

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at [ocw.mit.edu](http://ocw.mit.edu)

**JOHN TSISIKLIS:** So here's the agenda for today. We're going to do a very quick review. And then we're going to introduce some very important concepts. The idea is that all information is-- Information is always partial. And the question is what do we do to probabilities if we have some partial information about the random experiments. We're going to introduce the important concept of conditional probability. And then we will see three very useful ways in which it is used. And these ways basically correspond to divide and conquer methods for breaking up problems into simpler pieces. And also one more fundamental tool which allows us to use conditional probabilities to do inference, that is, if we get a little bit of information about some phenomenon, what can we infer about the things that we have not seen?

So our quick review. In setting up a model of a random experiment, the first thing to do is to come up with a list of all the possible outcomes of the experiment. So that list is what we call the sample space. It's a set. And the elements of the sample space are all the possible outcomes. Those possible outcomes must be distinguishable from each other. They're mutually exclusive. Either one happens or the other happens, but not both. And they are collectively exhaustive, that is no matter what the outcome of the experiment is going to be an element of the sample space.

And then we discussed last time that there's also an element of art in how to choose your sample space, depending on how much detail you want to capture. This is usually the easy part. Then the more interesting part is to assign probabilities to our model, that is to make some statements about what we believe to be likely and what we believe to be unlikely. The way we do that is by assigning probabilities to subsets of the sample space. So as we have our sample space here, we may have a subset  $A$ . And we assign a number to that subset  $P(A)$ , which is the probability that this event happens. Or this is the probability that when we do the experiment and we get an outcome it's the probability that the outcome happens to fall inside that event.

We have certain rules that probabilities should satisfy. They're non-negative. The probability of

the overall sample space is equal to one, which expresses the fact that we're certain, no matter what, the outcome is going to be an element of the sample space. Well, if we set the top right so that it exhausts all possibilities, this should be the case.

And then there's another interesting property of probabilities that says that, if we have two events or two subsets that are disjoint, and we're interested in the probability, that one or the other happens, that is the outcome belongs to A or belongs to B. For disjoint events the total probability of these two, taken together, is just the sum of their individual probabilities. So probabilities behave like masses. The mass of the object consisting of A and B is the sum of the masses of these two objects. Or you can think of probabilities as areas. They have, again, the same property. The area of A together with B is the area of A plus the area B.

But as we discussed at the end of last lecture, it's useful to have in our hands a more general version of this additivity property, which says the following, if we take a sequence of sets--  $A_1$ ,  $A_2$ ,  $A_3$ ,  $A_4$ , and so on. And we put all of those sets together. It's an infinite sequence. And we ask for the probability that the outcome falls somewhere in this infinite union, that is we are asking for the probability that the outcome belongs to one of these sets, and assuming that the sets are disjoint, we can again find the probability for the overall set by adding up the probabilities of the individual sets.

So this is a nice and simple property. But it's a little more subtle than you might think. And let's see what's going on by considering the following example. We had an example last time where we take our sample space to be the unit square. And we said let's consider a probability law that says that the probability of a subset is just the area of that subset. So let's consider this probability law. OK.

Now the unit square is the set --let me just draw it this way-- the unit square is the union of one element set consisting all of the points. So the unit square is made up by the union of the various points inside the square. So union over all  $x$ 's and  $y$ 's. OK? So the square is made up out of all the points that this contains.

And now let's do a calculation. One is the probability of our overall sample space, which is the unit square. Now the unit square is the union of these things, which, according to our additivity axiom, is the sum of the probabilities of all of these one element sets. Now what is the probability of a one element set? What is the probability of this one element set? What's the probability that our outcome is exactly that particular point? Well, it's the area of that set, which

is zero. So it's just the sum of zeros. And by any reasonable definition the sum of zeros is zero. So we just proved that one is equal to zero.

OK. Either probability theory is dead or there is some mistake in the derivation that I did. OK, the mistake is quite subtle and it comes at this step. We're sort of applied the additivity axiom by saying that the unit square is the union of all those sets. Can we really apply our additivity axiom. Here's the catch. The additivity axiom applies to the case where we have a sequence of disjoint events and we take their union. Is this a sequence of sets? Can you make up the whole unit square by taking a sequence of elements inside it and cover the whole unit square? Well if you try, if you start looking at the sequence of one element points, that sequence will never be able to exhaust the whole unit square.

