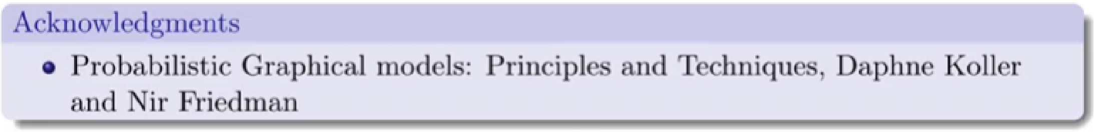
Lecture - 17.1

Markov Networks: Motivation

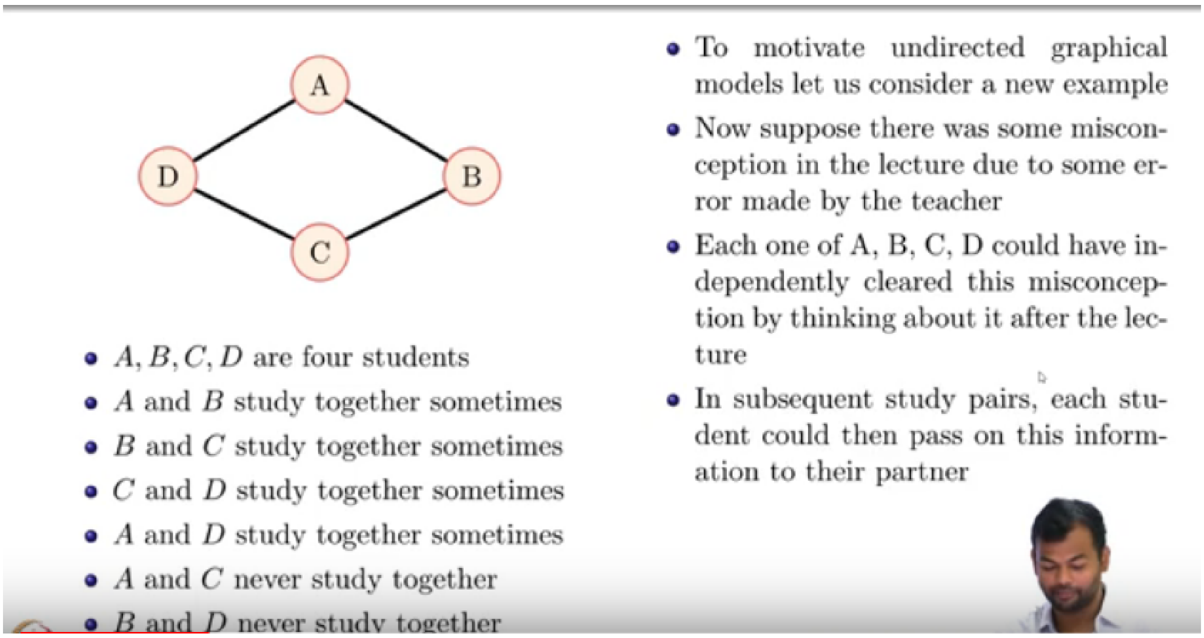
Markov networks, which is so, so what we saw? So for is, known as, Bayesian networks or directed graphical models, what we're going to see now, are known, as Markov networks or undirected graphical models. Okay?

Refer slide time :(0:28)



