PROFESSOR: So we're going to start now with a new chapter. We're going to talk about Markov processes. The good news is that this is a subject that is a lot more intuitive and simple in many ways than, let's say, the Poisson processes. So hopefully this will be enjoyable.

So Markov processes is, a general class of random processes. In some sense, it's more elaborate than the Bernoulli and Poisson processes, because now we're going to have dependencies between difference times, instead of having memoryless processes. So the basic idea is the following. In physics, for example, you write down equations for how a system evolves that has the general form. The new state of a system one second later is some function of old state.

So Newton's equations and all that in physics allow you to write equations of this kind. And so if that a particle is moving at a certain velocity and it's at some location, you can predict when it's going to be a little later. Markov processes have the same flavor, except that there's also some randomness thrown inside the equation. So that's what Markov process essentially is. It describes the evolution of the system, or some variables, but in the presence of some noise so that the motion itself is a bit random.

So this is a pretty general framework. So pretty much any useful or interesting random process that you can think about, you can always described it as a Markov process if you define properly the notion of the state. So what we're going to do is we're going to introduce the class of Markov processes by, example, by talking about the checkout counter in a supermarket. Then we're going to abstract from our example so that we get a more general definition. And then we're going to do a few things, such as how to predict what's going to happen n time steps later, if we start at the particular state. And then talk a little bit about some structural properties of Markov processes or Markov chains.

So here's our example. You go to the checkout counter at the supermarket, and you stand there and watch the customers who come. So customers come, they get in queue, and customers get served one at a time. So the discussion is going to be in terms of supermarket checkout counters, but the same story applies to any service system. You may have a server, jobs arrive to that server, they get put into the queue, and the server processes those jobs one at a time.

Now to make a probabilistic model, we need to make some assumption about the customer arrivals and the customer departures. And we want to keep things as simple as possible to get started. So let's assume that customers arrive according to a Bernoulli process with some parameter b. So essentially, that's the same as the assumption that the time between consecutive customer arrivals is a geometric random variable with parameter b.

Another way of thinking about the arrival process-- that's not how it happens, but it's helpful, mathematically, is to think of someone who's flipping a coin with bias equal to b. And whenever the coin lands heads, then a customer arrives. So it's as if there's a coin flip being done by nature that decides the arrivals of the customers. So we know that coin flipping to determine the customer arrivals is the same as having geometric inter-arrival times. We know that from our study of the Bernoulli process.

OK. And now how about the customer service times. We're going to assume that-- OK. If there is no customer in queue, no one being served, then of course, no one is going to depart from the queue. But if there a customer in queue, then that customer starts being served, and is going to be served for a random amount of time. And we make the assumption that the time it takes for the clerk to serve the customer has a geometric distribution with some known parameter q.

So the time it takes to serve a customer is random, because it's random how many items they got in their cart, and how many coupons they have to unload and so on. So it's random. In the real world, it has some probability distribution. Let's not care exactly about what it would be in the real world, but as a modeling approximation or just to get started, let's pretend that customer service time are well described by a geometric distribution, with a parameter q.

An equivalent way of thinking about the customer service,mathematically, would be, again, in terms of coin flipping. That is, the clerk has a coin with a bias, and at each time slot the clerk flips the coin. With probability q, service is over. With probability 1-q, you continue the service process.

An assumption that we're going to make is that the coin flips that happen here to determine the arrivals, they're all independent of each other. The coin flips that determine the end of service are also independent from each other. But also the coin flips involved here are independent from the coin flips that happened there. So how arrivals happen is independent with what happens at the service process.

OK. So suppose now you want to answer a question such as the following. The time is 7:00 PM. What's the probability that the customer will be departing at this particular time? Well, you say, it depends. If the queue is empty at that time, then you're certain that you're not going to have a customer departure. But if the queue is not empty, then there is probability q that a departure will happen at that time.

So the answer to a question like this has something to do with the state of the system at that time. It depends what the queue is. And if I ask you, will the queue be empty at 7:10? Well, the answer to that question depends on whether at 7 o'clock whether the queue was huge or not. So knowing something about the state of the queue right now gives me relevant information about what may happen in the future.