So there's a deeper reason behind that. And the reason is that infinite sets are not all of the same size. The integers are an infinite set. And you can arrange the integers in a sequence. But the continuous set like the unit square is a bigger set. It's so-called uncountable. It has more elements than any sequence could have. So this union here is not of this kind, where we would have a sequence of events. It's a different kind of union. It's a Union that involves a union of many, many more sets. So the countable additivity axiom does not apply in this case. Because, we're not dealing with a sequence of sets. And so this is the incorrect step.

So at some level you might think that this is puzzling and awfully confusing. On the other hand, if you think about areas of the way you're used to them from calculus, there's nothing mysterious about it. Every point on the unit square has zero area. When you put all the points together, they make up something that has finite area. So there shouldn't be any mystery behind it.

Now, one interesting thing that this discussion tells us, especially the fact that the single elements set has zero area, is the following-- Individual points have zero probability. After you do the experiment and you observe the outcome, it's going to be an individual point. So what happened in that experiment is something that initially you thought had zero probability of occurring. So if you happen to get some particular numbers and you say, "Well, in the beginning, what did I think about those specific numbers? I thought they had zero probability. But yet those particular numbers did occur."

So one moral from this is that zero probability does not mean impossible. It just means extremely, extremely unlikely by itself. So zero probability things do happen. In such

continuous models, actually zero probability outcomes are everything that happens. And the bumper sticker version of this is to always expect the unexpected. Yes?

**AUDIENCE:** [INAUDIBLE].

**JOHN TSISIKLIS:** Well, probability is supposed to be a real number. So it's either zero or it's a positive number. So you can think of the probability of things just close to that point and those probabilities are tiny and close to zero. So that's how we're going to interpret probabilities in continuous models. But this is two chapters ahead. Yeah?

**AUDIENCE:** How do we interpret probability of zero? If we can use models that way, then how about probability of one? That it's extremely likely but not necessarily for certain?

**JOHN TSISIKLIS:** That's also the case. For example, if you ask in this continuous model, if you ask me for the probability that  $x, y$ , is different than the zero, zero this is the whole square, except for one point. So the area of this is going to be one. But this event is not entirely certain because the zero, zero outcome is also possible. So again, probability of one means essential certainty. But it still allows the possibility that the outcome might be outside that set. So these are some of the weird things that are happening when you have continuous models. And that's why we start to this class with discrete models, on which would be spending the next couple of weeks.

**OK.** So now once we have set up our probability model and we have a legitimate probability law that has these properties, then the rest is usually simple. Somebody asks you a question of calculating the probability of some event. While you were told something about the probability law, such as for example the probabilities are equal to areas, and then you just need to calculate. In these type of examples somebody would give you a set and you would have to calculate the area of that set. So the rest is just calculation and simple.

Alright, so now it's time to start with our main business for today. And the starting point is the following-- You know something about the world. And based on what you know when you set up a probability model and you write down probabilities for the different outcomes. Then something happens, and somebody tells you a little more about the world, gives you some new information. This new information, in general, should change your beliefs about what happened or what may happen. So whenever we're given new information, some partial information about the outcome of the experiment, we should revise our beliefs. And conditional probabilities are just the probabilities that apply after the revision of our beliefs, when we're given some information.

So let's make this into a numerical example. So inside the sample space, this part of the sample space, let's say has probability  $3/6$ , this part has  $2/6$ , and that part has  $1/6$ . I guess that means that out here we have zero probability. So these were our initial beliefs about the outcome of the experiment. Suppose now that someone comes and tells you that event B occurred. So they don't tell you the full outcome with the experiment. But they just tell you that the outcome is known to lie inside this set B.

Well then, you should certainly change your beliefs in some way. And your new beliefs about what is likely to occur and what is not is going to be denoted by this notation. This is the conditional probability that the event A is going to occur, the probability that the outcome is going to fall inside the set A given that we are told and we're sure that the event lies inside the event B. Now once you're told that the outcome lies inside the event B, then our old sample space in some ways is irrelevant. We have then your sample space, which is just the set B. We are certain that the outcome is going to be inside B.

