WILLIAM GREEN: So today we're going to talk about Bayesian prior estimation and prior estimation in general. So the last time we were writing down the expressions for the probability of observing a mean measurement if you know what the model is. So let's try to do that again. So suppose I have a model that predicts some observable and it depends on some knobs, and it depends on some parameters.

And suppose because I have great powers of faith that I believe this model is 100% correct with every core of my being. And also because I have tremendous confidence in all the people who built my apparatus, and the knobs that I turn actually correspond to the real values, and I have tremendous confidence in all the literature that reports the parameter values. And so I'm absolutely certain that this is the truth. So we'll start from a position of absolute certainty, and then we'll degrade into doubt as the collector goes on.

So let's start from the position of someone who has absolute faith that this is the truth. Is true. So I have a model, I really believe this model. So for example, I believe that the kilogram weight in the SI Institute in Paris weighs exactly one kilogram. I believe that with every core of my being. I'm completely confident that model is correct. So there are some things I'm really confident on. That's one. And maybe guys have some things you really believe, too. So let's go with things we really believe.

So I plan to conduct some experiments that measure this observable and are related to this model. And so I'm going to do 10 repeats of measuring y. So I'm going to get to the kilogram blob that's in Paris, and I'm going to stick it on my really expensive scale that I really believe is great, and I'm going to measure its weight. And then I'm going to put it back, and then put it back, and put it back, and put it back. And I'm going to get another really great scale that I really believe is great, and I'm going to measure it there, too. So I've got a lot of repeats of measuring the weight of this kilogram, and I believe it's really a kilogram.

But the stupid measurements don't say a kilogram. They say, you know, 1.0003, 0.99995, all

kinds of numbers not equal to one kilogram. So now I'm going to try to figure out what the probability is that it would have measured some particular value y. So what is the probability that my experimental is between some value, say, y and y plus dy. So that's a question for you. So what's the probability? Sorry, what?

[INAUDIBLE]

OK, so we think that the probability that y is in this interval given that the model is true. And I know the theta values perfectly, and I know the x values perfectly, is equal to some integral of what? The bounds integral, probably y to y plus dy, believe that? What's the integrand? Sorry, what?

AUDIENCE: The probability [INAUDIBLE] AC function of y.

WILLIAM GREEN: Right, so what is it?

AUDIENCE: [INAUDIBLE]

You wrote it down last time, I think. So this [INAUDIBLE] is large? Standard normal, right? So it should be one over sigma root of 2 pi. Does that sound OK? I mean here. It's probably the same, it's fine.

Yep.

AUDIENCE: What does that notation mean, if your model is true?

WILLIAM GREEN: So this means, given that the model is true, and I know these data values are exactly certain numbers, and the x values are actually certain numbers, what's the probability that I would make a measurement whose average would fall in this interval? So this line means given that this is true, what's the probability of that? OK, is this right? Is this surprising? This is OK?

So this is what I mean. So we say that our probability distribution converges to a Gaussian distribution, this is what we expect. So we expect it to have been large enough for this to be true. Yeah? This is very important. This is like the whole course, actually. This is the whole section, is this one equation. So I just wanted to make sure you really get what this says. And if you don't like the integral, you can make dy really small, and then it's just this times dy. OK?

Actually, this notation's like [INAUDIBLE] I think I should do this way. I should do this. This is a number. Let's get rid of the integral. Let's make dy really small. I'll make it [INAUDIBLE]. That

all right? So this is the probability density that we would observe, this is the experimental value y that we observe from the mean, and this is the little with of our tiny little interval. Is that all right? Yes?

AUDIENCE: So is sigma the [INAUDIBLE] on there?

WILLIAM GREEN: Ah, what is sigma? That's a great question. We didn't write down what sigma was. What is sigma?

AUDIENCE: Standard deviation?

