

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).

**PROFESSOR:**

Let's get started. Drop date is coming up. So that you didn't do very well on the midterm, and you haven't been doing your homework, and you haven't been going to recitation, I would strongly consider drop date. And you can see your TAs to get some idea of what your grades are, and what you're likely to do if you keep on giving the same performance you've had so far.

Now, today, we're going to switch gears again and cover the last topic for the course. For the remainder of the course, we're going to talk about probability. It's one of the most important subjects in all of math or computer science. In fact, I was at a frat last night for dinner, and they asked me what my favorite algorithm was.

And I couldn't really think about my favorite algorithm was, but in terms of one of the most important subjects for you to know about, after induction, it's probably probability. It's really important across math and computer science. Most of the upper level courses you take in computer science are going to use probability. They're going to expect you to know the basics.

For example, when you're designing algorithms in 006 or 6046, you're going to be designing in some cases probabilistic algorithms-- algorithms that flip coins or generate random numbers to get the result. And by using randomness, they run faster. If you're doing software engineering later, you're going to use probability when you analyze system performance or build hash tables. Probability is used in information theory, in coding theory, in cryptography.

In fact, our nation's codes, their security depends on probabilistic analysis. Because the bad guys are using probabilistic algorithms to break the crypto-system. And we'll give an example of that in another week or so.

Probability is obviously important if you're doing game theory or if you like to gamble, really critical in gambling. In fact, we're going to analyze a lot of different gambling games over the next few weeks. And for fun, we'll play some in class. In fact, we'll play one later today. And it's amazing how often the intuition is wrong. You think the answer goes one way, and then it's

actually the other.

Probability is important in the study of fault tolerance. For example, how do you figure out the probability there's going to be some critical failure that causes the space shuttle to crash? Or a critical failure that causes a nuclear plant to meltdown? That's done through probabilistic analysis. And we'll do some examples of that.

Probability comes up in everyday life. We just had an election. And running up to the election, you get all these polls that say that's so-and-so is ahead of so-and-so by three points with a margin of error of five points. So it's a statistical dead heat, is what they say. Well, what the heck does that mean? In fact, usually they've got it screwed up. And we're going to spend the recitation where you're going to do a margin of error analysis for a poll, and see how many people do you need to poll to get a pretty accurate result?

Comes up in the medical studies. Does cholesterol cause heart disease? Does smoking cause lung cancer? How effective is a certain drug? Say you get tested for a disease and it comes back positive. Well, there's false positives and false negatives. What are the chances you really have the disease? So the list goes on and on. In fact, we're going to see all these examples. We've got eight lectures and eight recitations. And there's just no end of interesting examples where probability comes into play.

The bad news is, probability is probably the most misunderstood subject in all of mathematics and the sciences. In fact, Mark Twain once said, there's three kinds of lies-- lies, damned lies, and statistics, meaning probability. Because it's so easy to lie with statistics. And we're going to see lots and lots of examples of that.

In fact, newspaper articles, you know, have all sorts of things that are just horribly fallacious based on statistics. And we'll see examples. In fact, we're going to even debunk a very famous article-- published research paper-- by some famous computer scientists at Berkeley. Where they got to a conclusion using a very standard method you see all the time-- completely fallacious. And we'll see that.

So you can think of that as, I'm going to teach you how to lie with statistics. Even better, how not to be taken in by incorrect arguments using probability theory. Now, at MIT, it's probably the subject that most often trips up students in their PhD qualifying exams. You know, there's no better way to strike terror in the heart of a student than to be sitting there and ask them a

question while they're at the board, what's the probability that foo happens? It's like panic sets in when they've got to do a probabilistic analysis.

And now, MIT PhD students are good. You know, they can crank out the most horrendous calculation, no problem, come up with really clever ideas to solve problems. But you get into probability, it gets into trouble. In fact, even the other faculty sitting there start squirming around when this question's being posed. You know, and probability questions are-- and we'll see some examples today-- are always the best at faculty parties, just can ruin people's whole night. You know, what's the probability this happens? And it's some weird answer.

Now, the strange thing about this-- well, first strange thing is that our human intuition is terrible. And you will see that. Your intuition will be wrong, just the way our brains are created genetically. But the good news is that probability can be very simple. If you force yourself to just go to the basic principles-- which we're going to cover-- and go through the basic template of how to solve a problem-- which we're going to cover-- it is easy, and you can't go wrong.

The key is forcing yourself to just actually write it down, follow the basic principles, and get to the answer. As opposed to saying, oh, the answer's that and here's why. Because it's amazing how often you'll be wrong if you just try to rely on your intuition.

Now, we're going to start today our study of probability with a famous, simple game. And the game is known as the Monty Hall problem. Now, do you all know who Monty Hall was? You're probably too young. Yeah, OK, not many people know who Monte Hall was. But back in the 1970s, there was a very popular daytime TV show-- a game show-- called *Let's Make a Deal*. And it was a great show. It was the precursor of *Wheel of Fortune*-- if you're old enough to have seen that-- and then *Deal or No Deal* after that. There's always some game show that's really popular.

Anyway, back in the '70s, it was *Let's Make a Deal*, and Monty Hall was the host. How many people have ever heard of *Let's Make a Deal*? Raise your hand if you've ever heard of that. OK, so it still has some fame. Good. All right. Now, Monty Hall, of course, was the precursor of Pat Sajak of *Wheel of Fortune* and Howie Mandel, you know, from *Deal or No Deal*. And his assistant was Carol Merrill. She was also famous as the beautiful assistant.

Now, in *Let's Make a Deal*, they TV audience would come dressed up in wild costumes. And Monty would go into the audience and pick out a contestant from the audience based on how wild their costume was-- you know, somebody that would look interesting for TV. And what he

would do is he'd buy a piece of the costume-- maybe their hat or shirt or something-- and give them \$100 for a piece of their costume. And then he would say, hey look, you know, if you give me back the \$100 I just gave you, I'm going to let you have your choice of what's behind door number one, door number two, or door number three. Or sometimes it would be box number one, box number two, and box number three.

Now, because people have seen the game show before, they knew that behind one of the doors was a brand new car or underneath one of the boxes was a diamond ring. And behind the other doors were goats or donkeys or something totally useless. So one of the three doors has a big prize, the other two got nothing. But because anybody would pay \$100 for a one-in-three chance at the car, they played. And they're on national TV. So they give the \$100 back to Monty, and they pick a door, say door number one.