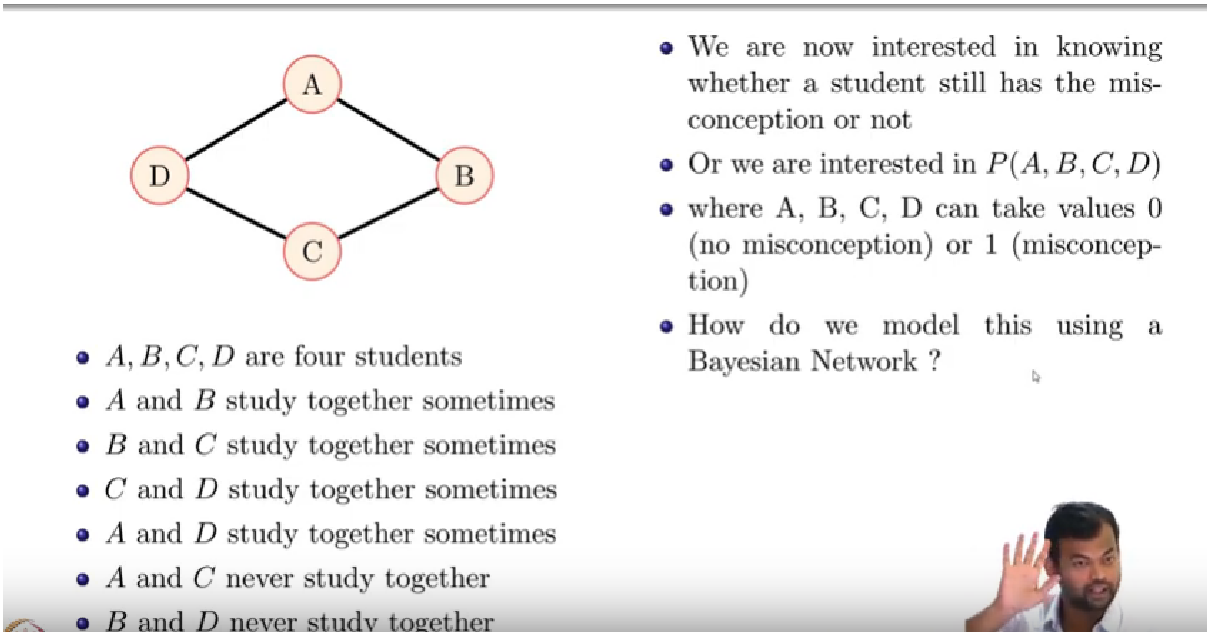
So let's start with that, so a lot of this material and even the previous lecture have been taken from this text book, from the relevant chapters.

Refer slide time :(0:38)



So to motivate these, undirected graphical models or the Markov networks Right? Let us consider a new example, now so we have dealt with the student example. Now, we look at a different set of students, so suppose this is the situation that I have, I have four students A, B, C, D .Okay? A and B study together sometimes, A and B and C study together sometimes, D, N, C and a and d .Okay? so, every as indicates whether these two students ,study together or not, for now for whatever reason I am NOT drawing directions on these edges .Okay? And if you don't see an edge, I mean that these students never study together, for whatever reason, it’s A and C we see don't study together and B and D don't study together. Now, let us consider this situation that all of these are attending some lectures and there was some misconception in class, because of some mistake with the instructor made .Okay? And so there's some concept, which everyone in the class has not understood properly. Now, once they go back and being good students, they go back and read the material, lecture, slides and other things and also think about the subject, instead of doing other things, so in the course of time they of course, may or may not clear this minkins, misconception .Right? because, I would have, I mean, not I but some other instructor of those, would have made some mistake and then, they went and thought about it and they realized oh this is not correct, this is incorrect and that misconception again is a random variable because, it could or could not have been clear .Okay? Now, what happens after that? is that, these people are also going to now study together, so each of them independently thought of the problem and may be or may not clear the misconception .but, now when they study together, now this is further chance, that if one of those students in the group, had cleared the misconception, then maybe the other students misconception would also be clear. Because, hopefully and desirably that student passes on this information to the other student, there is no relative grading