So what is the state of the system? Therefore we're brought to start using this term. So the state basically corresponds to anything that's relevant. Anything that's happening right now that's kind of relevant to what may happen in the future. Knowing the size of the queue right now, is useful information for me to make predictions about what may happen 2 minutes later from now. So in this particular example, a reasonable choice for the state is to just count how many customers we have in the queue. And let's assume that our supermarket building is not too big, so it can only hold 10 people.

So we're going to limit the states. Instead of going from 0 to infinity, we're going to truncate our model at ten. So we have 11 possible states, corresponding to 0 customers in queue, 1 customer in queue, 2 customers, and so on, all the way up to 10. So these are the different possible states of the system, assuming that the store cannot handle more than 10 customers. So this is the first step, to write down the set of possible states for our system. Then the next thing to do is to start describing the possible transitions between the states.

At any given time step, what are the things that can happen? We can have a customer arrival, which moves the state 1 higher. We can have a customer departure, which moves the state 1 lower. There's a possibility that nothing happens, in which case the state stays the same. And there's also the possibility of having simultaneously an arrival and a departure, in which case the state again stays the same.

So let's write some representative probabilities. If we have 2 customers, the probability that during this step we go down, this is the probability that we have a service completion, but to no customer arrival. So this is the probability associated with this transition. The other possibility is that there's a customer arrival, which happens with probability p, and we do not have a customer departure, and so the probability of that particular transition is this number.

And then finally, the probability that we stay in the same state, this can happen in 2 possible ways. One way is that we have an arrival and a departure simultaneously. And the other possibility is that we have no arrival and no departure, so that the state stays the same. So these transition probabilities would be the same starting from any other states, state 3, or state 9, and so on. Transition probabilities become a little different at the borders, at the boundaries of this diagram, because if you're in a state 0, then you cannot have any customer departures.

There's no one to be served, but there is a probability p that the customer arrives, in which case the number of customers in the system goes to 1. Then probability 1-p, nothing happens. Similarly with departures, if the system is full, there's no room for another arrival. But we may have a departure that happens with probability q, and nothing happens with probability 1-q. So this is the full transition diagram annotated with transition probabilities.

And this is a complete description of a discrete time, finite state Markov chain. So this is a complete probabilistic model. Once you have all of these pieces of information, you can start calculating things, and trying to predict what's going to happen in the future. Now let us abstract from this example and come up with a more general definition. So we have this concept of the state which describes the current situation in the system that we're looking at.

The current state is random, so we're going to think of it as a random variable Xn is the state, and transitions after the system started operating. So the system starts operating at some initial state X0, and after n transitions, it moves to state Xn. Now we have a set of possible states. State 1 state 2, state 3, and in general, state i and state j. To keep things simple, we assume that the set of possible states is a finite set.

As you can imagine, we can have systems in which the state space is going to be infinite. It could be discrete, or continuous. But all that is more difficult and more complicated. It makes sense to start from the simplest possible setting where we just deal with the finite state space.

And time is discrete, so we can think of this state in the beginning, after 1 transition, 2 transitions, and so on. So we're in discrete time and we have finite in many states. So the system starts somewhere, and at every time step, the state is, let's say, here. A whistle blows, and the state jumps to a random next state. So it may move here, or it may move there, or it may move here, or it might stay in the place. So one possible transition is the transition before you jump, and just land in the same place where you started from.

Now we want to describe the statistics of these transitions. If I am at that state, how likely is it to that, next time, I'm going to find myself at that state? Well, we describe the statistics of this transition by writing down a transition probability, the transition probability of going from state 3 to state 1. So this transition probability is to be thought of as a conditional probability. Given that right now I am at state i what is the probability that next time I find myself at state j? So given that right now I am at state 3, P31 is the probability that the next time I'm going to find myself at state 1.

Similarly here, we would have a probability P3i, which is the probability that given that right now I'm at state 3, next time I'm going to find myself at state i. Now one can write such conditional probabilities down in principle, but we need to make-- so you might think of this as a definition here, but we need to make one additional big assumption, and this is the assumption that to make a process to be a Markov process. This is the so-called Markov property, and here's what it says. Let me describe it first in words here.