Monty wouldn't open door number one right away. What he would do is say, OK, you're sure you like that? All right. Carol, can you please reveal what's behind door number three? So Carol opens up door number three, and there's one of the donkeys sitting there or a goat sitting there. Well, now there's only two doors left. One of them has the car. So he'd ask the contestant, well, you've seen the goat. There's a car behind one of the two doors. Do you want to stick with door number one? Or do you want to switch and get door number two?

And then the audience is screaming, you know, switch, stick, you know, door number one, door number two. You know, the guy's wife's there screaming at him and everything. You know, he's got to pick, so he picks one. And you know, if he picks the car, well, his wife's thrilled. They live happily ever. He's a hero on national TV. And if he picks the goat, you know, he's looking at a divorce and his career's ruined. So it's pretty important that he gets the right strategy here.

Now, the question we're going to look at today is whether or not the contestant should switch, playing this game. Now, before we do the analysis, we're going to actually play the game in class. I've got three boxes here. The game we're going to play is the same idea. The prizes are a little different. Instead of a diamond ring or a brand new car, we've got \$100 gift certificates. You know, your tuition only goes so far here.

Instead of donkeys or goats, we've got something much better, the ever-popular nerd pride pocket protector. And just to be clear, this is not the prize, the gift certificate here-- this is much, much better.

[AUDIENCE LAUGHTER]

Now, in terms of Monty Hall, we don't have him, but you have me. But I am very happy to report that we're bringing Carol Merrill out of retirement. She's coming and making a special appearance today, I hope, at 6042. Do we have Carol?

**GUEST SPEAKER:** Hi, Tom.

**PROFESSOR:** Oh, yes, we have Carol. That's scary.

[AUDIENCE LAUGHTER]

Oh, Jesus. OK.

[APPLAUSE]

As you can see, the years have not been kind to Carol. And your tuition only does goes so far. Now, we need three volunteers. We're going to play the game three times. All right, got a couple there. Why don't you come down? We've got three volunteers here. All right, we'll get the boxes set up. And Carol, why don't you set those up and load them up with-- these are the gift certificates. That's the pocket protector there. All right, now, don't look. You've got to get back over there so you can't see--

[AUDIENCE LAUGHTER]

And no, no, no, now, you've got the film going, you know, here we go. All right. So Carol is going to put the prize in one of the boxes and the pocket protectors in the other. Now, I want you all to be thinking, you know, what's the right strategy here? Is it better to switch? Does it matter? What are the chances of winning? One in three? One in two?

All right, so who's our first volunteer here? All right, come on over. We're going to write your name on the board because we're going to keep track of what happens here. What's your name?

**TERRANCE:** Terrance.

**PROFESSOR:** Terrance. All right. Now, first thing you've got to do, Terrance, is pick a box as your basic choice to start.

**TERRANCE:** Two.

**PROFESSOR:** Number two, OK. All right, you picked number two. Hm, interesting choice. Now, Carol, can you show him one of the other boxes? All right, which one was that you showed there? You revealed number three. And what was under number three?

**AUDIENCE:** Pocket protector.

**PROFESSOR:** Pocket protector. Oh, God, you may never live this down, Carol. OK. Now, Terrance, as you know, one of these has the gift certificate. One of them has another pocket protector. And now, you've got to decide, are you going to stick with number two, or are you going to switch? In fact, class, you know, Terrance may need some help here, figuring this out, so what do you recommend?

[AUDIENCE CHATTER]

**TERRANCE:** I'll switch.

**PROFESSOR:** He is going to switch. All right, and Carol, can you reveal his prize? What did you get? Ah, he gets the prize. So he wins. Very nice, OK. So you get \$10 at Tosci's. Well done.

Who's next? What's your name?

**CHINUA:** Chinua.

**PROFESSOR:** Chinua. How do I spell that?

**CHINUA:** Chin-U-A.

**PROFESSOR:** All right, easy. OK. Now, Chinua, which-- have we got new prizes loaded in there? All right. Now, it may not be the same as last time. OK. What box do you like?

**CHINUA:** Number three.

**PROFESSOR:** Number three. Which one is number three? Is that the one close to me? Yeah, that is. OK. Number three. All right, Carol, can you show him one of the other boxes, please?

Carol shows you number two. All right, Chinua, what do you think? Do you want to stick or switch? Do you want to stick with three or switch to one? One had it last time, right?

[AUDIENCE CHATTER]

**CHINUA:** Switch.

**PROFESSOR:** He's going to switch. OK, so that means you are picking number one. Carol, can you show him what he's won? Another gift certificate. All right, now, Carol it is OK to use a different box for the gift certificate.

[LAUGHTER]

We can move them around. All right, who's our last contestant? What's your name?

**TY:** Ty.

**PROFESSOR:** T-Y, OK. That's good. Oh, yeah, yeah, turn around. No peeking here. OK. Ty, what's your first choice?

**TY:** One.

**PROFESSOR:** Box number one. Yeah, that's been the one with the gift certificate, for sure. Carol, can you show him another box? Way too many cameras in this room. This is not good for Carol.

[LAUGHTER]

OK, Carol has revealed box number two.

[AUDIENCE CHATTER]

Switch has worked twice in a row.

**AUDIENCE:** Now you gotta stay.

**AUDIENCE:** Take them both.

**PROFESSOR:** Just one.

**AUDIENCE:** Well, if you switch, there's two-thirds chance of winning, and they already won so--

[LAUGHTER]

**PROFESSOR:** Oh, getting some theory here now. What do you think, Ty? Switch. All right. What happened? What do you know? All right, now, gentlemen, I'd like you to explain why-- each in turn, why did you switch?

**TERRANCE:** I switched-- it seemed like the majority vote.

**PROFESSOR:** The majority knew what they were talking about, OK. You have faith in these guys, huh? All right. How about you, Chinua, why did you switch?

**CHINUA:** Um, I was told that it's more likely to be if you switch.

**PROFESSOR:** And you believe what you hear, huh?

[LAUGHTER]

**TY:** I don't know who said it, but it sounded good.

**PROFESSOR:** Sounded good Yeah, that two-thirds thing. Well, as you can see, you're guaranteed to win if you switch, right?

