PROFESSOR PATRICK WINSTON: I was in Washington for most of the week prospecting for gold.

Another byproduct of that was that I forgot to arrange a substitute Bob Berwick for the Thursday recitations.

I shall probably go to hell for this.

In any event, we have many explanations, none of them good.

But today we'll try to get back on track and you'll learn something fun.

In particular you will learn how a graduate student of mine Mark [? Phillipson ?], together with a summer UROP student, Brett van Zuiden, one of you-- managed to pull off a tour de force and recognize in these two descriptions the pattern that we humans commonly call "revenge." It was discovered.

The system didn't have a name for it, of course.

It just knew that there was a pattern there and sat waiting for us to give a name to it.

That's where we're going to end up.

But it'll be a bit of a journey before we get there, because we've got to go through all that stuff on the outline.

And in particular, we want to start off by a little tiny bit of review.

Because some of the stuff we did last time went by pretty fast.

In particular, you may remember they had this wonderful joint probability table, which tells us all we want to know, all we want to know.

We can decide what the probability of the police being called is given the this and the that, and all that sort of stuff, by clicking the appropriate boxes.

The trouble is, gee, there are only three variables there.

And when there are lots of variables it gets pretty hard to make up those numbers or to even collect them.

So we're driven to an alternative.

And we got to that alternative just at the end of the show a week ago.

And we got to the point where we were defining these inference nets, sometimes called "Bayes nets." And the one

we worked with looked like this.

There's a burglar, a raccoon, the possibility of a dog barking, the police being called, and a trash can being overturned.

So more variables than that.

That only has three.

This has got five.

But we're able to do some magic with this because we, as humans, when we define-- when we draw this graph we're making an assertion about how things depend or don't depend on one another.

In particular, there's something to break down and memorize to the point where it rolls off your tongue.

And that is that any variable on this graph is said by me to be independent of any other non-descendant given its parents.

Independent of any non-descendant given its parents.

So that means that the probability of the dog barking, given its parents, doesn't depend on T, the trash can being overturned.

Because the intuition is all of the causality is flowing through the parents and can't get to this variable D without going through the parents.

So that is [? inserted ?] property of the nets that we draw.

And we tend to draw them in a way that reflects causality.

So it tends to make sense.

So somehow this thing is going to be-- we're going to use this thing instead of that thing.

But wait.

We may need that thing in order to do all the computations we want to perform.

So we need to be able to show that we can get to that thing by doing calculations on this thing.

So what to do?

Well, we're going to use the chain rule.

And remember that the chain rule came to us by way of the basic Axioms of Probability plus the definition plus a little colored chalk.

So we got to the point last time where we sort of believed this.

It's a really magical thing.

It says that the probability of all this stuff happening together is given as the product of a bunch of conditional probabilities.

And the conditional probabilities in this product are arranged such that this first guy depends on everybody else.

The second guy doesn't depend on the first guy but depends on everything else.

So that list of dependencies gets smaller and smaller as you go down here until it depends only one thing.

There's no conditional at all.

So that's going to come to our rescue because it enables us to go from calculations in here to that whole table.

But first I have to show you a little bit more slowly how that comes to be.

One thing I'm going to do before I think about probability is I'm going to make a linear list of all these variables.

And the way I'm going to make it is I'm going to chew away at those variables from the bottom.

I've taken advantage of a very important property of these nets.

And that is there no loops.

You can follow the arrows in any way so as you get back to yourself.

So there's always going to be a bottom.

So what I'm going to do is I'm going to say, well, there are two bottoms here, there's C and T. So I have a choice.

I'm going to choose C. So I'm going to take that off and pretend it's not there anymore.

Then I'm going to take this guy.

That's now a bottom because there's nothing below it.

I've already taken C out.

So we'll take that out next.

And now I've got this guy, this guy, and this guy.

This guy no longer has anything below it.

So I can list it next.

Now over here I've got raccoon and trashcan.

But trashcan is at the bottom.

So I've got to take it next because I'm working from the bottom up.

I want to ensure that there are no descendants before me in this list.

So finally I get to raccoon.

So the way I constructed this list like so ensures that this list arranges the elements so that for any particular element, none of its descendants appear to its left.

And now that's the magical order for which I want to use the chain rule.