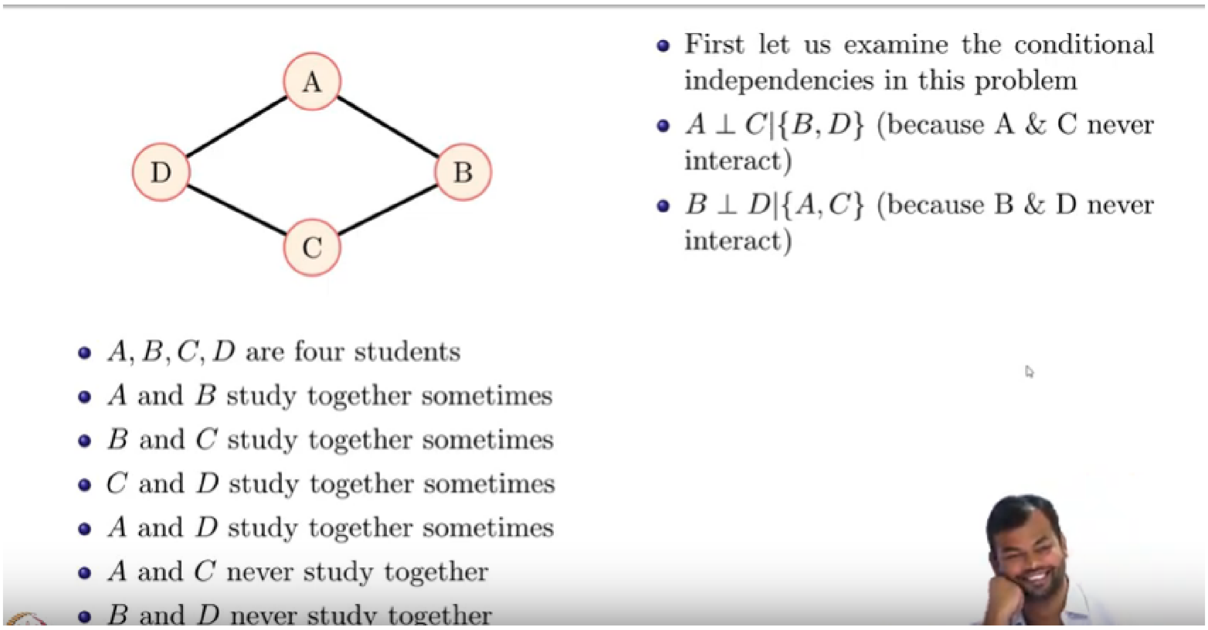
Refer slide time :(2:37)



And they're, now, interested in knowing, whether a student still has the misconception .Okay? And we are no interest in knowing this for all the four students. So, what's the joint distribution that we are interested in? And, what are the random variables that we have? We are interested in knowing for every student, whether that student still has the misconception or not? So, what are the random variables and even, so think of that diagram that we have, your set is students and you want a random variable to right? So, when what we'll do is, we'll have this random variable which A, B, C, D? Which tells us whether student a has a misconception or not, student B has a misconception or not, students C as a misconception or not and student D has a misconception or is that fine .Okay? Everyone is fine with the setup .Okay? Now, what's the joint distribution, that I'm interested in, p of a

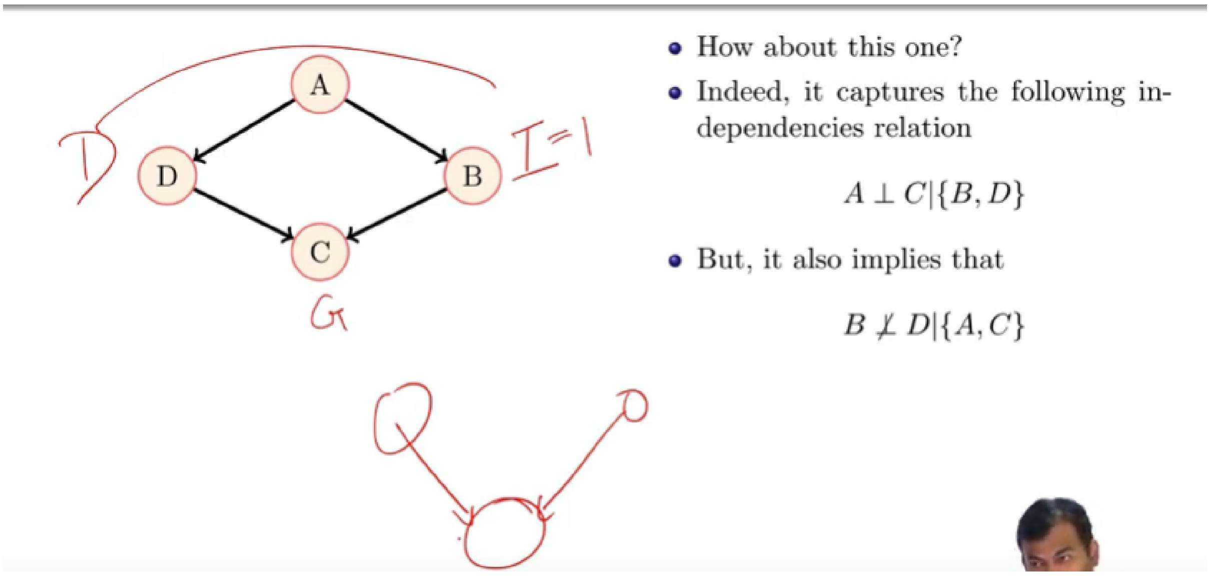
gamma, B gamma, C gamma d? Right? That's the joint distribution, that I am interested, in .Okay? Fine, now, A, B, C, D can take on values, 0 for no misconception and one for misconception. So, it's slightly no need to do, but one means, there is a misconception .Okay? How do we model this using a patient? So, now why this question? Because, you can see that there is some dependence. Right? Just as, grade dependent on the recommend of sorry, the recommendation letter dependent on the grade and so on. Here whether, D would have a misconception or not? Depends on whether D has resolved it himself or herself and also whether any of their partners have resolved it Right?
So there is dependence between the partners also, so how will you, represent this joint distribution. now ,here's where the discussion of I maps, helps, write that I am asking you, that given a joint distribution ,so I have not given you the table ,but I have given you an intuition about, what kind of in dependencies exist in this joint distribution? What is the Independence's which exist in this joint distribution? Is, independent of C, is a independent of C, given B or D. Right? So, he is independent of C given B and D, what is the other independence which holds? Yeah. So, I don't need to actually give you a Table, I just gave you the problem setting and from that problem setting you are able to make these assumptions that these Independence's hold in the network .Right? So, you see that what happens in an IMAP, you don't need the table, you can just understand your world and based on that make some assumptions.