WILLIAM GREEN: It's not the standard deviation exactly. Standard deviation of the mean, right? So there's two sigmas. We have the sigma of y, of the measurements, and that's equal to average value of y squared minus. So we just figure that for how many experiments we do, we just compute the average of y squared, the average y, subtract them. That's the variance. And then sigma that I used in that equation there is 1 over n times sigma y. And we call this the variation of the mean, it's the uncertainty in the mean value of y.

And the central limits theorem said that as long as n gets really large, we expect that this should converge to this. And we talked last time about how when n get bigger, these averages don't really change when it gets big. They're just the average. But this number declined as n gets big because of this one over n formula.

And to understand that, suppose I measure the weight, and I measure, it should be around one kilogram. But in fact my measurements are all over here. Lots of measurements. So they have a variance something like this. But if I make a plot of-- as I run, I compute the running average. So when I run the first two points, I get some average value here. After I run 27 points more, the average value is here. After I run 1,000 repeats, the average value is here. It's getting pretty close to this, and the uncertainty in this number's getting smaller and smaller as I'm doing better and better averages. The average more and more repeats. Does that make sense? OK.

So from this key equation, I can derive a lot of things. And it depends what you want to do. So one thing people do a lot is it was called model validation. And what does this mean? It means I have a model, I believe it's true. I have some parameters, I believe they're true. But there are some foolish skeptics out there who don't have the faith that I do. And they think that my model's baloney, or my parameter values are wrong, or something. And so to prove I'm right,

I'm going to make some experiments. And I'm going to show that I make a plot that looks like the experiment and model agree. Some of you might have done this in your life, yes? Everybody might make a parity plot or something. You've seen these things before.

Now, this is like a confidence builder. You're trying to get the skeptics out there to believe that there's some evidence to back up your faith that this model is perfect. And what you really want to know is like, if the measurement that I measure, the average for my 10,000 repeated measurements, I expect that this quantity should be pretty big somehow in some way. By then quantitatively saying what that means exactly what's a good fit, what's a bad fit, this is actually kind of a difficult question, and we'll come back to this one. But that's a very common use of this equation is to try to do validation.

Now because it's kind of complicated, most people don't actually do it. So instead what they do is they just plot some data points, and they plot your model curve. And as long as they look good, then you're done. So that's the normal way that it's done in the literature currently. But of course, that's completely unquantitative. It doesn't really say whether the model and the data really agree, it just means they look sort of like each other. So that's like a human qualitative thing.

Now, if the purpose of validation is just to convince humans, then you've done the purpose. Now, if your purpose is to try to quantitatively say something, then you really have to get into this equation, which usually is not done but would be the right thing to do for validation.

Now the alternative view is disproving a model. But I just say that there's several ways this can happen. You can try to disprove a model, but you might also show that the theta values are incorrect. Or you might show that the experiment is wrong. These are all possibilities, reasons why the model and the data might not agree with each other.

So this equation, it only holds if the model is really true, if the parameter values are all perfectly correct, if we know exactly what all the knob values are perfectly. If any of those things are not true, then you should have some discrepancy, and there should be a way to show it. And really what you're showing is that you'd observed some y that is very unlikely to be observed. So probably observing that y is very extremely unlikely if all these other things were true. So if all these things are true, and you compute this value, and this value's very tiny, then it makes you think that it's unlikely that you would have observed that. And therefore, you might try to use that as an argument to say that something must be wrong. The model's

wrong, parameters are wrong, the knobs are wrong, something's wrong. My y values are wrong. It could be any of those things.

So this is often the most exciting papers to publish. You publish a paper, you take some model that a lot of people believe. You tell them they're full of baloney, it's completely wrong. My great experiment shows you are completely wrong. And so you'll see a lot of these in *Nature.* I should warn you, a lot of those get retracted later, a very high retraction rate in *Nature.* Because they want to publish papers like that that show that the common view is incorrect, and sometimes it's true. But oftentimes the common view is actually correct, and there's something wrong with the experiment, or the interpretation, or how they computed this equation, or whatever. And so actually it turns out the common view is perfectly fine, and it's just that the foolish authors went off on a tangent. And then they have to six months later publish a retraction, by the way, sorry, paper was completely wrong. And so you see a lot of that.