For example, what is this conditional probability? It should be one. Given that I told you that B occurred, you're certain that B occurred, so this has unit probability. So here we see an instance of revision of our beliefs. Initially, event B had the probability of  $(2+1)/6$  -- that's  $1/2$ . Initially, we thought B had probability  $1/2$ . Once we're told that B occurred, the new probability of B is equal to one. OK.

How do we revise the probability that A occurs? So we are going to have the outcome of the experiment. We know that it's inside B. So we will either get something here, and A does not occur. Or something inside here, and A does occur. What's the likelihood that, given that we're inside B, the outcome is inside here? Here's how we're going to think about. This part of this set B, in which A also occurs, in our initial model was twice as likely as that part of B. So outcomes inside here collectively were twice as likely as outcomes out there.

So we're going to keep the same proportions and say, that given that we are inside the set B, we still want outcomes inside here to be twice as likely outcomes there. So the proportion of the probabilities should be two versus one. And these probabilities should add up to one because together they make the conditional probability of B. So the conditional probabilities should be  $2/3$  probability of being here and  $1/3$  probability of being there. That's how we revise our probabilities. That's a reasonable, intuitively reasonable, way of doing this revision. Let's translate what we did into a definition.

The definition says the following, that the conditional probability of A given that B occurred is calculated as follows. We look at the total probability of B. And out of that probability that was inside here, what fraction of that probability is assigned to points for which the event A also occurs? Does it give us the same numbers as we got with this heuristic argument? Well in this example, probability of A intersection B is  $2/6$ , divided by total probability of B, which is  $3/6$ , and so it's  $2/3$ , which agrees with this answer that's we got before. So the former indeed matches what we were trying to do.

**One** little technical detail. If the event B has zero probability, and then here we have a ratio that doesn't make sense. So in this case, we say that conditional probabilities are not defined.

Now you can take this definition and unravel it and write it in this form. The probability of A intersection B is the probability of B times the conditional probability. So this is just consequence of the definition but it has a nice interpretation. Think of probabilities as frequencies. If I do the experiment over and over, what fraction of the time is it going to be the case that both A and B occur? Well, there's going to be a certain fraction of the time at which B occurs. And out of those times when B occurs, there's going to be a further fraction of the experiments in which A also occurs.

So interpret the conditional probability as follows. You only look at those experiments at which B happens to occur. And look at what fraction of those experiments where B already occurred, event A also occurs. And there's a symmetrical version of this equality. There's symmetry between the events B and A. So you also have this relation that goes the other way.

OK, so what do we use these conditional probabilities for? First, one comment. Conditional probabilities are just like ordinary probabilities. They're the new probabilities that apply in a new universe where event B is known to have occurred. So we had an original probability model. We are told that B occurs. We revise our model. Our new model should still be legitimate probability model. So it should satisfy all sorts of properties that ordinary probabilities do satisfy.

So for example, if A and B are disjoint events, then we know that the probability of A union B is equal to the probability of A plus probability of B. And now if I tell you that a certain event C occurred, we're placed in a new universe where event C occurred. We have new probabilities for that universe. These are the conditional probabilities. And conditional probabilities also satisfy this kind of property. So this is just our usual additivity axiom but the applied in a new

model, in which we were told that event C occurred. So conditional probabilities do not taste or smell any different than ordinary probabilities do. Conditional probabilities, given a specific event B, just form a probability law on our sample space. It's a different probability law but it's still a probability law **that has all of the desired properties.**

OK, so where do conditional probabilities come up? They do come up in quizzes and they do come up in silly problems. So let's start with this. We have this example from last time. Two rolls of a die, all possible pairs of rolls are equally likely, so every element in this square has probability of  $1/16$ . So all elements are equally likely. That's our original model. Then somebody comes and tells us that the minimum of the two rolls is equal to zero. What's that event? The minimum equal to zero can happen in many ways, if we get two zeros or if we get a zero and-- sorry, if we get two two's, or get a two and something larger. And so this is our new event B. The red event is the event B.