[LAUGHTER]

All right, Thank you very much. Enjoy the ice cream. OK. So how many people think it helps to switch? Raise your hand if you think it helps to switch. How many people think it does not help to switch? Raise your hand. Anybody? Oh, everybody likes the switch thing here. Does this data help you decide? No. Anything could have happened with three trials. In fact, we'll talk about statistical inference and statistical sampling later.

If you switch, what do you think the odds of winning are?

**AUDIENCE:** 100%

**PROFESSOR:** 100%? OK, who likes 100%? Raise your hand. Well, we've got a couple. How many people think it's 50-50 if you switch? Yeah, a lot. How many people think it's worse than 50-50 if you switch? Not many. How many people think it's better than 50-50? Ah, that's I think most of you there, pretty close. All right, now, the answers to these questions are very easy to work out-- and we're going to do it. We're going to figure out the probability exactly-- if you use mathematics. But it's really tricky if you try to use intuition or hand waving arguments.

Now, for example, this game-- Carol wants to leave-- yeah, Carol, let's give Carol a hand here. She did a wonderful job. Well done.

[APPLAUSE]

Yikes. I'm going to get in trouble here, too, now. You know? OK. For example, this game, and this particular version of it, was the subject of a series of articles in *Parade Magazine*. This is an old version of *Parade Magazine*. And what I'm going to do is hand out the series of articles. And the TAs are going to help me hand these out. There was three of them with a lot of write-ins here. I'm going to give you a bundle here. There you go. There you go. I'll give the rest to Oscar. Great.

Now, Marilyn has a column every week in *Parade*. And people write in questions, usually about their love life and weird things. And she answers them. And she's well qualified to do this because she's listed in the Guinness Book of World Records as having the highest IQ of a human being ever, at 228. And so, you know she's in some sense qualified to answer any question that is written in.

Now, in September of 1990-- and we're passing out the articles now and it shows the first article-- this guy named Craig Whitaker writes in and asks Marilyn if you should switch in this game. So you can all read that. It says, suppose you're on a game show and you're given the choice of three doors. Behind one door is a car. Behind the others, goats. You pick a door, say number one, and the host, who knows what's behind the doors, opens another one, say door three, which has a goat. He says to you, do you want to pick door number two? Is it to your advantage to switch?

Now, Marilyn writes back saying, very plainly, yes, you should switch. The first door had a one-in-three chance of winning. The second door has a two-in-three chance. Here's a good way to visualize what's going on. Suppose there are a million doors. You pick door number one. Then the host, who knows what's behind the doors, reveals every other door but one and 777,777. Well, it would sort of look like 777,777 is the one with the car, and you'd switch pretty fast.

All right that's her argument. That is not exactly a proof, OK, that the answer is you should switch or that the probability is  $2/3$ . All right? Now, because she's giving another problem with a million doors, but you've got a problem with three doors. And as we'll see in recitation tomorrow, the number of doors actually matters in terms of some strategies here.

So people weren't convinced by this. And so, if you look to the next page, you see people writing in. And there's a lot a write-ins. The first one's by Robert Sachs, who's actually-- I

looked him up on Google, and he's on the faculty at George Mason University. I also noticed that in his history, he got his BA from Harvard. So that should tell you something about the quality of his response here. He says, since you seem to enjoy coming straight to the point, I'll do the same. You blew it. Let me explain. If one door is shown to be a loser, that information changes the probability of either remaining choice, neither of which has any reason to be more likely. And so, the answer is one half of having the car. Then he says, as a professional mathematician, I'm very concerned with the general public's lack of mathematical skills. Please help by confessing your error. And in the future, be more careful. Just a little gratuitous extra statement there.

Next we've got a PhD from Florida, Scott Smith, who says, you blew it, and you blew it big. Since you seem to have difficulty grasping the basic principles at work here, I'll explain. After the host reveals a goat, you now have a one-in-two chance of being correct. Whether you change your selection or not, the odds are the same. There is enough mathematical illiteracy in this country, and we don't need the world's highest IQ propagating more. Shame. Well, now, this guy is right about that one point about mathematical illiteracy in the country.

The next guy, Barry Pasternack, you know, he lists himself as California Faculty Association. He actually was president of the California Faculty Association. He says, your answer to the question is an error, but if it's any constellation, many of my academic colleagues have also been stumped by this problem.

OK. So next page, Marilyn writes back. And she says, you know, good heavens, with so much learned opposition, I'll bet this is going to keep math classes all over the country busy on Monday. And it did. It was a huge uproar at the time about this. So now she tries to explain it some more with shells and stuff under the shells and so forth, trying to argue it. And then she says, just try the experiment six times or something, and you'll see. You know, but, it's not really convincing. It's not a proof. She's sort of waving hands at it to give you some idea of why she thinks she's right. But when I look at the hand waving here versus the hand waving on those faculty, well, it's all hand waving. All right.

So you go to the next page, and now the letters are getting a little nastier. We got another University of Florida guy. He says, may I suggest you obtain and refer to a standard textbook on probability before you try to answer a question of this type again. Then Robert Smith, Georgia State, I am sure you will receive many letters on this topic from high school and college students. Perhaps you should keep a few addresses for help with future columns.

These are great.

Now this next guy, Ray Bobo, now, he's a real guy. I looked him up, too. It's not fake. That's his real name. And I've got to say, if my name was Bobo, I'd think twice before writing into *Parade Magazine*, you know, especially if I'm wrong. Anyway, here's Bobo. You are utterly incorrect about the game show question. And I hope this controversy will call some public attention to the serious national crisis in mathematical education. Well, that's fair. If you could admit your error, you will have contributed constructively towards a solution of a deplorable situation. Now, here's [INAUDIBLE]. How many irate mathematicians are needed to get you to change your mind? Well that's proof by intimidation, right?

[LAUGHTER]

You know, it's not, here's the reason why. It's just, so many irate mathematicians said it, they must, must be right.

