is simply the difference between-- for a particular individual, the causal effect is simply the difference between the potential outcome in one state of the world versus the other one. So for example, in the pill examples, there are four possibilities.

You could have no headache if you take the aspirin and a headache if you don't take the aspirin. You could have a headache if you take the aspirin and also if you don't take it. It is that sad state of the world. You could have no headache if you take the aspirin and no headache if you take no aspirin-- so nice state of the world. And you could have a headache if you take the aspirin but not if you don't take it. OK? So there are four possibilities.

So in this case-- for one particular person. So in this case, if we are in the first case for this particular person, aspirin makes the headache go away. Case two or three, there is no effect. Case four, aspirin prevents-- sorry-- the headache from going away. So those are the possible treatment effects in this case. Does that make sense for every person?

So the definition of treatment effect depends on the potential outcomes but not on what we actually observe. For every given person, either she has taken the aspirin or she hasn't. But her causal effect depends on those potential outcomes, only one of which I get to see.

So the causal effect for a particular unit involves the comparison between something I see and something I don't see. Because it's, what would have happened to this person had she taken the aspirin if what we observe is just a world where she has not taken the aspirin? So this is what Holland referred to as the fundamental problem of causal inference-- is that, at most, we can only see one of the potential outcomes that can be realized and just observed.

So we all are missing a big part of the data that we do need. But for the estimation of treatment effect, we are going-- what can we do? We cannot-- so the treatment effects are nicely defined as a function of potential outcome. But so unless we're just willing to make them up, if we want to estimate causal effect, we are going to have to manage with the world as we see it. So that's the problem.

And that means that we will need different-- we will need many units. So one person, of course, will only give us one piece of information. If we-- so with one person, we will never be able to get to any estimate of causal effect because we don't even have the second part of the comparison. At a minimum, we'll need two units. And it's going to be pretty critical to know or to make assumptions about-- and in a lot of cases, that's going to be the name of the game-- to make assumptions about the way that some potential outcome got realized and not others.

So this is the decision of what we call assignment mechanism, which is, if I observe a group of people, how come some of them took the headache pill and some didn't? Was it-- how did it get to be the case? Is it the case that the people who felt particularly lousy took the pill? Is it the case that people who felt particularly lousy just went to bed and didn't take the pill? Is it the case that some happened to have aspirin in their cabinet and some didn't? et cetera. So this is what is going to determine whether or not-- in the data that I observe who gets what kind of treatment.

But unfortunately, even before we get there, the problem with many units is that, if we allow for all the possibility in the world, things can quickly become more complicated. So suppose that I'm with Sara in her office. And we are both in our office. And our office are in the living room.

We basically hesitate, I think. Because before, I put my husband, who is also a faculty. So we are frequently preparing class notes in the living room at the night figured, actually, Sara would be a better example. And we would be more likely to be in our office than in our living room, either of our living room.

[CHUCKLING]

Although, it's not impossible. So suppose that I'm in our office. And we both have a headache. We've been doing a lot of work. And we both have the option to take aspirin because it's in our office. So now each of us has-- if we want to fully describe the world without any loss of generality, there are four potential outcomes for each of us.

There is a potential outcome where I take the aspirin and she doesn't, the potential outcome where I take the aspirin and she does-- should be an "SE" here. There is a potential outcome where neither of us take the aspirin. There is a potential outcome where I don't take the aspirin and she does.

So in this situation with two people, there are four choose two, six different comparisons depending on which of the potential outcomes are compared. So it is very nice to add more units because we now have more things to compare. But we also have more people that we can start comparing. But at the same time, we added more potential scenarios that we might want to compare to each other.

So if we are not willing to make any assumptions on the relevant/interesting state of the world, then as we add more units, we also add more state of the world that we might want to compare to each other. And therefore, we have not solved any problem. And we are not even in a position to start thinking about assignment mechanism.

So what is a reasonable assumption to be made in the case of the headaches in this particular example?

AUDIENCE: [INAUDIBLE]

ESTHER DUFLO: Is "independent" the word you want to use?

AUDIENCE: That both subjects are-- you can kind of think that both subjects are one subject but being subjected to an experiment at the same time.

ESTHER DUFLO: Yes. So I think what you're trying to say is that-- it's exactly that. I think what you're trying to say is that what happens to me doesn't depend on what she does. So that means that, for example, the potential outcome, I taking the aspirin and she doesn't, is the same as the potential outcome, I take the aspirin and she does. So that's the assumption that you have in mind of the potential outcome? What happens to my headache does not depend on what happens to-- on what she decides to do.