And now we want to calculate probabilities inside this new universe. For example, you may be interested in the question, questions about the maximum of the two rolls. In the new universe, what's the probability that the maximum is equal to one? The maximum being equal to one is this black event. And given that we're told that B occurred, this black event cannot happen. So this probability is equal to zero. How about the maximum being equal to two, given that event B? OK, we can use the definition here. It's going to be the probability that the maximum is equal to two and B occurs divided by the probability of B. The probability that the maximum is equal to two.

OK, what's the event that the maximum is equal to two? Let's draw it. This is going to be the blue event. The maximum is equal to two if we get any of those blue points. So the intersection of the two events is the intersection of the red event and the blue event. There's only one point in their intersection. So the probability of that intersection happening is  $1/16$ . That's the numerator. How about the denominator? The event B consists of five elements, each one of which had probability of  $1/16$ . So that's  $5/16$ . And so the answer is  $1/5$ .

Could we have gotten this answer in a faster way? Yes. Here's how it goes. We're trying to find the conditional probability that we get this point, given that B occurred. B consists of five elements. All of those five elements were equally likely when we started, so they remain equally likely afterwards. Because when we define conditional probabilities, we keep the same proportions inside the set. So the five red elements were equally likely. They remain equally likely in the conditional world. So conditional event B having happened, each one of these five

elements has the same probability. So the probability that we actually get this point is going to be  $1/5$ . And so that's the shortcut.

More generally, whenever you have a uniform distribution on your initial sample space, when you condition on an event, your new distribution is still going to be uniform, but on the smaller events of that we considered. So we started with a uniform distribution on the big square and we ended up with a uniform distribution just on the red point.

Now besides silly problems, however, conditional probabilities show up in real and interesting situations. And this example is going to give you some idea of how that happens. OK. Actually, in this example, instead of starting with a probability model in terms of regular probabilities, I'm actually going to define the model in terms of conditional probabilities. And we'll see how this is done. So here's the story. There may be an airplane flying up in the sky, in a particular sector of the sky that you're watching. Sometimes there is one sometimes there isn't. And from experience you know that when you look up, there's five percent probability that the plane is flying above there and 95% probability that there's no plane up there.

So event A is the event that the plane is flying out there. Now you bought this wonderful radar that's looks up. And you're told in the manufacturer's specs that, if there is a plane out there, your radar is going to register something, a blip on the screen with probability 99%. And it will not register anything with probability one percent. So this particular part of the picture is a self-contained probability model of what your radar does in a world where a plane is out there. So I'm telling you that the plane is out there.

So we're now dealing with conditional probabilities because I gave you some particular information. Given this information that the plane is out there, that's how your radar is going to behave with probability 99% is going to detect it, with probability one percent is going to miss it. So this piece of the picture is a self-contained probability model. The probabilities add up to one. But it's a piece of a larger model.

Similarly, there's the other possibility. Maybe a plane is not up there and the manufacturer specs tell you something about false alarms. A false alarm is the situation where the plane is not there, but for some reason your radar picked up some noise or whatever and shows a blip on the screen. And suppose that this happens with probability ten percent. Whereas with probability 90% your radar gives the correct answer.

So this is sort of a model of what's going to happen with respect to both the plane -- we're



given probabilities about this -- and we're given probabilities about how the radar behaves. So here I have indirectly specified the probability law in our model by starting with conditional probabilities as opposed to starting with ordinary probabilities. Can we derive ordinary probabilities starting from the conditional number ones? Yeah, we certainly can.

Let's look at this event, A intersection B, which is the event up here, that there is a plane and our radar picks it up. How can we calculate this probability? Well we use the definition of conditional probabilities and this is the probability of A times the conditional probability of B given A. So it's 0.05 times 0.99. And the answer, in case you care-- It's 0.0495. OK. So we can calculate the probabilities of final outcomes, which are the leaves of the tree, by using the probabilities that we have along the branches of the tree. So essentially, what we ended up doing was to multiply the probability of this branch times the probability of that branch.

Now, how about the answer to this question. What is the probability that our radar is going to register something? OK, this is an event that can happen in multiple ways. It's the event that consists of this outcome. There is a plane and the radar registers something together with this outcome, there is no plane but the radar still registers something.