So that's a second kind of thing. And we'll talk more about that a little bit later, too. And then another thing is I'll relax my assumptions. So I'll say, well, I'm sure that the model is true, and I'm sure that my knob settings are perfect, and I know what they are. But I'm not really sure about all the parameters. And therefore I want to use the experiment to try to refine parameter values.

So I'm trying to take my y's that I measure and somehow infer something about the thetas. And this is a very common thing to do. So in my group we've tried to measure the rate coefficient for a reaction. We believe there is value of that theta, and in fact, we probably have an estimate of what it is. But we're not sure of the exact number, and we'd like to do an experiment to refine the number and get it more accurately determined. So that's another useful thing to do.

And this leads into two somewhat different points of view about this. One you've probably done already called least squares fitting. That's one view. And the other is this Bayesian view that I'll tell you about next. So there's sort of A and B. There's one that I'll call Bayesian, and one I'll call least squares. They're sort of related to each other, but not exactly the same conceptually. So I'll try to explain that.

So the Bayesian view is probabilistic, so it's actually pretty straightforward to write down. Remember that we wrote that the probability of A and B is equal to the probability of A times

the probability of B given A, and it's also equal to the probability of B times the probability of A given B. And what we have here is one of these conditional probabilities, if the thetas have a certain value, this is a certain probability.

So I should be able to use that formula somehow. So I can write down that the probability of measuring y given theta is equal to the probability of y times the probability of theta given y divided by the probability of theta. So I just took this formula, and I plugged in y's and thetas instead of A's and B's. So I said, these two are equal to each other. Rearranged it so then I can rewrite this.

This is the way we have it in here, probably of measuring y given theta. Let's flip it around. So probability of theta given that we measured y is equal to the probability of theta times the probability of observing y if theta was true divided by the probability of y. Terrible handwriting there. That's just algebra.

So this is what we want to know. We want to know, what's the probability distribution of the parameter values y? Because some of them are uncertain. Now, before we started the experiment, we had some idea of what the ranges were for all the primary values. Like I'm trying to measure a rate coefficient. I know from experience with other similar reactions, from a quantum chemistry calculation, from some indirect evidence, from some other more complicated experiment, I have some idea that this rate coefficient has to be in a certain range.

Now, it could be pretty uncertain. It might be five orders of magnitude uncertain. But I know it's not less than 0. I know it can't be faster than the diffusion limit, how fast things can come together. So for sure I know some range, and oftentimes I know a much narrower range than that. So I have some information about these parameter values before I even start. Some of the parameters of the model I know perfectly, or pretty well. So you know maybe there's a Planck's constant, or the heat of formation of one of my chemicals or something like that shows up in the numbers, and I might know that parameter really pretty accurately. Whereas the particular rate coefficient I care about is the thing I really don't know very well.

So some of these have tight probability distributions ahead of time, and some of them have loose ones. And this thing has a name. It's called the prior. And it's our prior information before we did the experiment. And this one, after we've done the experiment, we're going to change it. So we're going to say previously people thought that the parameters all lied in these certain

ranges. And now I'm going to get a tighter range, because I have some additional experimental information. So this is called the posterior. This means before, it means after. So this is what I know about parameter values before and after the experiment.

This is the formula that I have over there. It's a probability that if the thetas had a certain value, probably would have observed what I saw. Yeah?

AUDIENCE: Which one refers to which?

WILLIAM GREEN: Sorry, this is the prior, this is the posterior. And those of you who are paying attention to notation realize I'm not doing this very nicely. Because these are continuous variables, and I'm writing capital PRs, and they should be not be capital PRs, it should be probably density functions instead. So let's rewrite it nicely.