Now the next guy. I am in shock that after being corrected by at least three mathematicians, you still did not see your mistake. As if three mathematicians, that's sort of the criteria for correctness. Oh, this guy is bad. Maybe women look at math problems differently than men. Ooh. Well, maybe it's a good thing in this case. I don't know. Oh, then we have Glen Calkins, you are the goat. Finally, this guy, US Army Research Institute, a little scary. you made a mistake, but look at the positive side. If all those PhD's were wrong, the country would be in some very serious trouble.

All right. Now Marilyn writes an even longer response because she's getting zillions of letters, 90% voting against her, that she's wrong. She begins now, suggests a nationwide experiment, so that, nationwide, you'd do the sample a million times and get the answer, which really wouldn't prove it either. Then we go to the last page. There's one last letter. And this is what finished the controversy in the press.

Comes from MIT. You are indeed correct. My colleagues at work had a ball with this problem, and I dare say that most of them, including me at first, thought you were wrong. So maybe that \$50,000 a year somebody's paying for education is worth something, because MIT came in and agreed with her.

OK. It turns out that Marilyn was correct in her statement that you should switch. And if you switch, you have a  $\frac{2}{3}$  chance of winning. So if you don't switch, you've got a  $\frac{1}{3}$  chance of

winning, provided that Monty is guaranteed to open a door with goat. You know and that's the assumption here. And the proof is simple. Although Marilyn's reasoning wasn't so convincing, I didn't think. And all those PhDs that wrote in with those stupid letters, they're probably intelligent people. After all, they got advanced degrees in mathematics, and most of them were faculty teaching mathematics. A little scary.

But they just weren't following the basic principles. They were following their intuition. They didn't go through the basic steps to figure out the probability. So that's what we're going to do. We're going to go through the basic steps now and see how to solve it. And we will use these steps to solve pretty much every problem in probability. OK?

So the key, the first step, is to look at the sample space of possible outcomes. Let me define sample space and outcome. The sample space for an experiment, or a probability game-- you can treat this Monty Hall game as an experiment-- is simply the set of all possible outcomes, things that could happen.

In fact, let me define an outcome, also known as an atomic event or a sample point. An outcome, also known as a sample point-- we'll use those terms interchangeably-- consists of all the information about the experiment, including-- and that's after it's been performed-- including the values of all random choices.

For example, say we want to know the probability of winning in Monty Hall if you switch. All right, let's figure that out. So in that case, we're going to define an outcome of the Monty Hall game or a sample point-- and I'm assuming in this case, we know you're going to switch and the contestant switches-- consists of three things. First, the box with the prize, where did Carroll put it? Second, the box chosen by the contestant. And third, the box that was revealed. Once you know those three things, you know everything about what happened. All right?

For example, let's look at a typical sample point. Say I took sample point, or outcome, 2,1,3. This is the outcome where the prize is the box 2. The player picks first box 1. And Carol reveals box 3. OK, now in this case, in this outcome, did the contestant win? Yes, because we're assuming-- for henceforth, for most of the rest of the day, we're assuming the switch case, because we're going to analyze that. So that's fixed, they switch. They come in knowing they're going to switch.

So the prize is in 2, all right? The prize is here. The player started here. Carol revealed 3. And when they switched, they got the prize. So in this outcome, the player won. OK. Now, does

every 3-tuple here correspond to a sample point, to our possible outcome? No. What's an example of a sample point with numbers 1, 2, and 3 that's not. Yeah?

Very good. Yeah, so for example, 1, 2, 1 is not a sample point, is not in the set, the sample space, because we can't reveal the box with the prize. Similarly, 2, 1, 1 is not a sample point, because we can't reveal the box that the player picked. Yeah?

**AUDIENCE :** I think the first two pieces of data actually define the third one, because once, or-- actually, no, if they're the same.

**PROFESSOR:** Right, so if the player picked the box with the prize, which never happened in our case. For example, 1, 1 2 and 1, 1, 3 are both possible because you could put the prize here, the player could pick it, now Carol has a choice of revealing 2 or 3. OK? So these are OK.

All right, so let's figure out a nice way of characterizing all outcomes in the sample space. Now, to do this, we're going to set up a tree. And we're going to use a thing called the tree method to construct the sample space, getting all the outcomes. All right. So the first thing we do is figure out where the prize is. And we'll represent that as sort of like a decision tree. There's three choices. It's in box 1, box 2, or box 3. Then the next thing is the box that the player picked. And we'll continue our decision tree here. Each of these now has three branches. And they're labeled 1, 2, or 3.

And then, the last step is which box is revealed. Now, in this case, if the prize is in one and the player chose one, there's two possibilities. We could reveal 2 or reveal 3. So we get 1, 2, 3 here and 1, 2, 3 here. What about here? Prize is in box 1, box 2 is revealed-- sorry, player picked box 2. Prize is in 1, player picks 2, how many choices for which box is revealed? One. And which box is going to be revealed? 3. All right, this is sample point 1, 2, 3. And here we've got 1, 3, 2 is the only choice. All right.

And let's do-- maybe I'll, we ought to fill the whole thing out here. 2, 1, 3-- I'm going to run out of space, but-- 2, 2, now this splits into 1 and 3. All right? So we have here 2, 2, 1, 2, 2, 3. Now let me do the last one, 2, 3, 1 here. And now I'll bring these up over here. I've got 3, 1, 2. Then I've got 3, 2, 1. And finally, I've got 3, 3 with two choices, 1 and 2. All right? That is now the construction of the entire sample space, all the possible outcomes. How many sample points are there here? How many possible outcomes?

**AUDIENCE :** 12

**PROFESSOR:** 12. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, OK. Let's see which ones correspond to victory for the player. Let's label each one as a win for the player or a loss for the player. What's this one? Yeah. Prize is in 1, the player chose 1 where the prize was, and then switched away from it. Loss. What about this one? Loss, same reason. What about this one? That's a win. The prize is here, the player went there and switched, so they wound up there. Win. And these are all wins for the same reason. Whoop, not this one. 2, 1, 1, the player got it right to start with, and then switched away. That's a loss. Loss. Win, win, win. And here the player got it right, they switched away, it's a lose.

All right, so we've labeled every outcome as to whether the player who's switching wins or loses. How many sample points lead to victory? 6. 1, 2, 3, 4, 5, 6 out of 12. Uh-oh. Was Marilyn wrong? Is it really Bobo was right? It's 50-50? Half the sample points lead to victory. Hm, what's missing here? Yeah?