So to find the probability of this event, we need the individual probabilities of the two outcomes. For the first outcome, we already calculated it. For the second outcome, the probability that this happens is going to be this probability 95% times 0.10, which is the conditional probability for taking this branch, given that there was no plane out there. So we just add the numbers. 0.05 times 0.99 plus 0.95 times 0.1 and the final answer is 0.1445. OK.

And now here's the interesting question. Given that your radar recorded something, how likely is it that there is an airplane up there? Your radar registering something -- that can be caused by two things. Either there's a plane there, and your radar did its job. Or there was nothing, but your radar fired a false alarm. What's the probability that this is the case as opposed to that being the case? OK. The intuitive shortcut would be that it should be the probability-- you look at their relative odds of these two elements and you use them to find out how much more likely it is to be there as opposed to being there.

But instead of doing this, let's just write down the definition and just use it. It's the probability of A and B happening, divided by the probability of B. This is just our definition of conditional probabilities. Now we have already found the numerator. We have already calculated the denominator. So we take the ratio of these two numbers and we find the final answer -- which

is 0.34. OK.

There's this slightly curious thing that's happened in this example. Doesn't this number feel a little too low? My radar -- So this is a conditional probability, given that my radar said there is something out there, that there is indeed something there. So it's sort of the probability that our radar gave the correct answer. Now, the specs of our radar we're pretty good. In this situation, it gives you the correct answer 99% of the time. In this situation, it gives you the correct answer 90% of the time. So you would think that your radar there is really reliable.

But yet here the radar recorded something, but the chance that the answer that you get out of this is the right one, given that it recorded something, the chance that there is an airplane out there is only 30%. So you cannot really rely on the measurements from your radar, even though the specs of the radar were really good. What's the reason for this? Well, the reason is that false alarms are pretty common.

Most of the time there's nothing. And there's a ten percent probability of false alarms. So there's roughly a ten percent probability that in any given experiment, you have a false alarm. And there is about the five percent probability that something out there and your radar gets it. So when your radar records something, it's actually more likely to be a false alarm rather than being an actual airplane. This has probability ten percent roughly. This has probability roughly five percent

So conditional probabilities are sometimes counter-intuitive in terms of the answers that they get. And you can make similar stories about doctors interpreting the results of tests. So you tested positive for a certain disease. Does it mean that you have the disease necessarily? Well if that disease has been eradicated from the face of the earth, testing positive doesn't mean that you have the disease, even if the test was designed to be a pretty good one. So unfortunately, doctors do get it wrong also sometimes. And the reasoning that comes in such situations is pretty subtle.

Now for the rest of the lecture, what we're going to do is to take this example where we did three things and abstract them. These three trivial calculations that's we just did are three very important, very basic tools that you use to solve more general probability problems. So what's the first one? We find the probability of a composite event, two things happening, by multiplying probabilities and conditional probabilities. More general version of this, look at any situation, maybe involving lots and lots of events.

So here's a story that event A may happen or may not happen. Given that A occurred, it's possible that B happens or that B does not happen. Given that B also happens, it's possible that the event C also happens or that event C does not happen. And somebody specifies for you a model by giving you all these conditional probabilities along the way. Notice what we move along the branches as the tree progresses. Any point in the tree corresponds to certain events having happened.

And then, given that this has happened, we specify conditional probabilities. Given that this has happened, how likely is it for that C also occurs? Given a model of this kind, how do we find the probability or for this event? The answer is extremely simple. All that you do is move along with the tree and multiply conditional probabilities along the way. So in terms of frequencies, how often do all three things happen, A, B, and C? You first see how often does A occur. Out of the times that A occurs, how often does B occur? And out of the times where both A and B have occurred, how often does C occur? And you can just multiply those three frequencies with each other.

What is the formal proof of this? Well, the only thing we have in our hands is the definition of conditional probabilities. So let's just use this. And-- OK. Now, the definition of conditional probabilities tells us that the probability of two things is the probability of one of them times a conditional probability. Unfortunately, here we have the probability of three things. What can I do? I can put a parenthesis in here and think of this as the probability of this and that and apply our definition of conditional probabilities here. The probability of two things happening is the probability that the first happens times the conditional probability that the second happens, given A and B, given that the first one happened.