Every time that I find myself at state 3, the probability that next time I'm going to find myself at state 1 is this particular number, no matter how I got there. That is, this transition probability is not affected by the past of the process. It doesn't care about what path I used to find myself at state 3.

Mathematically, it means the following. You have this transition probability that from state i jump to state j. Suppose that I gave you some additional information, that I told you everything else that happened in the past of the process, everything that happened, how did you get to state i? The assumption we're making is that this information about the past has no bearing in making predictions about the future, as long as you know where you are right now. So if I tell you, right now, you are at state i, and by the way, you got there by following a particular path, you can ignore the extra information of the particular path that you followed.

You only take into account where you are right now. So every time you find yourself at that state, no matter how you got there, you will find yourself next time at state 1 with probability P31. So the past has no bearing into the future, as long as you know where you are sitting right now. For this property to happen, you need to choose your state carefully in the right way. In that sense, the states needs to include any information that's relevant about the future of the system. Anything that's not in the state is not going to play a role, but the state needs to have all the information that's relevant in determining what kind of transitions are going to happen next.

So to take an example, before you go to Markov process, just from the deterministic world, if you have a ball that's flying up in the air, and you want to make predictions about the future. If I tell you that the state of the ball is the position of the ball at the particular time, is that enough for you to make predictions where the ball is going to go next? No. You need to know both the position and the velocity. If you know position and velocity, you can make predictions about the future.

So the state of a ball that's flying is position together with velocity. If you were to just take position, that would not be enough information, because if I tell you current position, and then I tell you past position, you could use the information from the past position to complete the trajectory and to make the prediction. So information from the past is useful if you don't know the velocity. But if both position and velocity, you don't care how you got there, or what time you started. From position and velocity, you can make predictions about the future.

So there's a certain art, or a certain element of thinking, a non-mechanical aspect into problems of this kind, to figure out which is the right state variable. When you define the state of your system, you need to define it in such a way that includes all information that has been accumulated that has some relevance for the future. So the general process for coming up with a Markov model is to first make this big decision of what your state variable is going to be. Then you write down if it may be a picture of the different states. Then you identify the possible transitions.

So sometimes the diagram that you're going to have will not include all the possible arcs. You would only show those arcs that correspond to transitions that are possible. For example, in the supermarket example, we did not have a transition from state 2 to state 5, because that cannot happen. You can only have 1 arrival at any time. So in the diagram, we only showed the possible transitions.

And for each of the possible transitions, then you work with the description of the model to figure out the correct transition probability. So you got the diagram by writing down transition probabilities.

OK, so suppose you got your Markov model. What will you do with it? Well, what do we need models for? We need models in order to make predictions, to make probabilistic predictions. So for example, I tell you that the process started in that state. You let it run for some time. Where do you think it's going to be 10 time steps from now? That's a question that you might want to answer.

Since the process is random, there's no way for you to tell me exactly where it's going to be. But maybe you can give me probabilities. You can tell me, with so much probability, the state would be there. With so much probability, the state would be there, and so on. So our first exercise is to calculate those probabilities about what may happen to the process a number of steps in the future. It's handy to have some notation in here.

So somebody tells us that this process starts at the particular state i. We let the process run for n transitions. It may land at some state j, but that state j at which it's going to land is going to be random. So we want to give probabilities. Tell me, with what probability the state, n times steps later, is going to be that particular state j?

The shorthand notation is to use this symbol here for the n-step transition probabilities that you find yourself at state j given that you started at state i. So the way these two indices are ordered, the way to think about them is that from i, you go to j. So the probability that from i you go to j if you have n steps in front of you. Some of these transition probabilities are, of course easy to write. For example, in 0 transitions, you're going to be exactly where you started. So this probability is going to be equal to 1 if i is equal to j, And 0 if i is different than j.

That's an easy one to write down. If you have only 1 transition, what's the probability that 1 step later you find yourself in state j given that you started at state i? What is this? These are just the ordinary 1-step transition probabilities that we are given in the description of the problem. So by definition, the 1-step transition probabilities are of this form. This equality is correct just because of the way that we defined those two quantities.