So now I can write-- I can pick C to be my variable n.

And I can say that the chain rule says that the joint probability of all these variables P of C, D, B, T, and R-- the probability of any particular combination of those things is equal to the probability of C given everybody else.

Next in line is D given everybody else.

Next in line is T-- next in line is B given everybody else.

And next in line is T given everybody else.

And finally, just R. So this combination of things has a probability that is given by this chain rule expression.

Ah.

But first of all, none of those expressions condition any of the variables on anything other than non-descendants,

all right?

That's just because of the way I've arranged the variables.

And I can always do that because are no loops.

I can always chew away at the bottom.

That ensures that whenever I write a variable, it's going to be conditioned on stuff other than its descendants.

So all of these variables in any of these conditional probabilities are non-descendants.

Oh wait.

When I drew this diagram, I asserted that no variable depends on any non-descendant given its parents.

So if I know the parents of a variable I know that the variable is independent of all other non-descendants.

All right?

Now I can start scratching stuff out.

Well, let's see.

I know that C, from my diagram, has only one parent, D. So given its parent, it's independent of all other non-descendants.

So I can scratch them out.

D he has two parents, B and R. But given that, I can scratch out any other non-descendant.

B is conditional on T and R. Ah, but B has no parent.

So it actually is independent of those two guys.

The trashcan, yeah, that's dependent on R. And R over here, the final thing in the chain, that's just a probability.

So now I have a way of calculating any entry in that table because any entry in that table is going to be some combination of values for all those variables.

Voila.

So anything I can do with a table, I can do in principle with this little network.

OK?

But now the question is, I've got some probabilities I'm going to have to figure out here.

So let me draw a slightly different version of it.

So up here we've got the a priori probability of B. Well, that's just probability of B. Down here with the dog, I've got a bigger table because I've got probabilities that depend on the values of its parents.

The probability of dog barking depends on the condition of the parents, nothing else.

So let's see.

I've got to have a column for B. I've got to have a column for the burglar and the raccoon.

And there are a bunch of possibilities for those guys.

But once I get those then I'll be able to calculate the probability of the dog barking.

So there are two of these variables.

So there are four combinations.

There's T T. There's T R, R T, and-- whoa, what am I doing?

Wake up!

T false.

False true.

And false false.

So what I really want to do is I want to calculate all of these probabilities that give the probability of the dog condition of the burglar and the raccoon.

Similarly, I want to calculate the probability of B happening doesn't depend on anything else.

So I don't know what to do.

Well, what I'm going to actually do is I'm going to do the same thing I had to do up there.

I'm going to keep track of-- I'm going to try a bunch of-- I'm going to get myself together a bunch of data.

Maybe I do a bunch of experiments.

Maybe somebody hands it to me.

But I'm going to use that data to construct a bunch of tallies which are going to end up giving me the probabilities for all of those things.

So I don't know, let's see.

How should we start?

Step one, find colored chalk.

Step two, I'm going to extend these tables a little bit so I can keep track of the tallies.

So this is going to be all the ones that end up in a particular row.

And these are going to be the ones for which dog is true.

Similarly, I'm going to extend this guy up here in order to keep track of some tallies.

This is going to be the ones for which B is true.

And this one will be all.

So that's my set up.

And now suppose that my first experiment comes roaring in.

And it's all T's.

So I have T T T. That's my first experimental result, my first data item.

So let's see.

The arrangement here is burglar, raccoon, dog.

So burglar as a true.

And there's one tally count in there.

Likewise, the T T, that's the burglar and the raccoon, that brings me down to this first row.

So that gives me one tally in there and dog is true so that gives me a tick mark in that one.

All right?

Are you with me so far?

And now let's suppose that the next thing happens be all false.

Well, burglar is false.

But there is one experiment.

Everybody's false.

So we come down here to false false.

And that's the row we're going to work on.

We get a tally in there.

Do we put one in here?

No, because that's false.

Dog is false.

That's what our data element says.

So that's cool.

Maybe one more.

Let's suppose we have T T F. Well in that case, we have a tick mark here and a tick mark here because the burglar element is true.

Then we have T T. That brings us to the first row again.

So we get a tick mark there.

But dog as false, so no tick mark there.

That's how it works.

I suppose you'd like to see a demonstration, right?