So the probability of theta given y, probability density is equal to the probability density of theta initially times the probability of y given theta, [INAUDIBLE] density divided by the probability density of-- all right? And what I just basically did was this is the correct equation the previous other one was all multiplied by d theta dy, it shouldn't be done that way. So this is OK?

Now this is the prior information I have about the parameter values. I know that they have to fall into some ranges. And really all I'm doing is I'm correcting that information. I'm improving the information to tighten the distribution. So initially I know that my rate constant, here's my rate constant. I know that it's got to be greater than zero, and I don't think it's really down there at zero anyway. I think it's somewhere in here. I really don't know much. And I really don't think it's all up at the diffusion limit, and no way it's higher than the diffusion limit. So that's my initial information that I have about the probability distribution of K. So it's the rate coefficient I want to know, and I know it's bigger than zero, and I know it's less than infinity. And actually I know there's some physical limit, it can't be higher than something or other.

And you can do this for any problem, right? I give you any parameter, you should be able to tell me something about it. You might be uncertain by 20 orders of magnitude, but at least you have an error bar some width. It can't be anything, right? A lot of parameters have to be positive, for example. You know that. And you usually know something. you might not think you know anything, but you do, you actually do know before you start.

So you actually know, this is the P of theta to start with. And after I've done the experiment, hopefully you're going to know more information about it. I might know that this quantity here is

going to be like a Gaussian distribution. It might have a kind of goofball dependence on theta. I should comment that. Notice how theta appears inside F. So theta's up in the exponent. It's sort of inside a Gaussian, but it's like processed by F and so the observable might have a pretty goofball dependence on this rate coefficient. So this thing could be some weird thing.

But for sure, when I change theta so this changes a lot, it's going to make a pretty big difference. Because up inside the exponent of a Gaussian, so it's going to drop off a lot somewhere. So I should get something that looks something like this maybe for my experiment. So this one is P of K initially, the prior. This one is P of yk. And what this equation says is I want to multiply those two together. And so I'm going to multiply this times this, and I'm going to get some new thing that's something like that when I multiply this time that. Is that OK?

And so that's my new numerator of this equation. Now this denominator doesn't make too much sense. This says, what's the probability that I measured the mean I measured, given nothing? So this is sort of like the prior probability that I would have measured it or something. I don't know what this is. So instead what people do, is they say, forget this. But instead, let's multiply this by a constant that's going to normalize it to make it probability density so that it integrates to one.

So that's the way Bayes' theorem is used. This is called Bayesian analysis. And so what it's telling you is how to take your experimental information as expressed in this formula and use all your previous information about the parameters, put them all together, now we have a cumulative information about everything. So we have some parameters that came into our problem into my experiment, but from previous work, I also knew something about those parameters. Now I put it all together and I get a new value of probability distribution of those parameters. And if my expert was really good, it would make this really tight [WHOOSHING SOUND]. And then when I multiply these two together, it's going to make this really sharp, and we have a really good value of k. So that's like the ideal case if I have a really great, well-designed experiment executed perfectly with great precision, then I can do this.

More generally, when I don't think about it, I get some distribution like this. I still learn something compared to what I had before, but it might not be much. So now I can end up with some distribution that's a little tighter than before. So is this OK so far? All right, now this is super simple. I didn't have to solve anything, all I had to do was multiply two distributions together.

So in some respects, this is what you should always do. All you do is you take your experiment, you multiply the probability distribution corresponds to your experiment times the prior, and you get some posterior, and that's why new information about the distribution. And if I have a distribution like this, suppose this is my new distribution here, I can still get it central value, that's my mean value, k. I can get an estimate of the range of k. So I end up with a k plus or minus dk maybe, from just looking at the plot. In fact, I never even have to evaluate what this constant is in order to do this. I can just go look at the plot, see where the peak is, figure out the width, and I can report now because in my experiment, k plus or minus dk is more precisely determined than it was before.