Note that sometimes it's an innocuous assumption-- for example, if we were each in our separate offices, it would be a particularly innocuous assumption-- and sometimes it isn't. For example, if we are in fact both in our room, and she doesn't take aspirin, and she keeps complaining about a headache, then maybe I'll feel like I'll get a headache. So in that sense, there might be-- so typically, we don't think that much about this. We will see that we kind of often sort of make the assumption. But it may or may not be worth-- it may or may not be a good assumption. The point is that it is an assumption. You had a question. Yeah?

AUDIENCE: What's the y?

ESTHER DUFLO: It's the outcome. So it's the potential-- so in this particular example, it's a random variable that can be either headache or no headache. It takes-- yeah?

AUDIENCE: [INAUDIBLE]

ESTHER DUFLO: Sorry?

AUDIENCE: [INAUDIBLE]

ESTHER DUFLO: So I take-- Esther takes the aspirin; Sara does not take the aspirin. Esther takes the aspirin-- this one should be, Sara takes the aspirin. So it should be SA. Esther doesn't take the aspirin; Sara doesn't take the aspirin. Esther doesn't take the aspirin; Sara does take the aspirin. OK?

So these are all the potential state of the world in principle. And if I allow Sara's treatment status to affect my headache through her whining, for example-- which she doesn't, by the way-- then this would be-- this is the set of the universe with two people.

So-- so we discussed that perhaps a natural assumption in the headache example is that she does not influence mine. And as we said, there are ways in which it will fail. I think I have some-- yes, I will give you some examples.

So here's the assumption more formally. So it has this complicated name for some reason-- Stable Unit Treatment Value Assumption, or SUTVA. And it says that the potential outcome for any unit do not vary with the treatment assigned to other units. So the potential outcome for me, my headache, as a function of the two possible treatment stages for me, treated or not, depends only on that and not on her-- on what she or anybody else is doing.

And the second part of it is, for each unit, there are no different form of versions of each treatment unit leading to different outcomes. What this means is that there is something that is defined as a treatment. Otherwise, we don't know what we're talking about. So this part of the SUTVA assumption is just to be able to define a treatment. Otherwise, if every person could have different versions of the treatment, it would get very confusing.

But the second part is the one that could be more controversial-- is, the potential outcome for any unit does not vary with the treatment assigned to other units. So to go back to what we had here, it says that we can remove-- for me, we could only worry about my treatment status. And for Sara, we can only worry about her treatment status.

So besides the headache example, I want you to think about example where this particular assumption is likely to fail-- real-world example. Yeah?

AUDIENCE: So is it something that's independent and the treatment is identical?

ESTHER DUFLO: So the treatment doesn't have to be identical across people. Instead, for each unit, there is something called a treatment.

AUDIENCE: The example could be if you're vaccinating people. Then if you vaccinate more people, then the other people are less likely to get sick.

ESTHER DUFLO: Exactly. So immunization would be a key cases where SUTVA will fail, which is if we are defining the treatment status at the level of the individual. Because anybody who lives in an area where most people are immunized will not get sick with that disease.

De-worming medicine is one example that we studied in development, where if you take de-worming medicine, the worms that are in your-- all the worms dies. And therefore, they are not going to infect another kid. And conversely, if another kid doesn't take de-worm medicine, the worm could travel from them to you. So that creates a violation of SUTVA. What could be another-- a non-medical example?

AUDIENCE: Heating. So in my [INAUDIBLE] heating unit place, then it is-- if 10 of us have it, then the 11th person will automatically have it.

ESTHER DUFLO: Right, exactly. It could be that. You could have-- an example on which I work is job training-- like, a program that helps long-term unemployed apply for jobs. You could think that SUTVA might be violated if the people who get this program become very good at interviewing and they get the few jobs that exist.

The other people who don't get treated get hurt by the fact that someone else took the job that existed. It could be violated or it could not. If the only thing that-- if the number of jobs were actually infinite for qualified people, then all I would do by helping long-term unemployed honing their interview skills is make sure that there are people to take those jobs. So it depends on the state of the world.

So how are you-- in the case of immunization, if I wanted to think of the causal effect of introducing a new vaccine, how would I solve-- how would I design an experiment that does not run into the SUTVA-- that doesn't run into an obvious violation of SUTVA?

AUDIENCE: Could you look at different groups that are spread apart geographically?