Always like to see a demonstration.

So here's what it actually looks like.

So on the left you see the network as we've constructed it, with a bunch probabilities there.

And what I'm going to do now is I'm going to start simulating away so as to accumulate tick marks, tally marks, and see what kinds of probabilities that they indicate for the table.

I happen to be using a process for which the model on the left is a correct reflection.

So there's one simulation.

So the dog barking-- let's see, the burglar is false.

The raccoon is true.

I get one tick mark.

So the probability there is one.

Of course, I'm not going to just go with one.

I want to put a whole bunch of stuff in there.

So I'll just run a bunch more simulations.

No [? dice. ?] I don't even have an entry at all yet for T F here.

That's because I haven't run enough data.

So let me clear it instead of doing it one at a time.

Let me run 100 simulations.

See, it's still not too good.

Because it says this T T probability true.

This just because I'm feeding it data, right?

And I'm keeping track of what the data elements tell me about how frequently a particular combination appears.

Yes, [INAUDIBLE] STUDENT: So when you're doing one simulation, is that [INAUDIBLE] variables?

PROFESSOR PATRICK WINSTON: When I'm doing one simulation, I'm just keeping track of that combination in each of these tables.

Because it's going to tell me something about the probabilities that I want reflected in those tables.

So it's pretty easy to see when I go up here to burglar.

If I have a lot of data elements, they're all going to tell me something about the burglar as well as the other variables.

So if I just look at that burglar thing, the fraction of time that it turns out true over all the data elements is going to be its probability.

So now when I go down to the joint tables, I can still get these probability numbers.

But now they're conditioned on reticular condition of its parents.

So that's how I get these probabilities.

So I didn't do too well here because that T T combination gave me an excessively high probability.

So maybe 100 simulations isn't enough.

Let's run 10,000.

So with that much data running through, the probabilities I get-- let's see, I've got 893 here, instead of 0.9, 807 instead of 0.8, 607 instead of 0.6.

And that one's dead-on at 0.01.

So if I run enough of these simulations, I get a pretty good idea what the probabilities ought to be given that I've got a correct model.

OK, so that takes care of that one.

And of course, I didn't draw the other things in here.

But by extension, you can see how those would work.

Oh.

But you know what?

I think I will put a little probability of raccoon table in here.

Because the next thing I want to do is I want to go the other way.

This is recoding tallies from some process so I can develop a model.

But once I've got these probabilities, of course, then I can start to simulate what the model would do.

All right?

How would I do that?

Well, do I want to use the same table?

I think just to keep things sanitary, what I'll do is I'll go over here and do it again.

Here's B. It's got a probability of B. Here's R.

Here's a table probability of R. That comes down into a joint table for dog.

And it's got four elements.

Depending on the burglar condition and the raccoon condition, we get a probability of dog.

And now, imagine these have all been filled in.

So what do I want to do if I want to simulate this system generating some combination of values for all the variables?

Well, I do the opposite of what I did when I was working around with this chain rule showing that I could go from the table to those probabilities.

Now I've got the probabilities.

I'm going to go the other direction.

Instead of chewing away from the bottom, I'm going to chew away from the top.

Because when I go into the top and chew way, everything I need to know to do a coin flip is there.

So in particular, when I go up in here, I've got the probability of burglar now.

So I'm going to use that probability to flip a coin.

Say it produces a T. So that takes care of this guy.

And I can now scratch it off since it's no longer in consideration.

It's no longer a top variable.

So now I go over into raccoon and I do the same thing.

I take this probability.

I do a flip.

And say it produces an F. Whatever its probability is, I flip a biased coin and that's what I happen to get.

But now, having dealt with these two guys, that uncovers this dog thing.

And I've got enough information, because I've done everything above, to make the calculation for whether to dog is going to be barking or not.

But wait.

I have to know that I've got a T and a T and a T and an F and an F and a T and an F and an F. Because I have to select the right row.

So I know that B is T. And I know that R is F. So that takes me into the table into the second row.

So now I get this probability.

I flip that coin and I get some result, say, T. Voila.

I can do that with the other two variables.

And I've got myself an experimental trial that is produced in accordance with the probabilities of the table.

OK?

Of course-- yeah, in fact, how did I get those numbers?

Actually what I did is I used the model on the left to generate the samples that were used to compute the probabilities on the right.