Now we want to say something about the n-step transition probabilities when n is a bigger number. OK. So here, we're going to use the total probability theorem. So we're going to condition in two different scenarios, and break up the calculation of this quantity, by considering the different ways that this event can happen. So what is the event of interest? The event of interest is the following. At time 0 we start i. We are interested in landing at time n at the particular state j.

Now this event can happen in several different ways, in lots of different ways. But let us group them into subgroups. One group, or one sort of scenario, is the following. During the first n-1 time steps, things happen, and somehow you end up at state 1. And then from state 1, in the next time step you make a transition to state j. This particular arc here actually corresponds to lots and lots of different possible scenarios, or different spots, or different transitions. In n-1 time steps, there's lots of possible ways by which you could end up at state 1. Different paths through the state space.

But all of them together collectively have a probability, which is the (n-1)-step transition probability, that from state i, you end up at state 1

And then there's other possible scenarios. Perhaps in the first n-1 time steps, you follow the trajectory that took you at state m. And then from state m, you did this transition, and you ended up at state j. So this diagram breaks up the set of all possible trajectories from i to j into different collections, where each collection has to do with which one happens to be the state just before the last time step, just before time n. And we're going to condition on the state at time n-1.

So the total probability of ending up at state j is the sum of the probabilities of the different scenarios -- the different ways that you can get to state j. If we look at that type of scenario, what's the probability of that scenario happening? With probability Ri1(n-1), I find myself at state 1 at time n-1. This is just by the definition of these multi-step transition probabilities. This is the number of transitions.

The probability that from state i, I end up at state 1. And then given that I found myself at state 1, with probability P1j, that's the transition probability, next time I'm going to find myself at state j. So the product of these two is the total probability of my getting from state i to state j through state 1 at the time before. Now where exactly did we use the Markov assumption here? No matter which particular path we used to get from i to state 1, the probability that next I'm going to make this transition is that same number, P1j.

So that number does not depend on the particular path that I followed in order to get there. If we didn't have the Markov assumption, we should have considered all possible individual trajectories here, and then we would need to use the transition probability that corresponds to that particular trajectory. But because of the Markov assumption, the only thing that matters is that right now we are at state 1. It does not matter how we got there.

So now once you see this scenario, then this scenario, and that scenario, and you add the probabilities of these different scenarios, you end up with this formula here, which is a recursion. It tells us that once you have computed the (n-1)-step transition probabilities, then you can compute also the n-step transition probabilities. This is a recursion that you execute or you run for all i's and j's simultaneously. That is fixed. And for a particular n, you calculate this quantity for all possible i's, j's, k's. You have all of those quantities, and then you use this equation to find those numbers again for all the possible i's and j's.

Now this is formula which is always true, and there's a big idea behind the formula. And now there's variations of this formula, depending on whether you're interested in something that's slightly different. So for example, if you were to have a random initial state, somebody gives you the probability distribution of the initial state, so you're told that with probability such and such, you're going to start at state 1. With that probability, you're going to start at state 2, and so on. And you want to find the probability at the time n you find yourself at state j.

Well again, total probability theorem, you condition on the initial state. With this probability you find yourself at that particular initial state, and given that this is your initial state, this is the probability that n time steps later you find yourself at state j. Now building again on the same idea, you can run every recursion of this kind by conditioning at different times. So here's a variation. You start at state i. After 1 time step, you find yourself at state 1, with probability pi1, and you find yourself at state m with probability Pim. And once that happens, then you're going to follow some trajectories. And there is a possibility that you're going to end up at state j after n-1 time steps.

This scenario can happen in many possible ways. There's lots of possible paths from state 1 to state j. There's many paths from state 1 to state j. What is the collective probability of all these transitions? This is the event that, starting from state 1, I end up at state j in n-1 time steps. So this one has here probability R1j of n-1. And similarly down here. And then by using the same way of thinking as before, we get the formula that Rij(n) is the sum over all k's of Pik, and then the Rkj(n-1).