So this is just the definition of the conditional probability of an event, given another event. That other event is a composite one, but that's not an issue. It's just an event. And then we use the definition of conditional probabilities once more to break this apart and make it  $P(A)$ ,  $P(B \text{ given } A)$  and then finally, the last term. OK.

So this proves the formula that I have up there on the slides. And if you wish to calculate any other probability in this diagram. For example, if you want to calculate this probability, you would still multiply the conditional probabilities along the different branches of the tree. In particular, here in this branch, you would have the conditional probability of C complement, given A intersection B complement, and so on. So you write down probabilities along all those tree branches and just multiply them as you go.

So this was the first skill that we are covering. What was the second one? What we did was to calculate the total probability of a certain event  $B$  that consisted of-- was made up from different possibilities, which corresponded to different scenarios. So we wanted to calculate the probability of this event  $B$  that consisted of those two elements.

Let's generalize. So we have our big model. And this sample space is partitioned in a number of sets. In our radar example, we had a partition in two sets. Either a plane is there, or a plane is not there. Since we're trying to generalize, now I'm going to give you a picture for the case of three possibilities or three possible scenarios. So whatever happens in the world, there are three possible scenarios,  $A_1$ ,  $A_2$ ,  $A_3$ . So think of these as there's nothing in the air, there's an airplane in the air, or there's a flock of geese flying in the air. So there's three possible scenarios.

And then there's a certain event  $B$  of interest, such as a radar records something or doesn't record something. We specify this model by giving probabilities for the  $A_i$ 's-- That's the probability of the different scenarios. And somebody also gives us the probabilities that this event  $B$  is going to occur, given that the  $A_i$ -th scenario has occurred. Think of the  $A_i$ 's as scenarios.

And we want to calculate the overall probability of the event  $B$ . What's happening in this example? Perhaps, instead of this picture, it's easier to visualize if I go back to the picture I was using before. We have three possible scenarios,  $A_1$ ,  $A_2$ ,  $A_3$ . And under each scenario,  $B$  may happen or  $B$  may not happen. And so on. So here we have  $A_2$  intersection  $B$ . And here we have  $A_3$  intersection  $B$ . In the previous slide, we found how to calculate the probability of any event of this kind, which is done by multiplying probabilities here and conditional probabilities there.

Now we are asked to calculate the total probability of the event  $B$ . The event  $B$  can happen in three possible ways. It can happen here. It can happen there. And it can happen here. So this is our event  $B$ . It consists of three elements. To calculate the total probability of our event  $B$ , all we need to do is to add these three probabilities. So  $B$  is an event that consists of these three elements. There are three ways that  $B$  can happen. Either  $B$  happens together with  $A_1$ , or  $B$  happens together with  $A_2$ , or  $B$  happens together with  $A_3$ .

So we need to add the probabilities of these three contingencies. For each one of those

contingencies, we can calculate its probability by using the multiplication rule. So the probability of A1 and B happening is this-- It's the probability of A1 and then B happening given that A1 happens. The probability of this contingency is found by taking the probability that A2 happens times the conditional probability of A2, given that B happened. And similarly for the third one. So this is the general rule that we have here. The rule is written for the case of three scenarios. But obviously, it has a generalization for the case of four or five or more scenarios. It gives you a way of breaking up the calculation of an event that can happen in multiple ways by considering individual probabilities for the different ways that the event can happen.

OK. So-- Yes?

**AUDIENCE:** Does this have to change for infinite sample space?

**JOHN TSISIKLIS:** No. This is true whether your sample space is infinite or finite. What I'm using in this argument that we have a partition into just three scenarios, three events. So it's a partition to a finite number of events. It's also true if it's a partition into an infinite sequence of events. But that's, I think, one of the theoretical problems at the end of the chapter. You probably may not need it for now.

OK, going back to the story here. There are three possible scenarios about what could happen in the world that are captured here. Event, under each scenario, event B may or may not happen. And so these probabilities tell us the likelihoods of the different scenarios. These conditional probabilities tell us how likely is it for B to happen under one scenario, or the other scenario, or the other scenario.