Refer slide time :(5:34)



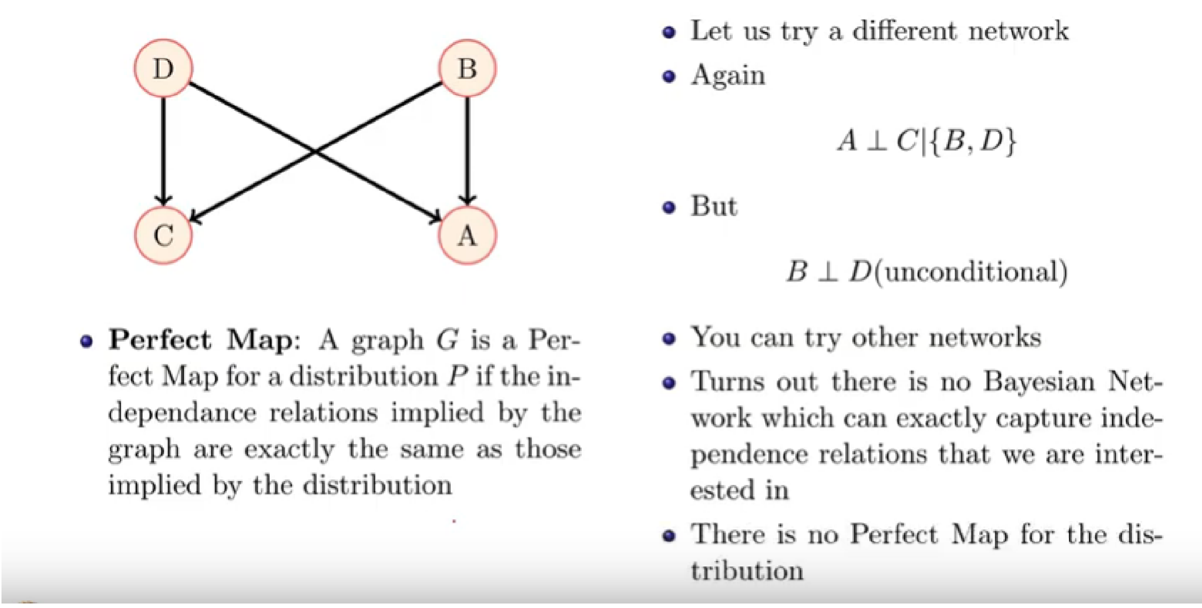
So, that's the first thing that you will do. Right? whenever I ask you to draw a Bayesian network, the first thing that you will do is, think about what are the independence assumptions which hold in the distribution ,either based on data which is this table, but most case ,most likely that will not be given to you or just your knowledge about the world. Right? So here, you are using your knowledge about the world and you are telling me, that these two independence is actually hold in the distribution. Right? Now, our job is simplified, what do we need to do now? This, draw Bayesian network, which encodes these independent sirs, that's what it means Right? I am asking you for an IMAP that means I just want a Bayesian network, which can encode these Independence's. So, this does this quickly draws that mission Network, how many nodes will that Bayesian network have? 4. Okay? So let's, just quickly draw it, how many of you are not falling for this? Okay good .so let's, try Right?