So this formula looks almost the same as this one, but it's actually different. The indices and the way things work out are a bit different, but the basic idea is the same. Here we use the total probability theory by conditioning on the state just 1 step before the end of our time horizon. Here we use total probability theorem by conditioning on the state right after the first transition. So this generally idea has different variations. They're all valid, and depending on the context that you're dealing with, you might want to work with one of these or another.

So let's illustrate these calculations in terms of an example. So in this example, we just have 2 states, and somebody gives us transition probabilities to be those particular numbers. Let's write down the equations. So the probability that starting from state 1, I find myself at state 1 n time steps later. This can happen in 2 ways. At time n-1, I might find myself at state 2. And then from state 2, I make a transition back to state 1, which happens with probability-- why'd I put 2 there -- anyway, 0.2. And another way is that from state 1, I go to state 1 in n-1 steps, and then from state 1 I stay where I am, which happens with probability 0.5.

So this is for R11(n). Now R12(n), we can write a similar recursion for this one. On the other hand, seems these are probabilities. The state at time n is going to be either state 1 or state 2. So these 2 numbers need to add to 1, so we can just write this as 1 - R11(n). And this is an enough of a recursion to propagate R11 and R12 as time goes on. So after n-1 transitions, either I find myself in state 2, and then there's a point to transition that I go to 1, or I find myself in state 1, which with that probability, and from there, I have probability 0.5 of staying where I am.

Now let's start calculating. As we discussed before, if I start at state 1, after 0 transitions I'm certain to be at state , and I'm certain not to be at state 1. If I start from state 1, I'm certain to not to be at state at that time, and I'm certain that I am right now, it's state 1. After I make transition, starting from state 1, there's probability 0.5 that I stay at state 1. And there's probability 0.5 that I stay at state 2. If I were to start from state 2, the probability that I go to 1 in 1 time step is this transition that has probability 0.2, and the other 0.8.

OK. So the calculation now becomes more interesting, if we want to calculate the next term. How likely is that at time 2, I find myself at state 1? In order to be here at state 1, this can happen in 2 ways. Either the first transition left me there, and the second transition is the same. So these correspond to this 0.5, that the first transition took me there, and the next transition was also of the same kind. That's one possibility. But there's another scenario. In order to be at state 1 at time 2 -- this can also happen this way. So that's the event that, after 1 transition, I got there. And the next transition happened to be this one.

So this corresponds to 0.5 times 0.2. It corresponds to taking the 1-step transition probability of getting there, times the probability that from state 2 I move to state 1, which in this case, is 0.2. So basically we take this number, multiplied with 0.2, and then add those 2 numbers. And after you add them, you get 0.35. And similarly here, you're going to get 0.65.

And now to continue with the recursion, we keep doing the same thing. We take this number times 0.5 plus this number times 0.2. Add them up, you get the next entry. Keep doing that, keep doing that, and eventually you will notice that the numbers start settling into a limiting value at 2/7. And let's verify this. If this number is 2/7, what is the next number going to be? The next number is going to be 2/7 -- (not 2.7) -- it's going to be 2/7. That's the probability that I find myself at that state, times 0.5-- that's the next transition that takes me to state 1 -- plus 5/7-- that would be the remaining probability that I find myself in state 2 -- times 1/5. And so that gives me, again, 2/7.

So this calculation basically illustrates, if this number has become 2/7, then the next number is also going to be 2/7. And of course this number here is going to have to be 5/7. And this one would have to be again, the same, 5/7. So the probability that I find myself at state 1, after a long time has elapsed, settles into some steady state value. So that's an interesting phenomenon. We just make this observation.

Now we can also do the calculation about the probability, starting from state 2. And here, you do the calculations -- I'm not going to do them. But after you do them, you find this probability also settles to 2/7 and this one also settles to 5/7. So these numbers here are the same as those numbers. What's the difference between these?

This is the probability that I find myself at state 1 given that I started at 1. This is the probability that I find myself at state 1 given that I started at state 2. These probabilities are the same, no matter where I started from. So this numerical example sort of illustrates the idea that after the chain has run for a long time, what the state of the chain is, does not care about the initial state of the chain.