**AUDIENCE :** The first two are identical, the box with the prize and the box chosen first. Each of those two sample points coming from that are half as likely as the single sample points.

**PROFESSOR:** That's right. What we're missing here is the likelihood of an outcome, the probability that this is what happens. We're missing one more thing. And this is really critical when you're making your sample spaces. You need to make sure you get the probabilities assigned to each event. They're not necessarily equally likely.

In fact, what we need to do is to construct the probability space. So let's define that. Your probability space consists of a sample space-- which we've already talked about. That's just all these outcomes-- and a probability function that maps-- and we call this function, usually,  $P$ . We'll have, sometimes it'll be  $f$  or something like that-- but  $P$  maps the sample space,  $S$ , to the real numbers such that first, for every outcome in the sample space, the probability is between 0 and 1 of that outcome. And 2, if you sum up the probabilities of all the outcomes, you get 1. So the sum over all outcomes of the probability of the outcome is 1. All right?

So the sample space is all the outcomes. The probability space adds in-- has this extra thing-- the probability of each outcome. And really, it could be anything, as long as the probability of every outcome is between 0 and 1 and they all add up to 1. That's it. All right. Now, there's a natural interpretation of what the probability function is. And that is the probability of a sample point, well, should be the probability of its outcome. And this is sort of where you get the to Marilyn saying, hey, try it a million times and you'll see  $2/3$  of the time, roughly, you get the

right, that you win.

So for every outcome, the probability of that outcome, that real number, should be the probability that when you run the experiment,  $w$  is what happens.  $w$  will be the outcome. In other words, if probability of  $w$  were a half, what that means is you should get that half the time, roughly speaking. We'll get into really details of that later. In other words, you toss a fair coin, you flip it. Well, you might say, a probability half of a heads. All right? And so, this function, then, the probability of heads is  $1/2$ .

Now, assigning the right probabilities to the outcomes is 95% of the work. In fact, if you screw that up, you're doomed because you can't get the right answer. So for example, if we just said, OK, each of these is probability  $1/12$ , we'd get Bobo's answer as the probability of winning. So that's not too good.

Now, to get the right probabilities for the Monty Hall problem, we have to make some assumptions. So let's write those down. So the first assumption we're going to make is that Carol put the prize in a random box. They're all equally likely. So the prize is in each box with probability  $1/3$ . The next assumption we're going to make is that the player didn't see which box has the prize. OK? In other words, and more specifically, no matter where the prize is, the player picks each box with probability of  $1/3$ . So the player's making, effectively, a random choice. All right, and there's one more assumption we're going to make. Anybody have any ideas about what's left to make as an assumption here? Yeah?

**AUDIENCE :** I guess, if Carol's the one who's opening the boxes, she knows where all the prizes are.

**PROFESSOR:** Yeah. OK. So what assumption should I make? Yeah, that she picks a box at random if she has a choice. All right, so we're going to make that assumption. So that no matter where the prize is, if Carol has a choice, she picks each box with probability  $1/2$ .

Now, why is this third one important? Why might that really matter? Can anybody think of a reason why that could be relevant? What if Carol had some strategy in mind that the contestant knew about? Can you think of why that might be useful.

**AUDIENCE :** If the contestant were able to produce information based off of that, then there's a higher chance a contestant could figure out which one to choose.

**PROFESSOR:** Yeah. In fact, several years ago, when I was lecturing, I said, in fact it could make a difference.

That there is a way that Carol can convey information to the player by of two she's choosing, which one to pick. Like, if she has a choice of 2 or 3, say she always picks bigger one, 3. Well, that might tell you information about, she didn't pick 2, so maybe it has to have, or she-- well, I don't know.

In fact, that's wrong, and it doesn't matter in the case of 2 boxes, this assumption. We're going to make it, because we're going to do the analysis with it. But for three boxes, you can show that it really doesn't matter for this assumption. But if you go to four boxes, it does. And so in recitation tomorrow, you're going to consider the four box version of this game. And if Carol has a strategy that the player knows about, you could improve the player's chances of winning. But not for three, it turns out.

In any case, when you're setting up a problem, you've got to set up your assumptions. Otherwise, you don't get anywhere. These are sort of like axioms, same kind of thing. It's fine to make assumptions. Just state what they are. Now, of course, you also want them to be reasonable. All right. So maybe we're going to go back now and figure out how to get the probabilities of the sample points, or the outcomes, for Monty Hall. Let's go do that. And maybe I'll draw them in a different color here.

What I'm going to do is label every edge with a probability associated with taking that step. So if the prize is placed in each box with a probability of  $1/3$ , that says this happens with probability  $1/3$ , that's  $1/3$ , and that's  $1/3$ . Same thing here. Assumption two says that no matter where the prize is, say it's in box 1, the player chooses each box with a probably of  $1/3$ . So given that the prize is in box 1, the player chooses 1 with  $1/3$ , 2 with  $1/3$ , 3 with  $1/3$ . Same if the prize is in box 2. So I label every edge here with the probability. All right.

Now I go to this level. And I'm, at this node, I'm in the state where prize is in box 1, player selected box 1, so Carol has a choice of 2 or 3 to reveal. And she's going to reveal each one with probability  $1/2$ . So I label  $1/2$  on each of these, because that's what we've assumed. What probability do I put on this edge here? Probability 1, because there's no choice. Carol has to do it. The same here. And I get  $1/2$ ,  $1/2$ , 1, 1, 1,  $1/2$ ,  $1/2$ . All right, so I've labeled every edge with its probability. Pretty simple, because it was all assumed.

Now I've got to get a probability for each outcome. And we do that with the following rule. The probability of a sample point is the product of the probabilities on the path leading to the outcome, the path in the tree leading to the outcome or the sample point. OK.

So what this says is just really, really simple. The probability of this sample point is  $1/3$  times  $1/3$  times  $1/2$ ,  $1/18$ . The probability of this sample point is  $1/3$  times  $1/3$  times  $1/2$ ,  $1/18$ . What's the probability of this sample point going to be here?  $1/9$ . I've got  $1/3$  times  $1/3$  times, well, this  $g$  was 1, is  $1/9$ . The next one here is  $1/3$  times  $1/3$  times 1,  $1/9$ . Same thing here.  $1/3$  times  $1/3$  times 1 is  $1/9$ . Then I get  $1/3$  times  $1/3$  times  $1/2$  for  $1/18$ . And I can keep on going. This would be  $1/9$ . This one has a 1, so it's a  $1/9$ . An  $1/18$  and  $1/18$ . OK.