So you've seen that a demonstration of this already.

Now of course-- I don't know, all of this sort of depends on having everything right.

I've written a thing to write it one more time.

Burglar, raccoon, dog, call the police, trashcan.

But somebody else may say, oh, you've got it all wrong.

This is what it really looks like.

The dog doesn't care about the raccoon at all.

So that's a correct model.

Now when I do a simulation, I could fill in the tables in either model, right?

I'm sure you'd like to see a demonstration.

So let me show you a demonstration of that.

So there are the two tables.

And I can run 10,000 simulations on those guys, too.

Now, look.

The guy on the left is a pretty good reflection of the probabilities in a model I used to produce the data.

But the guy on the right doesn't know any better. it just fills in its own tables, too.

So what to do?

I say this one's the right model.

And you say that one's the right model.

Who's right?

Maybe we'll never know.

And the guy on the left will get rich in the stock market and the guy on the right will go broke.

I would be nice if we could actually figure out who's right.

So would you to see how to figure out who's right?

Yeah, so would I. What we're going to do is we're going to look at naive Bayesian inference.

And that's our next chore.

So here's how it works.

We know, from the definition of conditional probability, we know that the probability of A given B is equal to the probability of A and B divided by the probability of B, right?

Equal to by definition.

So that means that the probability of A given B times the probability of B-- I'm just multiplying it out-- it equal to that joint probability.

Oh, but by symmetry, there's no harm in saying I can turn that around and say that the probability of B given A times the probability of B is also equal to that joint probability, right?

I've just expanded it a different and symmetric way.

If I've got to write a, b on B, b, a on A. Thank you.

Who was complaining?

Good work.

That would have been a major-league disaster.

But now, having written that, I can forget about the middle.

Because all I'm really interested in is how I've turned the probabilities around in that conditional.

Why would I care about doing that?

By the way, we're now talking about the work of the Reverend Bayes.

Because we can rewrite this yet again as the probability of A given B is equal to the probability of B given A times the probability of A divided by the probability of B.

That's just elementary algebra.

But now I'm going to do something magical.

I'm going to say I've got a classification problem.

I want to know which disease you have.

That's a classification problem.

Maybe you've got the swine flu.

Maybe you've got indigestion.

Who knows.

But I get all these symptoms.

I get all these pieces of evidence.

You've got a fever.

You're throwing-- oh, well, let's not go into too much detail, there.

But what I'm going to do is I'm going to say, well, let's suppose that A is equal to a class that I'm interested in, the disease you've got.

And B is equal to the evidence, the symptoms I observe.

Voila.

I may have a pretty hard time figuring out what the probability of the class is given the evidence.

But figuring out the probability of the evidence given the class might not be so hard.

Let me get another board in play and show you what I mean.

By plugging class and evidence into Bayes' rule, what I get is the probability of some class given the evidence is equal to the probability of the evidence given the class times the probability of the class divided by the probability of the evidence.

Now you've got to let that sing to you a little bit.

Suppose I've got several classes that I'm trying to decide between.

I'm trying to select the best out of that batch of classes.

Well, I've got the evidence.

And if I know the probability of the evidence given each of those classes, and if I know, a priori, the initial probability the class, then I'm done.

Because I've got the two elements in the numerator.

Why am I done?

Because the denominator is the same for all the classes.

It's just the probability of the evidence.

And then I could just sum everything up.

I know it adds to 1 anyway.

So that's cool.

But sometimes there's evidence-- actually there's more than one piece of evidence.

Let's say that there's some class.

some i, and we're trying to figure out if that's the correct class.

So we've got c sub i there and c sub i there.

And suppose that that evidence is actually a bunch of pieces of evidence.

So it could be e sub 1, e sub n, oops, premature right bracket.

All that evidence, given the class i times the probability of the class i over some denominator that we don't care

about because it's going to be the same for everybody.

So we'll just write that as d.

Now what if these pieces of evidence are all independent given the class?

So if you have the swine flu, the probability you have a fever is independent of the probability you're going to throw up, say.

Then can we write this another way?

An easier way?

Sure.

Because when things are independent, the joint probability is equal to the product of the individual probabilities.