Now, a practical challenge with this is that theta is usually a lot of parameters. And I only drew the plot here in one dimension, but really it's a multi-dimensional plot. So really what looked like, suppose I had two parameters. I had my k I care about, and I have some other parameter, theta 2, that also it shows up in my model. And say, before I started, I knew theta 2 fell in this kind of range, and I knew k fell in this kind of range.

So really before I started, if I think what it looks like, I really had sort of a blobby rectangular contour plot, where I think it's more likely that the k value and the theta 2 value are somewhere in this range. And the most likely one is maybe somewhere in the middle there. But I really didn't know much. So it could be anywhere in this whole blob.

Now, when I do the experiment, the experimental value depends of both k and theta 2. And commonly what'll happen is that the distribution from the experiment-- need color chalk here. Let's get rid of these guys. So this is my probably distribution, there's my prior. If I do the experiment, maybe I'll have something like this. That the experiment says that the guys have to be somewhere in the contour plot like this. Because I can get pretty good fits of the data with different values of k as long as I compensate with the value theta 2.

Now I multiply these two dimensional functions. The original is a blob function, and this is a stretched out blob. And I multiply a stretched out blob times a fat blob, I get some stretched out blob that looks something like the intersection of these guys. And so I end up with some kind of blob like that. I'll draw it really thick. So this is my posterior, some kind of blob like this.

So now I know a little bit more about these two parameters than I did before I started because of my experiment. Is this OK? I really can't say I know what the real value of k is, or the value of theta 2. But I know that combinations of k and theta 2 that are sort of in this range, all of

them will give me pretty good fits to my data, and also be consistent with all the previous information I have about those parameter values. Is that all right?

Now, I drew it with two parameters. In a lot of models we have, we have five parameters, six parameters, seven parameters, nine parameters, 14 parameters. We have a lot of parameters. And so then we try to make this plot, even how to display the plot is going to be a little problematic. But it's there, right? And somehow, we still narrowed down the hypervolume in the parameter space from whatever it was to begin with to now we know something a little bit better. We have a narrower range of the parameters that would be consistent with all the information available, including my new experiment.

And then the next guy does his experiment, and he does an experiment that shows that these guys have to be somewhere in this range in order to be consistent with his experiment. And so now I can narrow down the range to be something like that. And the next person does their experiment, and they get something else, and something else, and something else. And eventually by 2050, we have a pretty nice determination of the parameter values.

So that's the advance of science, as drawn in chalk by Professor Green at the board. So this is a very important way to think about it, is what you're doing when you do experiments, is you're generally restricting the range of parameter space that's still consistent with everything. And when we mean consistent, we mean that the probability that you would have observed what you did observe is reasonably high. We'll still have to come back to quantitatively figure out what reasonably high means.

Now, when you did this before when you were kids, nobody mentioned the word Bayes, or Bayesian, or conditional probabilities, right? So they just said, oh, just do a least squares fit. How many of you did that before? So somebody told me before, forget this stuff, we're going to never even mention this stuff. We're just going to do a least squares fit.

Now, where did the least squares fit idea came from? It came from looking at this formula and saying, you know, this is the deviations between the experiment and the model prediction, and I weigh them somehow, and I have the square. And that's the thing I want to make small. If I have a high probability that what I observed really happened, or the probably I'm going to observe this, it's got to be that these guys have to be reasonably close to each other. They're really different. And it's going to be very small, because it's inside an exponential. And if those guys are really different, and the squared thing is really large, then the probability is incredibly

small that I would have observed that.

So we think that this thing should be small. And in fact, if I want to get the very best fit I can get, which means like the probability was the highest of what I observed in the real observation or something, then if I'm free to adjust one of these thetas, I can adjust the theta, try to make this thing like equal to zero, or small as I can. So that's where the concept of least squares comes from.