ESTHER DUFLO: Exactly. What you would do is you would define your units differently. So you would say, well, my unit is not an individual within a community. My unit is the entire community. And now I can apply-- I can assume SUTVA the level of the community.

But note, it's still an assumption. There is no way to not have-- there's no way. You're never free from that assumption. It's always an assumption, but that you use your biological/economical/whatever knowledge to make it-- and that you need. Because otherwise, you have too many people to treat.

AUDIENCE: [INAUDIBLE]

ESTHER DUFLO: Yeah?

AUDIENCE: So by changing the units, we don't answer the original question. We just frame it differently [INAUDIBLE] different question in the same.

ESTHER DUFLO: I think you're still answering the question you're interested in. You're interested in the causal effect of immunization, say, against-- these new pneumococcal meningitis vaccines got introduced in the developing world not so long ago. So before you introduce it, you want to have some way of answering the causal question of how protective it is really going to be.

If you were doing your experiment at the individual level within a community, you would not answer your causal questions because SUTVA would be violated. So by just comparing people according to their own treatment status, you would have misleading comparisons because your treatment status affect my potential outcome.

If I go to a larger level and instead of randomizing at the-- or doing your experiment, let's say-- let's say randomizing or doing your assignment at the individual level. You assign at the community level, and you assume that kids don't move. Then you're going to vaccinate all the children within one area, and then not vaccinate all the children within one area, and do that across many areas.

Then you'll be able now to compare all the children that end up receiving the vaccines to all the children that end up not receiving the vaccine. And you know in this case that SUTVA holds. So you're still answering your original question. You just have to change the design.

So assume now that SUTVA holds. We solve problem number one-- by assumption, by the way. But sometimes it means that we will be thinking carefully about design. Then the aspirin example simplifies to two situations for Sara and I. Each of us can either take or not take the aspirin. And what the other does is not relevant. And of course, that extends to many units. If in fact we looked at the entire department, and we each were in our offices, and we each take aspirin or not take aspirin, we could-- that extends.

So now assume that we are not the two of us. We have a population of size big N, which is indexed by e, which takes value going from 1 to N. And define the observed value of the outcome Yi obs as the potential outcome that is actually observed. So W is the treatment. So the treatment for each person is 1 or 0.

So the Yi observe is Y of 0 if Yi is 0 and Y of 1 if Yi is 1. So the potential outcome that is actually observed is the potential outcome not treated for the not-treated population and the potential outcome treated for the treated population. And then the missing one is of course the complement, which is the potential outcome treated for the people who are controlled and the potential outcome control for the people who are treated.

Causal effect for person i is Yi of 1 minus Yi of 0. And the missing data problem that we already talked about, the Holland problem, is the fact that we only see Y obs. So we can never calculate the treatment effect for each person. So that's not happening. So let's forget that. It just won't happen.

So the only thing we can do is to try to infer something about Yi miss from the data we do observe. But in doing that, we'll need to know the assignment mechanism-- why are some people ended up treated, and why some people didn't end up treated? So in the case of the aspirin, what could be a potential outcome? What could be-- what are different assignment mechanism that we could consider?

AUDIENCE: The people with the worst headaches take the aspirin.

ESTHER DUFLO: The people with the worst headache would take the aspirin, for example. What could be another one?

AUDIENCE: Random assignment.

ESTHER DUFLO: Random assignment. Someone is actually conducting a study on this new form of aspirin, and has enrolled the population to give them or not give them the aspirin, and will do it randomly. So those are two candidates. So we're going to keep those two in the back of our mind for the time being-- hoping it's not going to give you a headache-- and think about the selection problem.

So imagine we have a large group of people who took aspirin and a group who did not. And we decide that something that might make sense-- a very sensible way to proceed is to say-- we have been discussing sample mean for a long time.

You're coming out of all these things where sample mean comes back and say, well, let's-- how about let's take the sample mean for people who take the aspirin and people who didn't? And we know from everything we did until today that the sample mean will converge to the actual population mean in the groups.

So we know that it's going to be a good estimator for the expectation of Yi given that you took the pill and that it's going-- the sample mean for the treated group is going to be a good estimate of the expectation of Yi given that you took the pill. The expectation for-- the sample mean among the people who did not take the pill is going to be a good estimator of Yi given that Wi equals 0-- that you didn't take the pill. Make sense? Are you with me until now? Very simple.