So that is to say-- it's easier to see it if you write it down than if you just say it-- this probability here from these two elements here is equal to the probability of e sub 1 conditioned on c sub i times the probability of e sub 2 conditioned on c sub i, all the way down to the probability of e sub n conditioned on c sub i divided by some denominator we don't care about.

See, what I'm going to try to do is I'm going to go through this for all the ci and see which one's the biggest.

STUDENT: That's the [INAUDIBLE] ci, right?

PROFESSOR PATRICK WINSTON: This is the probability of-- STUDENT: [INAUDIBLE] right-hand side [INAUDIBLE].

PROFESSOR PATRICK WINSTON: Right here?

Oh yes, you're quite right.

Oh yeah, thanks.

I can't write and think at the same time.

Thanks.

OK.

So I've just figure out which one of these is the biggest.

And I've identified the class.

Now you say to me, well, I would like to see an example.

So-- I don't know, does anyone have any spare change?

A nickel, a quarter.

This is not because of infinitesimally low raises here at MIT.

I just need it for a demonstration.

I need two coins.

Don't forget to get these back, I tend to be-- Now suppose these two coins are not exactly the same.

One of these points is a legitimate, highly-prized American quarter.

The other one is a fake.

And with this one, the probability of heads, let us say, is 0.8 instead of 0.5.

So I mix these all up.

And I pick one.

And I start flipping it.

And I get a head.

Then I flip it again.

And I get a tail.

Which coin did I pick?

Well, we're going to use this stuff to figure it out.

Here's what happens.

Before I forget.

Thank you very much.

So what we've done is we've selected these things from my hands.

And I can't draw hands.

So I'll draw a little cup here.

And there are two coins in here.

And we're going to pick one.

And one has a probability of heads equal to 0.8.

And this one has a probability of a head of 0.5.

So here's the draw.

I pick one.

Each has a probability of 0.5.

This one is the one with the 0.8 as the probability of head.

And this one is the one with the probability of 0.5 as a head.

OK?

So now suppose the first flips as it was is T. Well, that's a piece of evidence.

That's here.

Probably of evidence given the class.

Well in the case of having drawn this biased coin, the probability of coming up with a tail-- ah, let's say a head, just to make my numbers a little easier.

Probability of coming out there with a head is equal 0.8 given that it's up here in this choice.

The probability given that you have a fair coin is 0.5.

So now if we take the next coin and take it to be a tail then the probability of this guy given that evidence is 0.2.

And the probability of this guy given that evidence-- it's a fair coin, so it doesn't care.

It's still 0.5.

So now what's the probability of this class given this evidence?

It's the product 0.5 times 0.8 times 0.2.

And what's the probability of this guy?

It's 05 times 0.5 times 0.5, divided by a denominator which is the same in both cases.

So let's forget about this early 0.5 here.

Because it's the same in both cases.

And we just multiply those numbers together.

That gives us 0.8 times 0.2.

What's that?

0.16?

And this guy, 0.5 times 0.5, that's 0.25.

So it looks an awful lot like-- with this combination-- that I've picked the coin that's fair.

One more flip?

So let's flip it again, and suppose we come up with a head.

So that puts a 0.8 in here.

And 0.5 in here.

When you multiply those out that's 0.125.

And this is 0.128.

So it's about equal.

So you see how that works?

All right.

So we're using the coin flips as evidence to figure out which class is involved.

OK so I don't know, you'd probably like to see a demonstration of this, too, right?

You say to me, gosh, just two kinds of coins.

That's not very interesting.

Let's try five kinds of coins.

So what I want to show you is how the probabilities for all these coins-- there are five of them, color-coded-- how the probabilities vary with a series of flips.

Let's suppose I've got a head-- the grey line, by the way, is the fraction of heads-- so that's going to be one.

Because I'm just doing heads.

You see that black line rising?

Should look like a rocket.

That's the probability that the-- that's the coin which only shows heads, the probability of head is 1.

And I'm flipping a whole bunch of heads here.

Isn't that cool?

Now what happens if I suddenly put in a tail?

By the way, you'll no doubt, here one the extreme left-- the initial probability of the P=0 coin was 0.1.

As soon as I flipped a head that went to 0.

And it will never get off 0, right?

That makes sense.

Because if the probability that you'll get a head is 1 you should never see a tail.

If you ever do, that isn't your coin.