Now, when you're doing least squares, you almost always have multiple parameters, and therefore you're going to have to have multiple data. And they can't just be a repeat of one number. Can't be your data, it's not sufficient to determine the parameters. So normally when you do an experiment, you have to change the knobs. We have to make measurements in a couple different conditions. Like for example, kinetics. You often want the Arrhenius A factor and the EA. And so I got to run the experiment in more than one temperature or I'm never going to be able to figure that out. So I have to change the temperature in my reactor. Make some measurements at one temperature, and make some measurements at a different temperature.

And for almost everything in life that you want to measure, you're going to have to do this. You vary the concentration of your enzyme if you want to see how the enzyme kinetics depends on something. you can't just keep running exactly the same condition over and over. You'll get that number really precise, but it's not enough information to really fill out the parameters in your model. So you're going to have to run several different experiments with different knob settings.

Also, normally we don't just measure one quantity, one observable in each experiment. We usually try to measure as many things as we can. So we actually have several observables at each knob setting, and we have several knob settings, so we have quite a lot of data. And each one of those is repeated a whole bunch of times so that we're confident that we can use this Gaussian formula.

And so what we really have is the pi-th observable measured at the l-th knob position. Well, I'm sorry, l's not good either, it's used in your notes for something else. M, there you go. The m-th knob position. Now, normally you have several knobs, so that's a factor. And we have a lot of observables we can make at each position.

So this thing is a measurement. And we repeated this multiple times so I can get the average.

And we're also going to have a corresponding sigma I M, which is the variance of the mean. So it's variance, that's going to be divided by the square root of the number of repeats for that particular experiment and that particular observable.

So this is your incoming data set, and you also have your model which predicts y model, it predicts the observable i for the sequel to fi of xk theta. So if you have certain knob settings, like certain temperature, and you have your parameter values, then you can calculate what the model thinks should be the observable value, and then you can actually measure it and measure its variance. so that's the normal situation. And now you want to figure out, are there some values of the theta that make the model and the data agree? And that's the least squares fitting thing.

So what we can define as a new quantity, weight of the residual vector EJ, which is defined to be jk, to be consistent with Joe Scott's notes.

AUDIENCE: Is k the same as m?

WILLIAM GREEN: M is in oppositions, I'll tell you what k is in a second. m, sorry. Too many indices. OK, so this is the residual between the model position. And now-- oh man, I'm sorry, [INAUDIBLE]. K is an index over i and m. So k is just going to list all the data you got. Some of the data came from the same knob settings, some came from different knob settings. Yeah?

AUDIENCE: So is x the m the y model i [INAUDIBLE]?

WILLIAM GREEN: Thank you, yes. y model i, I guess this is now k. And so k is one of these indices that carry-- you can bind two indices together and put them together just like you did in your PD problems. All right. Now, I wrote down this sigma. But actually if you're measuring multiple things at the same experiment, you should expect them to be correlated. So really what we should worry about is the c, the covariance matrix, that we defined last time. So you should also compute that thing.

And so what you should expect is the probability density that we would measure any particular residuals if the model is true. And if we have these certain parameters, theta, this should be equal to 2 pi negative k over 2 the determinant of c negative 1.2 exponential of negative 1/2, epsilon transpose c inverse epsilon.

So this is the multi measurement version of the same equation here. So this is the quantity

that we think should be small if we have good parameter values and we did a good experiment. Actually, even when we did bad experiments, still should be small if we have good parameter values. And that's because the c's, if we did a bad experiment, we'll have a high variance or something, then we should see the c's will give us weightings that will reflect that. Yeah?

AUDIENCE: [INAUDIBLE]

WILLIAM GREEN: Is that--

AUDIENCE: So you have the next [INAUDIBLE] K [INAUDIBLE].

WILLIAM GREEN: Oh I'm sorry, this is the capital K, this is the number of data points. So little k is equal to 1 to capital K.

AUDIENCE: So does capital K count for both experiments?