Refer slide time :(6:32)



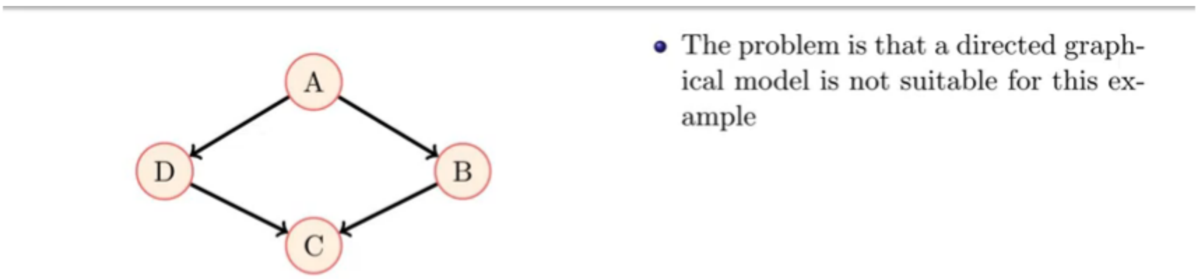
We try this one network, what are the independence is that it encodes? if now, bring a Bayesian network. Right? That's what I wanted to draw, what is the independence that it encodes? A independent of C given B, D .so, half of the job is done. Right? What the other thing which I needed, B's of D ,does that happen, B is independent of D given a gamma C, is that what I wanted Right? But, it implies that B is not independent of D given a Gamma C. Because, of what of what caused? Causal reasoning, aberrational reasoning or explaining away. If I know C, can be and D be independent of each other, so this think of it again as, this was the grade in the course, this was the student's intelligence and suppose this was the difficulty of the course. Right?
Now, if I know that the grade was bad, D and I cannot be independent of each other. Right? Because, if I know that, the grade was bad and if I know that, I was high, and then can I set the probability of D to be low to high. Right? Because, if the student was intelligent and if the grade is bad, then it has to mean that the course was difficult. Right? So, that's why I and D now no longer can be independent, if I know the value of I, there will be some belief of about D which will change, if I have given G. Right? So, that's the similar situation, so whenever you have this kind of a structure, where you have two parents and a child .Okay? There's always be some dependence between parents, in real life also .Okay? So, B is independent of D given a Gamma C that is the additional independence which is getting encoded, which we didn't want.

Refer slide time :(8:35)



So, then maybe let us try a different network. Right? How about this one? What are the independence is added encodes? And the second one, does it satisfy, we want B to be independent of D, given a Gamma C, does that happen. What is the independence that it encodes actually? B is independent of the D, unconditional a Respecter, whether I know a or C these are like a root nodes in the original graph. Right? I and D or whatever it was, they don't depend on anything else, so they are always going to be independent. Right? is that fine. so now ,you can go back and try any but Marco, any Bayesian network, that you want to draw and you will not be able to come up with the Bayesian network, which encodes these two exact set of Independence's that you want Right?
So, that's a problem .Right? So because, for this distribution, you don't really have a perfect map, so what's a perfect map? A perfect map is an eye map, which has exactly the same independence relations as implied in the distribution. So instead of that subset you'll have a equal to. Right? Now ,you cannot come up with a perfect match for this distribution ,this is just 4 random variables and you would have thought that I mean with 4 if I can't do this imagine if I have 20 or 30 or 100 or million or whatever. Right? So, that's why, Bayesian networks are not a solution for all the kind of in dependencies or all the kind of distributions that you might want to encode. Right? There are certain distributions, for which you cannot really come up with an equivalent Bayesian network or at least a perfect Bayesian network .Okay?