What happens now if I interrupt a series of heads and produce a tail?

STUDENT: [INAUDIBLE].

PROFESSOR PATRICK WINSTON: What's that?

STUDENT: [INAUDIBLE].

PROFESSOR PATRICK WINSTON: The black one will go to 0.

What else happens?

By the way, the blue one is the one with the highest probability of being a head.

[INAUDIBLE] Boom!

That blue one shot up.

Not going up slowly.

It shot up.

Because now the preponderance of evidence with all those heads is that I've flipped the coin with a bias of 0.75 towards heads.

So let's clear this.

Pick any probability you want.

0.25, 0.5, and so on.

I don't know, let's pick 0.25 since we've been at the upper end.

So orange is 0.25.

And sure enough, the probability that I've selected the 0.5 coin is going up and up and up and up after the original irregularity.

The Law of Large Numbers is setting in.

And a probability that I've got that 0.25 coin in play is pretty close to 1.

All right.

So that's cool.

Now you say to me, that's awfully nice but stop.

Awfully nice, but not very real-world-ish.

So let me give you another problem.

It's well-known that you are, with high probability, of the same political persuasion as your parents.

So if I wanted to figure out which party a parent belongs to, I could look at the party that their children belong to, right?

So it's just like flipping coins.

The particular coin I have chosen corresponds to the parent.

Individual flips correspond to the political party that the child belongs to.

So let's get up a little bit-- by the way, I wrote all this stuff over the weekend.

So who knows if any of it will work.

But let's see.

A parent party classifier.

There it is, Democrats and Republicans.

And now the prior for being a Republican given here is 0.5.

But I don't know, this is a little bit Democratic state.

So let's adjust that down a little bit.

Somewhere in there might be about right But let's just, for the sake of a classroom illustration, go down here.

So now the meter is showing the prior probability because that's the only thing in the formula so far.

I've got no evidence.

So now let's suppose that child number one is a Republican.

Back to neutral.

So I've got a low probability that the parent-- a priori probability that the parent is a Republican and a child who's a Republican.

I notice that 0.2 and 0.8, the conditional is 0.8.

And the prior is 0.2.

That's why it comes out to balance each other, right?

So now if we get another Republican in there it goes way up.

If I have a Democratic child it goes back down.

If I have an equal balance between children then it goes way back down because of that prior probability being low.

So if I make that high, even though the children are balanced, I'm still going to have a high probability of being a Republican.

Now let's see.

If I take that slider there, the conditional probability, and drive it to the left here-- let me make that equally in.

And let's make that one thing.

I don't know.

What am I doing now?

If I make the probability less than 0.5, what's that mean?

That means you're sore at your parents and you want to belong to a different party.

All right, so now, what's next?

Oh gosh.

What's next?

This is what's next.

What's next to somewhere?

Yeah, this is what's next.

This here.

We've got two models.

Remember when I said we wanted to decide between them?

Can we use that Bayesian hack to do that, too?

Sure.

Because we've got these two models.

We've got the probabilities in them.

So now I can take my data and calculate the probability of a left model given the data and the probability of the right model given the data, multiply that times their a priori probabilities, which I'll assume are equal.

Then I can do a model selection deal much in defiance to what I was hinting at before.

so let's try that.

Whoa.

There are my two models.

Yes, there they are.

We've already trained them up.

And they've got their probabilities.

Now what we're going to do is we're going to use the original model to simulate the data.

So what we're going to do is we're going to simulate draws, simulate events, similarly combinations of all variables using a model that looks like the one on the left, that is the one on the left except for the slight differences in probabilities, OK?

Then we're going to do this Bayesian thing and see where the meter goes.

So we'll run one data point.

Oops, went the wrong way.

Makes me nervous.

I just finished this at 9:15.

Maybe there's a bug.

Oops, two data points, swings to the left.

Three data points, back to the right.

Of course that's not much data.

So let's put some more data in.

Yeah.

Boom, there it goes.

Let's try that again.

That was cool.

So let's run 1,000 simulations and one data point.

It bobbles around a little bit and goes flat over to the left.

Because that is the model that reflects the one that the data is generated from.

So now we got Bayesian classification, except now the classification has gone one step more and it becomes structure discovery.

We've got two choices of structure.