Now, that's the rule. And you'll always do this. You can sort of see why it works. You'll understand after another week why, for sure, why it works. But intuitively, you know, to get to this outcome, these are the choices we had to make. First, the box had to be-- sorry, the prize had to be in box 1. There's a one-in-three chance of that happening. Now, in those one-out-of-three chances, the chance the player also picked box 1,  $1/3$  of the time of  $1/3$  of the time that happens. All right?  $1/3$  of the time, you're here. And of that,  $1/3$  of those times, you get to here. That's  $1/3$  times  $1/3$ , or  $1/9$ . And then,  $1/2$  of those times, Carol happen to reveal box 2. That's  $1/18$ .

So do you see why that might make sense, why you're multiplying these? Because at every step, you have a chance to diverge. And the probability you take the step that we're following here-- say, in this case, it was  $1/2$ -- that cuts down the chance you land here by that probability,  $1/2$ .

Any questions about that? All right. For now, you can take it on faith. We'll see why after we do a few more of these over the next week or so. Any questions there? Is that OK? OK. Oh, yeah, all right?

**AUDIENCE :** [INAUDIBLE]

**PROFESSOR:** On this one? Only one choice. So it happens with probability 1. Any other questions? OK. All right. So now we can figure out the probability of winning. Well, the probability of winning is we have  $1/9$ ,  $1/9$ , there's 3 of them, 4 of them, 5, 6. All right? So the probability of winning here is 6 times  $1/9$ .

Because you just sum up the probabilities of the winning sample points. There's six of winning. They're each  $1/9$ . And that's  $2/3$ . OK? So Marilyn was right. There's a  $2/3$  chance of winning if you switch. And now, this is a formal argument that gives you that answer. The nice thing is it works for pretty much every problem you ever think about.

Now, in general, if you want to know the probability of any event, you can follow this same method. Let me define what an event is. An event is simply a subset of outcomes, a subset of the sample space. For example, we could define the event of losing. Let's call  $E_L$ ,  $E_{\text{sub-L}}$ , to be the event that player loses, the player loses, with the switch strategy.

All right? Now, to compute the probability of an event, we just add the probabilities of the outcomes in that set. So let's get that down. All right, so formally, the probability that an event occurs is simply the sum, we'll call an event,  $e$ , let's say, event,  $e$ . It's a subset of the sample space. The probability that  $e$  occurs is the sum over the sample points in  $e$  of the probability of those sample points. So for example, the probability of the event,  $E_L$ , where the player loses, well, there's six sample points where the player loses, each with the probability  $1/18$ . So it's  $1/3$  is the probability the player loses on the switch strategy. And that's not surprising because you either win or you lose. The probabilities have to add to 1. So if won with probability  $2/3$ , you lose with probability  $1/3$ . OK? All right.

So Marilyn was right, much better to switch. We'll do a bunch more examples with Monty Hall tomorrow. Are there any other questions on Monty Hall before I change gears? Yeah?

**AUDIENCE:** So you showed that when you switch, you are more likely to win than to lose. But you didn't show that that's not also true when you don't switch.

**PROFESSOR:** Ah good question. What happens if you're playing the stay strategy. Right, maybe you win  $2/3$  of the time when you stay. Maybe. Maybe not. Right? Because if I win when I switch, what would have happened if I'd stayed? I'd have lost. Because if I win when I switch, it means the prize is over there. So if I stay, I'm going to lose. Which means that if I do the stay strategy, I'm going to lose  $2/3$  of the time. So you're right. I didn't cover that. But we can. It's not too hard in this case because the probability of winning on a switch equals the probability of losing on a stick.

So the probability that you win with switch equals the probability you lose with a stick strategy. All right? And they're both  $2/3$ , because they're the same and this is  $2/3$ . So in fact, what's, then, the chance you win with a stick strategy?  $1/3$ . And that's sort of you know, where Marilyn was coming from. Because she said, well, if you pick the box right the first time, which is what you need to get to win with a stick strategy, the chance you pick the right box is one-in-three. And showing, you know, a goat over here didn't matter. It's still one-in-three.

Now that's gets a-- that's intuition, but it can get you trouble, because say you picked this box

and Carol showed you both of these boxes, and they had pocket protectors. Then what do you know about the probability you won with a stick? It's 1. Ooh, so things did change there. All right? So you have to be a little careful about how you reason, OK? You've got to do this. Stay with that and you'll always be fine for your strategy. So really the key is to resist the temptation to just sort of use your intuition to think about it. All right.

Now, this is a pretty simple example. So now I'm going to do one that's a little more interesting and a little more devious in a sense. So I need a volunteer. We're going to do a gambling game. So I need a volunteer. All right. Now, he's raised his hands before I even said what the criteria are. First, you gotta be confident you understand probability. Well, you can fake it. But second, and more importantly, you've got to have some money in your pocket or you've got to borrow some on the way down. You got a few bucks there?

**AUDIENCE:** Yes.

**PROFESSOR:** All right, come on down. That's really all that matters, you know. OK, now, in this game, there's three dice. Well, that's 'cause you're smart. OK, so let me show you the dice in this game. And these, of course, are not normal dice. All right? So dice A has a 2, a 6, and a 7, and they're the same on the reverse side. So it's a six-sided die, and it's fair. When you roll it, the probability A comes up 2 is going to be  $1/3$ . All right? Same for 6 and for 7. Ah, we're just getting started.

Dice B has a 1, 5, and a 9 and dice C has a 3, 4, and an 8. So I got-- they're all different numbers, the numbers from 1 to 9. So just so we're really clear, the probability that A comes up 3 is  $1/3$ . The same for the probability A comes up 6 is  $1/3$  and so on. Ah, yes I do mean 2. Good. All right, good, you're paying attention. That's good. There's some hope. All right?

Now, here's the game. We're each going to pick a die. And we're going to roll our die. And the higher die wins. We're going to roll them once. The higher one wins. And the loser pays the winner \$1. And no tears, OK? You got that? That's good. He's already trying to figure out which one to pick. So let's put your dollar on the table. Here we go.