So if you start here, you know that you're going to stay here for some time, a few transitions, because this probability is kind of small. So the initial state does that's tell you something. But in the very long run, transitions of this kind are going to happen. Transitions of that kind are going to happen. There's a lot of randomness that comes in, and that randomness washes out any information that could come from the initial state of the system.

We describe this situation by saying that the Markov chain eventually enters a steady state. Where a steady state, what does it mean it? Does it mean the state itself becomes steady and stops at one place? No, the state of the chain keeps jumping forever. The state of the chain will keep making transitions, will keep going back and forth between 1 and 2. So the state itself, the Xn, does not become steady in any sense.

What becomes steady are the probabilities that describe Xn. That is, after a long time elapses, the probability that you find yourself at state 1 becomes a constant 2/7, and the probability that you find yourself in state 2 becomes a constant. So jumps will keep happening, but at any given time, if you ask what's the probability that right now I am at state 1, the answer is going to be 2/7.

Incidentally, do the numbers sort of makes sense? Why is this number bigger than that number? Well, this state is a little more sticky than that state. Once you enter here, it's kind of harder to get out. So when you enter here, you spend a lot of time here. This one is easier to get out, because the probability is 0.5, so when you enter there, you tend to get out faster. So you keep moving from one to the other, but you tend to spend more time on that state, and this is reflected in this probability being bigger than that one. So no matter where you start, there's 5/7 probability of being here, 2/7 probability being there.

So there were some really nice things that happened in this example. The question is, whether things are always as nice for general Markov chains. The two nice things that happened where the following-- as we keep doing this calculation, this number settles to something. The limit exists. The other thing that happens is that this number is the same as that number, which means that the initial state does not matter. Is this always the case? Is it always the case that as n goes to infinity, the transition probabilities converge to something?

And if they do converge to something, is it the case that the limit is not affected by the initial state i at which the chain started? So mathematically speaking, the question we are raising is whether Rij(n) converges to something. And whether that something to which it converges to has only to do with j. It's the probability that you find yourself at state j, and that probability doesn't care about the initial state. So it's the question of whether the initial state gets forgotten in the long run.

So the answer is that usually, or for nice chains, both of these things will be true. You get the limit which does not depend on the initial state. But if your chain has some peculiar or unique structure, this might not happen. So let's think first about the issue of convergence. So convergence, as n goes to infinity at a steady value, really means the following. If I tell you a lot of time has passed, then you tell me, OK, the state of the probabilities are equal to that value without having to consult your clock. If you don't have convergence, it means that Rij can keep going up and down, without settling to something. So in order for you to tell me the value of Rij, you need to consult your clock to check if, right now, it's up or is it down.

So there's some kind of periodic behavior that you might get when you do not get convergence, and this example here illustrates it. So what's happened in this example? Starting from state 2, next time you go here, or there, with probability half. And then next time, no matter where you are, you move back to state 2. So this chain has some randomness, but the randomness is kind of limited type. You go out, you come in. You go out, you come in. So there's a periodic pattern that gets repeated. It means that if you start at state 2 after an even number of steps, you are certain to be back at state 2. So this probability here is 1.

On the other hand, if the number of transitions is odd, there's no way that you can be at your initial state. If you start here, at even times you would be here, at odd times you would be there or there. So this probability is 0. As n goes to infinity, these probabilities, the n-step transition probability does not converge to anything. It keeps alternating between 0 and 1. So convergence fails.

This is the main mechanism by which convergence can fail if your chain has a periodic structure. And we're going to discuss next time that, if periodicity absent, then we don't have an issue with convergence. The second question if we have convergence, whether the initial state matters or not. In the previous chain, where you could keep going back and forth between states 1 and 2 numerically, one finds that the initial state does not matter. But you can think of situations where the initial state does matter. Look at this chain here.

If you start at state 1, you stay at state 1 forever. There's no way to escape. So this means that R11(n) is 1 for all n. If you start at state 3, you will be moving between stage 3 and 4, but there's no way to go in that direction, so there's no way that you go to state 1. And for that reason, R31 is 0 for all n.