WILLIAM GREEN: It's the number of distinct data values after you've already averaged over repeats. So you do m experiments, at each experiment you measure capital I observables. So it's like m times I, so K. If you measured everything in every experiment, it's equal to I times m.

Now there's two ways that people approach this in literature. The fancy way is you say, you know, this covariance matrix comes in in a pretty important way into this probability distribution function. And so maybe I need to worry a lot about whether I really know the covariance matrix. And my uncertainty in the mean drops pretty fast as I do averaging, but I'm not so confident that my answer in the covariance matrix was small.

So what people do sometimes is they'll try to vary both c and theta, and try to get a best fit where they're varying c. But then they have additional constraints on c that c has to satisfy the equations you gave last time about how you calculate the covariance matrix from data. And so I was saying, well, I want this c to personify these equations pretty well, but true covariance of the world of the system is not the same as what I actually measure by just measuring, say, five repeats of an experiment. And so I might want to vary the c.

You try to vary the c, turns out to be kind of complicated math, so not many people do it. Even though conceptually it makes some sense, you should worry about the fact you're not really sure about the covariance. So what a lot of people do is they say, let's just use the c that's computed from the formulas I gave you last time experimentally. So just say, let's just take c

experimental, put them in here. And now this is a constant. And now the only thing that varies in this problem is thetas which come into the epsilons. Because the epsilons depend on theta.

And so in that case, I can just try to maximize this probability. And what that happens to do is to minimize this quantity in the exponent. And so all I need to do is say, for example, theta best is equal to arg [INAUDIBLE] theta epsilon of theta c epsilon. And so this is the least squares fitting problem that you guys have probably done before. And probably what you did was you assumed I had perfectly uncorrelated data, and all my errors were the same. And so c, which is the identity matrix, and I took it out. Probably did that before? Yeah, OK. That's pretty dangerous to do, I'd say.

What people do a lot, which is a little bit less dangerous, is at least say, well, you know, when I measure the concentration of species x by GC, I have an error bar of plus or minus 5%. And when I measure the temperature with my thermocouple, I have an error bar of plus or minus 2 degrees. And so the variances of these guys should be a lot different, temperature and GC signal. And therefore I definitely need to weight my deviation somehow. And it's really what you do is you keep the diagonal entries of this. That's often done. And we just forget the fact that they might be covariant. But if you've done the experiments, you actually do have enough information to compute this thing anyway, so you might as well just use the experimental value.

So this is the least squares thing. And let's think, what the heck is this doing? We're saying, all of a sudden we grabbed all the parameters in the model, which might include things like the molecular weight of hydrogen or something. And we can find the very best values that would make our data match the experiment as best as possible. And in some sense, that's great, we know the best values and parameters for our experiment. But of course, if we vary the molecular weight of hydrogen, it's going to screw up somebody else's experiment. Because somebody else did some other experiment that depended on the molecular weight of hydrogen, and they had to get some other value to match their experiment.

So in these parameter set, anything I do to vary those parameters, I got to watch out that maybe some of those parameters are involved with somebody else's model and [INAUDIBLE] some other experiments. And I'm not really free to vary them all freely. So this is the idea from the Bayesian of having the prior intimation is so you know some of the ranges on these thetas already, and some of them you might know really sharp distributions, like the molecular weight of hydrogen. You might know that to a lot of decimal places.

And so when people do this, normally you don't vary all of the thetas. Usually what you do is you select a set of thetas that you feel free to vary because they're so uncertain, and other thetas that you think, oh, I better not touch them. Because if I adjust them, I may go to crazy values that are inconsistent with somebody else's experiment. So a lot of times like the molecular weights, you would not touch them. You would just say, I got to just stick to the recommended values and the tables. I'm not free to vary the molecular weight of hydrogen, even though if I did it would make my data match my experiment better. Makes my model and the experiment match more precisely.