So this is what we have here. Observe the expectation of Yi observed given that Wi equal 1 minus the expectation of Yi observed given Wi equals 0. We can replace because we know what this is. This is the potential outcome treated for the people who are treated and the potential outcome control for the people who are control-- not treated for the people who are not treated.

Now, what I'm going to do is that I'm going to add and subtract the same term. So I'm going to add and subtract the expectation for the potential outcome untreated for the population of given that you actually observed the treated group, that you are in a treated group. I'm going to-- this is something that I cannot see in the data but that exists perfectly well-defined. So I'm going to subtract it and add it.

And now what do we have? Well, this guy is kind of very promising. Because from everything you know about expectation, you know we can combine them. And this guy is Yi of 1 minus Yi of 0 given that Wi is 1, which is the average treatment effect for people who end up being treated. So that's nice. That's something that we care about.

That's the treatment effect on the treated. That's what we'd like. But unfortunately, there is another term that is on the side here. And this one is how-- I wrote here that we can call it the selection bias. But what's another way-- what's a way to think about this one? What does it tell us? And why did I call it the selection bias? Yeah?

AUDIENCE: Well, just differences-- the way I see it is, like, there's a difference between the group of people that was treated and the group of people that wasn't treated that were not dependent on what was changed-- so maybe different ages, or different racial backgrounds, or different educational backgrounds, or whatever [INAUDIBLE].

ESTHER DUFLO: Exactly. So what it says is that, in the event before they took the pill, the potential outcome untreated for the treated group may or may not be different. And the difference between the potential outcome untreated for the group that ends up taking the pill and the group that ends up not taking the pill is the selection bias.

So in the headache case, take with Lisa's example that it's the people who have the worst headache who take the pill anyway. Then the selection bias is going to be of what sign? Sorry? Positive. Exactly. It says that, in the absence of taking the pill, their headache would have been worse likely. Because that's why they took the pill in the first place. So in the absence of the pill, their headache would have been worse.

AUDIENCE: I think there's a typo in this.

ESTHER DUFLO: Yeah? That might be. So in the case of their headache-- this one is correct. In the case of the headache, this is going to be a positive selection bias. So that means that I might think that the pill has no effect because at the end, say, everybody's headache got resolved regardless. But this is made of the fact that there is a positive treatment effect for people who took the pill and-- or sorry, a negative treatment effect on having the headache and a positive selection bias.

Now, if we take the example of random assignment, what happens to the selection bias?

AUDIENCE: [INAUDIBLE]

ESTHER DUFLO: It goes to 0. That's the nice thing of treatment assignment. So thinking more about the selection effect, people take headache pills because their headache is pretty bad. People who go to college have all sorts of attributes-- that's what you were mentioning-- that are different than people who do not. So the bias disappear in case of random assignment. So randomization does solve the selection problem.

So what is a randomized experiment? So there are different ways to do a randomized experiment. But simplest case is what we call a completely randomized experiment, where you literally pick a random sample, Nt units. Using uniform random sampling, you take Nt unit from N to be in the treatment group and Nc you need from the control group. Then the probability of assignment does not depend on the potential outcomes.

The selection bias, which still has inherited this typo, becomes 0. And therefore, the difference between treated and control is the treatment effect. It's the treatment effect on the treated. But because the treated are randomly selected, it's also the treatment effect on the population because the fact that you are treated tells you nothing. You are-- there is nothing specific about you. Does it make sense?

So the name of the game in a lot of econometrics-- not all of it because sometimes we're interested in non-causal question. But in anything that have to do with causal questions, the name of the game is, when we don't have a randomized experiment, how can we make that guy go away anyway? What are the assumptions that are going to help us making this guy go away anyway? So that's kind of-- a lot of applied methods in social sciences are about that, except if you're lazy like me and you're thinking, well, it goes away with randomization, let's just stop there. And then you just-- the name of the game becomes to run randomized control trials.

But most people kind of take it one step further and say, well, what happens when we cannot run an experiment? Is there other way to make this guy go away? And that's kind of a lot of what we are going to do in the rest of the semester when we finish talking about RCTs, is thinking about other ways than randomized control trial to make these guys go away.

But for now, we can stop, pause in the luxury of randomized control trial, where it's gone, and think about how to-- so first of all, when do we have randomized control trials? We have a randomized control trial where some-- when someone has randomized something. So that is important. Because often, you have, let's say, casual use of randomization.

You don't have a randomized control trial when it's done every which way. "Random" means random-- means someone had actually used a randomization device to actually assign the treatment. So the assignment mechanism is controlled by the researcher or someone-- that the researchers know who they are.