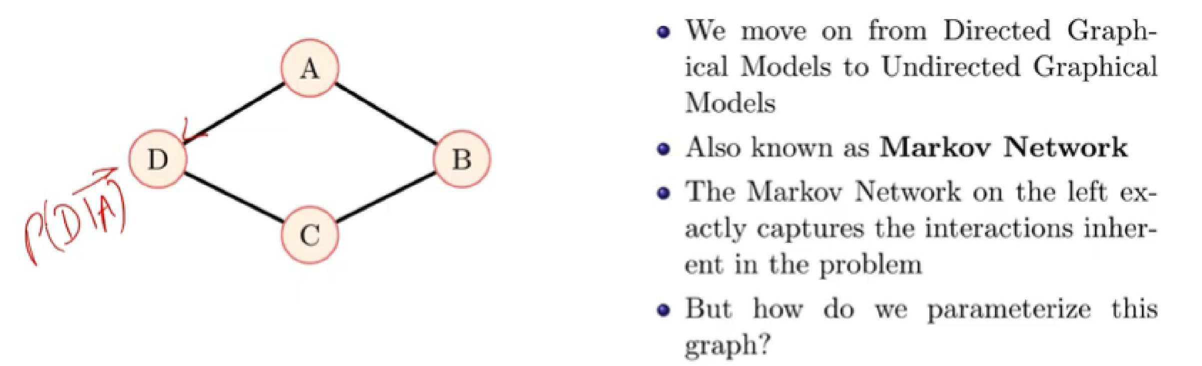
Refer slide time :(10:28)



Now, the problem here is that, what's the problem here actually? Why is a directed graphical model not suitable for this example? What is the problem here? Yeah, there's no sense of direction here

Right? Because you cannot say that a depends on D or D depends on A, it they both are equal partners. Right? Both of them study together, both of them contribute, equally to the discussion and so on .Right? So you can't really say that, one depends on the other. Right? So, I directed edge between two nodes is, not meaningful in this case. Right? And, what we actually want is not this direction? What we want is to capture the strength of the interaction between two. Right? So now, a and B may study together. But, maybe when they study, they typically, end up playing video games or whatever. Right? So, they probably have a much smaller Association, as compared to B and C, a B and C for example, they might have a stronger Association or things like that. Right? So, what we want to capture is these strengths of the association between two random variables is? What we desire and these kinds of distributions? Does that make sense .Okay?

Refer slide time :(11:46)



So now, let's see, so what we'll do is we'll move from directed graphical models to undirected graphical models? And, what's our aim? What's the basic aim that we have? When we are coming up with graphical models? What's our, one single most desired? Everyone, everyone, everyone come on, Factorization. Right? We want to come up with a simplified factorization of the Joint Distribution. Right? So we want to every Bayesian network was associated with what? Come on, independence assumptions and what else what did he have associated with the Bayesian network? The conditioners and the marginal's, which were the factors of the Joint Distribution. Right? So, our aim of coming up with a network, weather, Markov or Bayesian is to get a hold of these factors which are associated with the graph, that's what we care about, because once we have these factors, we want to believe that the Joint Distribution can be expressed as a factorization over this graph, that's what our aim is .Okay? so let's , be clear about that, that's what we ultimately want, we are not just interest, interest drying grass for the sake of it ,we want to eventually get to a factorization of the joint description ,that's the thing that we care about .Okay? So this is also known as, a Markov network and this monitor and exactly captures the interactions that we had Right? so we are not asking, the questions of Independence yet. But, what we are saying is that these are the interactions or these are the dependencies, if you may between the different random variables in the Marko network. Now, how do we parameterize this graph? Now, what does this question mean? If I had asked you this question, for a Bayesian network. How do we parameterize this graph? What would your answer be? The conditional parameterization. Right? So remember all this, we moved from the joint parameterization, to a conditional parameterization, when we are doing Bayesian networks. Right? So there, the answer was straightforward, that the parameters associated with the Bayesian network, are the conditionals or the marginal's or we were calling them as local probability distributions, for every node in the graph. Right? And in most cases, these local priority distributions were conditional probability distributions, except for the root, guys where we did not have a conditional property distribution.

So now, in the case of a mark of network or an undirected graphical model, what? So, I'm going to use these terms interchangeably, factors, parameters, conditional distributions and so on. What are the factors associated with the graph? Factors, parameters, conditional Distribution. So what do we need to have with every node? So, if this, were a directed graph, what would I have here? I would have the distribution D given a Right? Now, I'm asking you, what should I have in the case, when this is not a directed graph, I can again, have a conditional Distribution. Right? Why not? I can again write P of D given in A, how many if you say we can have a joint conditional distribution .Okay? Only a few honest people, only if you say we can't have a conditional distribution. So, I got two answers, for the first question and two answers, for the second question. I just don't understand the concept of a universal said .Okay?