So deciding which parameters to vary in this is a really crucial thing. And that's a lot of the art of doing this has to do with that issue. Also, you don't have to keep the thetas in the form you have them. You could do a transformation. So you could change to, say, W's that's equal to, say, some matrix times the thetas, and I could express the equation in terms of the W's. So I could transform my original representational parameters as some other parameters.

And often times, your experiment might be really good at determining some of these W's, even if it might be incapable of determining any of the thetas. So you often might know some linear combination of parameters, or maybe not linear combinations, some non-linear combination of parameters might actually be determinable very well from your experiment, even though you can't determine things separately.

And this gets into the idea of dimensionless numbers. So your experiment might depend on some dimensionless number very sensitively. And you can be quite confident from your external data what the value of that dimensionless number must be. But if you look inside the dimensionless number, it depends on a lot of different things. And you might not have any information about them separately. All you know is about your experiment just tells you the value of that one parameter very accurately.

So this is another big part of the art of doing the model versus data is setting up your model in terms of parameters that you really can determine, and getting out all the ones you can't determine and fixing them. So we're really going to generally change do this kind of thing. But we're going to say that some thetas are fixed, and also we might change to a different representation, change to W's instead. Yeah?

AUDIENCE: Can you explain where this transform-- I don't really know what's up with--

WILLIAM GREEN: Yeah, let's get an example. Suppose I was doing a reactor that had A equilibrium of B. And I was really interested in kf, the forward rate for A going to B. I'm a kineticist, I love to know A goes to B. However, if I setup the experiment wrong, it might be that this readction ran all the way to equilibrium. And what I see in the products is actually just the equilibrium ratio of A to B.

So what I'm measuring might be something that's dependent really on kf over kr, and that might be the quantity I can really determine. Because that's equilibrium constant. If I didn't think about it, I could just try to have the model fitting procedure, just optimize to find the very best value of kf. And in that situation, it might have a lot of trouble, because it might be quite indeterminant what the kf is, because really all that matters is the ratio.

Also I think about this some more, suppose I run at short times, and I measure the time dependence. What I'm really measuring is kf plus kr. Do you remember we did the analysis of A goes to B, one of the early homework problems? The time cost was actually kf plus kr, not kf separately. And so if I measure the exponential decay time constant, I'm really determining kf plus kr, I might be able to determine that very well. Actually, in my lab, I can do a great job for this. I have an instrument that can measure the time constant of the exponential of k really precisely, but it's determining the sum. It's not determining either one of them separately. And I might have to do a separate experiment, say a thermo-experiment to get the ratio. And then from the two I can put them together and get the two values distinctly.

So this will be an example of this would be a W. My W is kf plus kr, the matrix would be 1001, something like that. 1-1, something like that. Where I add these two guys, kf, kr. These are my two parameters, 1 plus 1. And I can determine W1 now very accurately. sorry, this is m, this is W.

So now in terms of W, this has two parameters now, W1 and W2. I can't determine W2 from my experiment, but I can determine W1 really well. So then when I do the least squares fitting, I should vary W1. I can fix it it for my experimental data, and just leave W2 fixed at some value. I can't do anything about it. That all right?

Now, do you get the difference in these two points of view? This is like, two completely different ways to look at the problem. You can think about it as, these parameters are free for me to vary, and I just have to be careful to select the ones I'm really free to vary. And that's the least squares fitting point of view. Or I could say, I'm not really determining anything in particular, all I'm doing is taking the whole range of uncertainty that we have about

parameters, and by my experiment, I narrowed it down in the Bayesian view. So it's the two different points of view.

To do this one, I need to make sure I have enough data to determine something. So I have to have enough determine some parameter, at least one, otherwise there's no point in doing this. This one I can do even if I can't determine anything, because I could still narrow down the range of parameters. But this might be harder to report in a table. Because all I have at the end is a new probably density function of multiple parameters. All right? OK, we're done. See you guys on Friday.