So the simplest type of randomized control trial is completely randomized, which is, you take your sample of people or units. It could be-- in my example of pneumococcal meningitis vaccines, you're going to take communities and not people because the super-assumption wouldn't apply. And you randomly assign them to a treatment or a control. So that's the completely randomized experiment.

Sometimes you decide that you're not going to do that. You're going to do stratified randomization. So you create blocks of some covariate X-- for example, race or gender, or the combination of the two; or education; et cetera. You make cells. And then within those cells, you randomize.

Or you can take that to the limit and say, well, I'm going to randomize by pairs. So I'm going to create pairs of people with very, very similar characteristics that I observe. And then I'm going to select one member of the pair to be in treatment, one [INAUDIBLE] control.

Cluster randomization is what we discussed with the vaccines, where the units to be randomized are not individuals but are villages, or classrooms, or stuff like that. And the result above, which is the randomization, removes the selection bias hold whatever the type of randomization.

If you are-- if the probability of assignment depends on the blocks, then you need to do the-- it only applies within each block. So you can do your comparison within each block and re-aggregate them. If the probability is the same in all the blocks, it actually applies in general. So why is that? Why it doesn't really matter how you randomize for the main result-- that the selection bias will go away with randomization?

AUDIENCE: Isn't it the answer that the selection bias goes to 0, and so that's why the result holds?

ESTHER DUFLO: Yes. But the point is it continues to go away, even if you've randomized by strata, et cetera. That's because of property of expectations that we've discussed in the class, which is the expectation-- if you have unbiased within each cell and you add them up, it's going to continue to be-- the expectation being linear, it's going to nicely add up. So you're going to add up a bunch of zero bias. So you're going to get still zero bias.

So intuitively, do you have an idea of why we might decide to have these blocks of X? Yeah?

AUDIENCE: Does that mean it's SUTVA.

ESTHER DUFLO: So I think the answer to SUTVA is why you would decide to do clusters. That's because you want to make sure that people do not interfere. Yeah?

AUDIENCE: I think it depends on your population and the intervention you're using. So if it's, like, a school-based intervention, if-- it doesn't make sense to look within one class and then divide the class in half?

ESTHER DUFLO: Exactly. So that's, again, a reason for why you would want to cluster. So in a lot of cases, the natural unit of randomization will not be the individual either because of violation of SUTVA, or because the intervention is defined at the level of a cluster, or because it's just politically/logistically/et cetera not feasible to not do cluster. So those would be reasons why you would cluster. That means take an entire cluster of people and randomize them together.

Now, stratifying, which is creating kind of cells of X's and randomizing within, is almost the opposite of clustering, right? We create cells of people. And for example, we could decide to stratify by village-- not for something where we're worried about SUTVA. But for, say, our headache example, we could stratify by village and then randomize. So why would we like to do that as opposed to just randomize in general? Oh, go ahead.

AUDIENCE: Well, I work on drug trials for rare diseases. And so any one hospital might not enroll that many people. So you don't want it to-- just by happenstance, the first four people who called get AAA. Well, if you don't get a fifth person, then you haven't tried the control on anyone. So you want to cluster within small groups instead of-- you might not-- your n is 20. But if you only end up in the one four, you're kind of screwed with those other 16 [INAUDIBLE].

ESTHER DUFLO: So I think that's absolutely right. And let me give a more general intuition to that. The reason why you might want to randomize is that, even though in expectation it is true that the treatment and the control-- in expectation, the potential outcome would be the same. In any particular sample that you realize, the potential outcome is whatever it would have been. When the samples are not infinite/when the samples are not very large, the actual realization for a particular population could be quite different than in fact.

And you never know. So you're going to try and make the-- if you think that there are variables that might affect potential outcome, that might be correlated with potential outcome-- for example, which hospital that they have been treated, their age, their gender, et cetera-- you're thinking, let's try and make our people as similar as possible on these observable characteristics by forcing the randomization to happen within people who have the similar characteristics. That way, we can hope that their potential outcome is also similar in reality, even though we know that in expectation it will be similar forever. So that is exactly why you would stratify.

Let me stop here. Actually, no, let me finish this. And then we'll do the little empirical example. I don't know which way it makes sense to do it.

I might stop here. We'll go back to that in a minute. I want to spend the next seven minutes to talk about Gabriel. This is Gabriel. And this is not a randomized control trial, but this is a causal question.