**AUDIENCE:** I have a \$5. Do you have change?

**PROFESSOR:** I might have change. All right. Let's see here. Yes, I do have change.

**AUDIENCE:** Oh, I do have a dollar.

**PROFESSOR:** You have a dollar. All right. Well, for now anyway. Let's put the dollar there on the table.

**AUDIENCE:** It's going to be a \$20 next.

**PROFESSOR:** We're just getting warmed up here. OK, now, there is-- what's your name, by the way?

**AUDIENCE:** Will.

**PROFESSOR:** Will. OK, nice to meet you, Will. There we go. All right. Now, there is one problem here. I don't really have those dice. So we're going to have to do a mathematical version of this bet.

**AUDIENCE:** You're always going to win, then.

**PROFESSOR:** So what we're going to do is pretend that we did a roll, head-to-head, and whichever one is more likely to win, we're going to assume the roll came out that way, all right? Now, I know you've been studying those three guys there, those three dice. I'm going to be fair and let you pick first which one you like.

Now, as you look at this, you know, you guys can help him out here if you can figure out which one's best. This has got a couple good sized numbers. You know, two of these beat these. This has the biggest number, but it also has the smallest. This has two small guys in it. You now and you want to pick the one that's likely to--

**AUDIENCE:** They all add up to the same thing.

**PROFESSOR:** They do all add up to the same thing, that's true. Yeah. Which one do you want? Which one do you want here Will? What's your best one there? A does look pretty good. You've got two solid numbers there. It does. A looks like it beats up on C pretty good.

**AUDIENCE:** I don't know. I think I'm going to lose a dollar.

**PROFESSOR:** That's probably true.

[LAUGHTER]

[AUDIENCE CHATTER]

I don't know. A looks pretty good to us. Which one do you like here? We'll try them out. We'll play more games here. A. All right, Will likes A. All right. And I'm going to pick-- I'm going to go with C. I'm going to give you a fighting chance, right? Because we figured A beat C for sure,

right? Because you've got a couple of 6 and 7 beat up 3 and 4 pretty good. Yeah, I'm going to pick C. All right.

And now we're going to do-- figure out which one's more likely to win. And we're going to make the tree diagram. All right? So here's A. You've got 2, 6, and 7 are your choices. And they're each  $1/3$ . And then, for mine, I've got 3, 4, and 8. 3, 4, and 8. 3, 4, and 8. And they're all  $1/3$  chance. All right, so I make my tree here. I got squished at the bottom, but they're all  $1/3$ . All right, now we're going to see who wins, C or A. There's nine outcomes. Who wins this outcome? Oh, I do. Yeah, I get all these. That's good. You're in trouble already.

What happens here, though? Who wins that? Your 6 beats 3, 6 beats 4. You win those. I got you here. Then you win 7 beats 3. A wins. 7 beats 4, A wins. It's tied, but, oh, my 8 beat your 7. All right. So let's see here. There's one, two, three, four, five for me, each at  $1/9$ . All right, so I'm going to win. All right?

Now, in this case, the sample space is uniform. Every probability is the same for the outcomes. They're all  $1/9$ .  $1/3$  times  $1/3$ .  $1/3$  times  $1/3$ . All right? So that's the case where we say the sample space is uniform. Don't go anywhere. We're going to do some more here. The sample space is uniform if every sample point has the same probability. In which case, the probability of the sample points would be  $1$  in the cardinality of the sample space. In this case, there's nine points. They each have probability  $1/9$ . And in this case, C beats A with probability of  $5/9$ . That means I win the dollar. All right.

So do you have some more money in there? In that wallet? I'll make change for you. So, what we're going to do is play again. And you've learned something now, that C beats A. So we play again, you're probably not going to want to pick A. In fact, C is surprisingly good there. Now, I'll let you wager more, if you want, to win your dollar back. Do you want to do double or nothing?

[LAUGHTER]

**AUDIENCE:** So you always get to pick second, though?

**PROFESSOR:** Yeah. I've got to be fair here. I'm going to let you pick first.

[LAUGHTER]

All right, you got a couple bucks? We'll do double or nothing. What do you say? I'll make

change for you. There you go. Here's \$5. I'll change your \$5. And then we'll bet \$2. How's that? I've actually got a \$2 bill here. There we go. You've got \$2. There we go. All right, double or nothing. Will, what do you like? You're obviously not going to pick A. That just got beat by C. What do you like, B or C?

Well, one of them's got to be best. One of them's got to be best here. You've just got to figure out which one it is.

**AUDIENCE:** I feel like everyone's going to have [INAUDIBLE]. If I choose C, let me see. That would be [INAUDIBLE] so I guess.

**PROFESSOR:** I won't pick A, that's for sure.

**AUDIENCE:** If you pick B, I get one from 4, I get two from the 8, which is four. And you get the other five. If I pick C and you pick A, that means I get one from the 3, I get one from the 4. If I get two from the 8, you still get 5.

**PROFESSOR:** That's good. You've figured out you're screwed here. That's very good. All right. So, yeah, so he's sort of-- he's doing great. He's figured out, because he went through this sort of like this method, what all the possible outcomes were. And he's discovered that, no matter what he picks, I'm going to beat him. So let's see why that's the case. So let's see. So because you figured it out, I won't take your money this time. But we'll do some more in a minute.

So say Will picked-- say he picked C. And say that then I will pick B. All right, so on C, there's three choices. 3, 4, and 8. And then, on my choice of B, there's 1, 5, and 9. All right, so I get 1, 5, and 9. And now we can see who wins here. Well, C wins there, loses here, loses there. C wins here, and B wins the next two. And C wins these. And B wins the tiebreaker. So B beats C with probability 5/9.

All right. So you don't want to pick A, you don't want to pick C, you want to try a bet with B? I mean, look, you've got B beats C. We just showed that. C beats A. I mean, that's a sure bet if I ever saw one.