And we can use this Bayesian thing to decide which of the two structures is best.

Isn't that cool?

Well, it's only cool if you could do what?

So if you had two choices-- you can select between them and pick the best one-- but there are-- gosh, for this number of variables, there are a whole lot of different networks that satisfy the no looping criteria and don't have very many parents.

There's an awful lot of them.

In fact, if you strict this network to two parents there are probably thousands and thousands of possible structures.

So do I try them all?

Probably not.

It's too much work when you get 30 variables or something like that.

So what do you do?

We know what to do, right?

We're almost veterans a 6034.

We have to search!

So what we do is we take the loser and we modified it.

And then we modify it again.

And we keep modifying it until we drop dead or we get something that we're happy with.

So let's see what happens if we change this problem a little bit and do structure discover.

We're starting out with nothing linked.

And we're going to just start running this guy.

So what's going to happen is that the good guy will prevail.

And the bad guy will be a copy of the good guy perturbed in some way.

So it's a random search.

You'll notice that score-- it's too small for you to read.

All these things are too small to read.

Let me make it a little bigger.

Too small to read, but that number on the right there is not the product of the probabilities, actually.

It's the sum of the logarithms of the probabilities.

They go together, right?

And the reason you use this instead of the probabilities is because these numbers get so small that was a 32-bit machine, you eventually lose.

So use the log of the probabilities rather than the product of the probabilities.

You use the sum of the logs instead of the product of the probabilities.

And eventually, you hope that this thing converges on the correct interpretation.

But you know what?

This thing is so flat as a space and so a large and so telephone pole-like that it's full of local maxima.

So what this program is doing is every once in awhile-- I think with probability 1 and 10; I forgot what parameters I used-- every once in awhile, it'll do a total radical rearrangement of the structures.

In other words, it's a random restart.

It keeps track of the best guy so far.

And every once in awhile it does a totally random restart in its effort to search the space.

So that's how you go from probabilistic inference to structure discovery.

Now when is this stuff actually useful?

Well, I hinted at a medical diagnosis, right?

That's a situation where you've got some symptoms.

And you want to know what the disease is.

So as soon as you use the keyword "diagnosis," you've got a problem for which this stuff is a candidate.

So what other kinds of diagnosis problems are there?

Well, you might be lying to me.

So I can put a lie detector on you.

And each of those variables that are measured by the lie detector are an independent indication whether you're telling the truth or not.

So it's this kind of Bayesian discovery thing.

Naive Bayesian Classification.

What other kinds of problems speak to the issue of diagnosis?

Well, we like to know how well you know the material!

So we can use quizzes as pieces of evidence.

Thank god we don't use exactly a naive Bayesian classifier, because then we wouldn't be able to do that combination.

We have to use a slightly more complex-- what you can think of as a slightly more complex Bayesian net to do that particular kind of diagnosis.

You might have a spacecraft or an airplane or other piece of equipment with all sorts of symptoms.

You're trying to figure out what to do next, what the cause is.

So using the evidence to go backward to the cause.

So maybe you've got some program that doesn't work.

Happens to me a lot.

So I use the evidence from the symptoms of the misbehavior to figure out what the most probable cause is.

But now to conclude the day-- last time there weren't any powerful ideas.

But if you take the combination of the last lecture and this lecture to be a candidate for gold star ideas, these are the ones I'd like to leave you with.

We got here is-- this Bayesian stuff, all these probabilistic calculations are the right thing to do.

They're the right way to work when you don't know anything, which would make it sound like you're not very useful, because you think you always-- well, in fact, there are a lot of situations where you either can't know everything, don't have time to know everything, or don't want to take the effort to know everything.

So in medical diagnosis all you've got is the symptoms.

You can't go in there and figure out in a more precise way exactly what's wrong.

So you use the symptoms to determine what the cause is.

And then all those other kinds of cases that I mentioned.

But now, what other kinds of structure discovery are there?

Well, the kind of structure discovery that I hinted at in the beginning will be the subject that we'll begin with during our next and sadly final conversation here in [? 10250 ?] on Wednesday.

It will feature not only a discussion of how this stuff can be used to discover patterns and stories, but we'll also talk about what's on the final, what kind of thing you could do next, that sort of thing to finish off the subject.

And that's the end of the story for today.