OK So this is a case where the initial state matters. R11 goes to a limit, as n goes to infinity, because it's constant. It's always 1 so the limit is 1. R31 also has a limit. It's 0 for all times. So these are the long term probabilities of finding yourself at state 1. But those long-term probabilities are affected by where you started. If you start here, you're sure that's, in the long term, you'll be here. If you start here, you're sure that, in the long term, you will not be there. So the initial state does matter here.

And this is a situation where certain states are not accessible from certain other states, so it has something to do with the graph structure of our Markov chain. Finally let's answer this question here, at least for large n's. What do you think is going to happen in the long term if you start at state 2? If you start at state 2, you may stay at state 2 for a random amount of time, but eventually this transition will happen, or that transition would happen. Because of the symmetry, you are as likely to escape from state 2 in this direction, or in that direction, so there's probability 1/2 that, when the transition happens, the transition happens in that direction.

So for large N, you're certain that the transition does happen. And given that the transition has happened, it has probability 1/2 that it has gone that particular way. So clearly here, you see that the probability of finding yourself in a particular state is very much affected by where you started from. So what we want to do next is to abstract from these two examples and describe the general structural properties that have to do with periodicity, and that have to do with what happened here with certain states, not being accessible from the others.

We're going to leave periodicity for next time. But let's talk about the second kind of phenomenon that we have. So here, what we're going to do is to classify the states in a transition diagram into two types, recurrent and transient. So a state is said to be recurrent if the following is true. If you start from the state i, you can go to some places, but no matter where you go, there is a way of coming back. So what's an example for the recurrent state? This one. Starting from here, you can go elsewhere. You can go to state 7. You can go to state 6. That's all where you can go to. But no matter where you go, there is a path that can take you back there.

So no matter where you go, there is a chance, and there is a way for returning where you started. Those states we call recurrent. And by this, 8 is recurrent. All of these are recurrent. So this is recurrent, this is recurrent. And this state 5 is also recurrent. You cannot go anywhere from 5 except to 5 itself. Wherever you can go, you can go back to where you start. So this is recurrent.

If it is not the recurrent, we say that it is transient. So what does transient mean? You need to take this definition, and reverse it. Transient means that, starting from i, there is a place to which you could go, and from which you cannot return. If it's recurrent, anywhere you go, you can always come back. Transient means there are places where you can go from which you cannot come back.

So state 1 is recurrent - because starting from here, there's a possibility that you get there, and then there's no way back. State 4 is recurrent, starting from 4, there's somewhere you can go and-- sorry, transient, correct. State 4 is transient starting from here, there are places where you could go, and from which you cannot come back. And in this particular diagram, all these 4 states are transients.

Now if the state is transient, it means that there is a way to go somewhere where you're going to get stuck and not to be able to come. As long as your state keeps circulating around here, eventually one of these transitions is going to happen, and once that happens, then there's no way that you can come back. So that transient state will be visited only a finite number of times. You will not be able to return to it. And in the long run, you're certain that you're going to get out of the transient states, and get to some class of recurrent states, and get stuck forever.

So, let's see, in this diagram, if I start here, could I stay in this lump of states forever? Well as long as I'm staying in this type of states, I would keep visiting states 1 and 2 Each time that I visit state 2, there's going to be positive probability that I escape. So in the long run, if I were to stay here, I would visit state 2 an infinite number of times, and I would get infinite chances to escape. But if you have infinite chances to escape, eventually you will escape. So you are certain that with probability 1, starting from here, you're going to move either to those states, or to those states.

So starting from transient states, you only stay at the transient states for random but finite amount of time. And after that happens, you end up in a class of recurrent states. And when I say class, what they mean is that, in this picture, I divide the recurrent states into 2 classes, or categories. What's special about them? These states are recurrent. These states are recurrent. But there's no communication between the 2. If you start here, you're stuck here. If you start here, you are stuck there.

And this is a case where the initial state does matter, because if you start here, you get stuck here. You start here, you get stuck there. So depending on the initial state, that's going to affect the long term behavior of your chain. So the guess you can make at this point is that, for the initial state to not matter, we should not have multiple recurrent classes. We should have only 1. But we're going to get back to this point next time.