[LAUGHTER]

**PROFESSOR:** What do you think?

**AUDIENCE:** So if I pick B and you pick C, that means [INAUDIBLE]

[AUDIENCE CHATTER]

But if he picks A, then I get [INAUDIBLE] Yeah, I'd still get 4, so--

**PROFESSOR:** So if you pick B, what am I going to do?

**AUDIENCE:** Oh, you're going to pick A.

**PROFESSOR:** I'm going to pick A. Let's see how that works out. All right, so B has 1, 5, and 9. A is, yeah, 2, 6, and 7. OK, so you're going to lose all three there. A wins those. A wins-- no, you win. B wins there, but A wins the next couple. And then, you win the last three.

So A wins, again, 5/9 of the time. So how is this possible? We've got A beats B, beats C, beats A. All 5/9. How is that possible? Did we make a mistake? Well, why isn't it possible? Whoever said probabilities were transitive? Right? I mean, you'd sort of think there's a best of the three die. But for any one that you pick, that he picks, I can find one that's better. And it goes in a cycle. All right?

So you'd think picking first, you've got a shot to pick the best one. But picking second always wins this game, if you pick the right one. All right. You know, since you did good and figured that out, I'm going to give you a special opportunity here. All right?

Now, I'm going to let you go second. All right? So I'm going to pick one, and I'm going to let you pick after me. And then we're going to bet. There's going to be a little twist to it. But you want to pick second, right? All right, well, let me tell you what I'm going to do. Just to make it-- I want to do a different analysis. We already did these three, so what I'm going to do is an analysis where we roll them twice. Otherwise, it's the same basic thing, but I want to do an analysis with more outcomes here. We've sort of beat this one to death.

So I'm going to go first. And let's see, what do I want to pick? Hm. Well, what should I pick, if I'm going to do this? I guess I'll pick B, all right? So I'm going to pick B here. All right. How much do you want to bet here? I'm going to let you win your money back. I'll let you bet up to \$10. Your choice. What do you-- what do you want to bet here? We've got to put some money on the table, make it a little interesting. Yeah, you pick--

**AUDIENCE:** Do you roll a different die or what?

**PROFESSOR:** No, the same ones. The die don't change. I'm picking B. And I'm going to roll it a couple times.

And I'm going to add up my score. And then, you get to pick whatever die you want. I'm going to let you pick. I mean, I'm going to let you pick the one that beats it. You roll yours a couple times. And now we see which one has the higher sum, who's more likely to win.

**AUDIENCE:** So it's the sum total, like you roll a 9 and 9 and I roll a 7 and 7.

**PROFESSOR:** Yeah, then I'm going to win, because I get 18 and you got 14. If I roll a 9 and a 1, I get a 10. You roll a 6 and a 7, you beat me at 13. All right? So which one do you want to pick? You could pick C. But you've got to put the money on the table here, whatever you want to bet. You know, I want to give you a chance to win your money back. What's the probability he's going to lose?

[LAUGHTER]

You got a T-Card? What've you got in there? Ah, I'll take the [INAUDIBLE] card. What's that worth? Ah, it's worth a few bucks here. All right, so I'm picking B. Which are you picking? I've really got his head messed up now.

**AUDIENCE:** There's \$4 on the table. Take that and run.

**PROFESSOR:** Yeah, really. We already saw-- I'll give you a hint here. A beats B.  $\frac{5}{9}$  of the time, A comes up higher than B. So if we roll them twice, your odds got even better that A beats B. All right, so Will goes with A. Will, that was a big mistake. All right. How can it be a mistake? A beats B  $\frac{5}{9}$  of the time. Now what we're going to do is make the decision tree. If we roll B twice, if we roll A twice. All right. Let's see what happens. This is a nasty one to write down. In fact, you'll start seeing this when you do these tree method things. Sometimes the trees get a little hairy, so we've got to sort of approximate it a little bit.

So here's B, right? I got 1, 5, and 9 for my first roll. And then I got 1, 5, and 9 for my second roll. And I add these up. See, if you read the notes ahead, sometimes that helps. 6-- here are my outcomes. These are the numbers I'm going to get. These are my nine values.

Your nine values, you got 2, 6, and 7. Then each 2, 6, and 7 again. So we add those up, you got 4, 8, 9, 8, 12, 13, 9, 13, and 14. And now, we have each one of these against each one of these. Because there's 81 outcomes. I got four levels to the tree. And rather than draw four levels of this [INAUDIBLE] tree to get 81 outcomes, you're going to say that each one of these corresponds to this tree, all right, that has nine leaves. And now we can just count up how many times I win.

Well, 2 got nothing. That's a 0. Let's see how many times 6 wins. One-- one time. 10-- one, two, three, four, five. All right, we've got that. 6 wins, we already decided, once. 10, we decided already, wins five times. 14 wins one, two, three, four, five, six, seven, eight, and ties once. Tie-- one. 10, we already got was five. 14, we already got, was eight and a tie. And finally, I get 18 wins every time. Let's add this up. I got 17, 22, 30, 35, 36, 41, 42 times plus half the ties out of 81. How much money's on this [INAUDIBLE] card? Ah, smart guy.

How did this happen? A is more likely to beat B, higher than B. But you roll them twice, and it's less likely to be as large on the sum? Well, why not? It's just our intuition's all screwed up. But the math is straightforward. In fact, an amazing thing happens. If you roll them twice, they reverse completely. A is worse than B, is worse than C, is worse than A if you roll them twice. If you roll them three times, something else happens.

In fact, it was known for a long time, this phenomenon happens with one roll. And when I was making lecture notes a few years ago, I noticed they reverse with two rolls. And then we did a class project to discover what happens with  $k$  rolls. And it turns out that you can get every possible combination. You know, on some number of rolls, A beats both B and C and B beats C. Other number of rolls, A beats them both and C beats B.

And then, motivated by that, some researchers out in San Diego showed that there's arbitrarily strange dice, now, not with just three sides, but with more sides. Such that, for any tournaments-- remember a tournament from graph theory, where everybody beats, you know, between a pair of players, one player wins. You can make the relation for any tournament show up for a particular set of dice by rolling a certain number of times. All right, which is just really fascinating that you can prove that. So these dice can get arbitrarily weird in terms of which die is likely to come up and be better than another die.

All right, that's it for today. Let me give you your money back here. And I've got a gift certificate you can have for being a good sport. Here you go. How much money did I take here?

**AUDIENCE:** I think only like \$2.

**PROFESSOR:** \$2? All right, there you go. Very good. Thanks. So we'll do a lot more examples over the next week or two.