Gabriel is a graduate student in the Department of Economics. And I want to advertise his research on-- as an example of something that is done it, you don't want to do it again but could be-- it's a nice, feasible research project that has a well-defined question he's trying to answer.

And this is using Google Map-- "API", not "IPA"-- using Google Map API to evaluate the causal effect of two congestion-reducing policies. On the website of the class, you will see all of his Python code for getting the data, et cetera. So there will be a little archive that you can study at your leisure to-- as an example.

So Google has set up an API to access Google Maps. So you don't have to do your own web-- they have a tool to do the web-scraping very easily. It's mostly used by smartphone apps, websites, et cetera. Google provides nice documentation libraries to access-- very simple access. There is nothing complicated.

So in kind of two lines of code, you can access Google Map data. It's not very expensive. The first 2,500 queries per day are free. And then it costs $1 for 2,000 queries. So even though it's mostly commercial, researchers can also take advantage.

So one thing that you could ask Google Maps, for example, is the travel time between two places. And it takes into account-- so when you go onto Google Maps and you're getting an ETA for a particular travel, it will take into account traffic conditions at different hours. So the departure time at now is a prediction based on the crowd-sourced live data from Google Android users as well as historical data on that route.

So when you Google how long it was going to take you, there is an algorithm. There is the machine-learning algorithm that is trying to guess how long it's going to take you to travel this time. So for example, on the-- you ask for a particular route at 8:34 AM. It's going to take you-- it's going to tell you how long it's going to take you to do it. There is no-- the historical data you can't get through this API. You can only get the predictions.

So what did Gabriel do? So he was interested in a policy that was introduced between 1 and 15 January 2016 in Delhi, which is the odd-even policy. So basically, it says that you can only drive on the day where-- if you have an odd license plate number-- it ends with an odd number-- you can only drive on odd days. And if you have an even license plate number, you can only drive on even days.

And he's interested in the short-run causal effect of that policy on travel time. Keep in mind, short run is very important. Because of course, in the long run, people could decide to buy another car, or something like that, or buy a different license plate. And that would change the policy.

But so what did he do? Very simple-- 15 days before, he started querying 93 routes every 20 minutes and gets-- and so he's basically set it up to run around Christmas time-- set it up a little bit before Christmas-- set it up to run and collect all this information of the travel time on these 93 routes. Here, you have the direction of the routes.

And then what is the research project? Very simple-- it's the average travel time between-- across all these routes for every single day in minutes. And what is the causal effect of the odd-even policy? It is simply the difference between the two lines in this case. What is this estimate of the causal effect of the odd-even policy? It's the difference between these two lines. You have a sense that maybe it does make sense. You have-- in a national holiday, you have almost nobody traveling, et cetera.

Here, is another example that is potentially interested. It's the-- yeah?

AUDIENCE: Sorry. Before the red line is-- the pilot has been enacted. But when is the odd-even policy being [INAUDIBLE]?

ESTHER DUFLO: So this is odd-even. This is the odd-even period. And this is the-- so sorry. I told you that there was pre-time. But in fact, there is no pre-time here. This is the odd-event period, and this is after it's finished. So during the odd-event, the time was shorter than after the odd-even-- makes sense because you have fewer cars on the roads. So this is the odd-event period. And this is the-- so the causal effect is the difference between the two. So it's about some minutes.

By the way, during that time, what he also did is he kept calling drivers during the prior time. He kept calling drivers regularly to ask them, how did you commute today? He had a sample of driver that he got just before. And he kept calling them to say, how did you commute today? to find out what people did instead. So that's another set of questions that are potentially interesting. But this one can be done from the comfort of your own office.

Here is another traffic policy in Jakarta. Jakarta has a 3-in-1 policy that they use during peak season. So during peak hours during the day, you can only-- on some roads, you can only travel if you have three people in a car. And this is the travel time on this road for people depending on the time of the day. And you can see that, during the restriction policy, the travel time is lower in the place that-- in the road that has the policy than in the place that doesn't.

And then immediately after, which is sort of interesting, you have a huge increase in the travel time. As soon as you leave the 3-in-1 period, you have a huge jump in the travel time. So that's kind of another example of something you can do. Again, here on this line, when the 3-in-1 is reinstated in the afternoon, immediately-- the moment you have the reinstatement of the 3-in-1 policy, the travel time goes down.

Thank you very much. We will continue with analyzing RCTs, power calculation and the like, on Monday.