The overall probability of B is found by taking some combination of the probabilities of B in the different possible worlds, in the different possible scenarios. Under some scenario, B may be very likely. Under another scenario, it may be very unlikely. We take all of these into account and weigh them according to the likelihood of the scenarios. Now notice that since A1, A2, and three form a partition, these three probabilities have what property? Add to what? They add to one. So it's the probability of this branch, plus this branch, plus this branch. So what we have here is a weighted average of the probabilities of the B's into the different worlds, or in the different scenarios.

Special case. Suppose the three scenarios are equally likely. So P of A1 equals 1/3, equals to P of A2, P of A3. what are we saying here? In that case of equally likely scenarios, the probability of B is the average of the probabilities of B in the three different words, or in the

three different scenarios. OK.

So to finally, the last step. If we go back again two slides, the last thing that we did was to calculate a conditional probability of this kind, probability of A given B, which is a probability associated essentially with an inference problem. Given that our radar recorded something, how likely is it that the plane was up there? So we're trying to infer whether a plane was up there or not, based on the information that we've got.

So let's generalize once more. And we're just going to rewrite what we did in that example, but in terms of general symbols instead of the specific numbers. So once more, the model that we have involves probabilities of the different scenarios. These we call them prior probabilities. They're our initial beliefs about how likely each scenario is to occur. We also have a model of our measuring device that tells us under that scenario how likely is it that our radar will register something or not. So we're given again these conditional probabilities. We're given the conditional probabilities for these branches.

Then we are told that event B occurred. And on the basis of this new information, we want to form some new beliefs about the relative likelihood of the different scenarios. Going back again to our radar example, an airplane was present with probability 5%. Given that the radar recorded something, we're going to change our beliefs. Now, a plane is present with probability 34%. The radar, since we saw something, we are going to revise our beliefs as to whether the plane is out there or is not there.

And so what we need to do is to calculate the conditional probabilities of the different scenarios, given the information that we got. So initially, we have these probabilities for the different scenarios. Once we get the information, we update them and we calculate our revised probabilities or conditional probabilities given the observation that we made. OK. So what do we do? We just use the definition of conditional probabilities twice. By definition the conditional probability is the probability of two things happening divided by the probability of the conditioning event.

Now, I'm using the definition of conditional probabilities once more, or rather I use the multiplication rule. The probability of two things happening is the probability of the first and the second. So these are things that are given to us. They're the probabilities of the different scenarios. And it's the model of our measuring device, which we assume to be available. And how about the denominator? This is total probability of the event B. But we just found that's it's

easy to calculate using the formula in the previous slide. To find the overall probability of event B occurring, we look at the probabilities of B occurring under the different scenarios and weigh them according to the probabilities of all the scenarios.

So in the end, we have a formula for the conditional probability, A's given B, based on the data of the problem, which were probabilities of the different scenarios and conditional probabilities of B, given the A's. So what this calculation does is, basically, it reverses the order of conditioning. We are given conditional probabilities of these kind, where it's B given A and we produce new conditional probabilities, where things go the other way.

So schematically, what's happening here is that we have model of cause and effect and-- So a scenario occurs and that may cause B to happen or may not cause it to happen. So this is a cause/effect model. And it's modeled using probabilities, such as probability of B given  $A_i$ . And what we want to do is inference where we are told that B occurs, and we want to infer whether  $A_i$  also occurred or not. And the appropriate probabilities for that are the conditional probabilities that A occurred, given that B occurred.

So we're starting with a causal model of our situation. It models from a given cause how likely is a certain effect to be observed. And then we do inference, which answers the question, given that the effect was observed, how likely is it that the world was in this particular situation or state or scenario.

So the name of the Bayes rule comes from Thomas Bayes, a British theologian back in the 1700s. It actually-- This calculation addresses a basic problem, a basic philosophical problem, how one can learn from experience or from experimental data and some systematic way. So the British at that time were preoccupied with this type of question. Is there a basic theory that about how we can incorporate new knowledge to previous knowledge. And this calculation made an argument that, yes, it is possible to do that in a systematic way. So the philosophical underpinnings of this have a very long history and a lot of discussion around them. But for our purposes, it's just an extremely useful tool. And it's the foundation of almost everything that gets done when you try to do inference based on partial observations. Very well. Till next time.