PROFESSOR: Probably the time has begun to begin. Welcome. I've given a guest lecture in this class for a few years running now. So I'm always happy to return. You're working on a good subject here.

I'm going to talk today about experimentation and robust design of engineering systems. I think you've already had a few lectures on design of experiments formally. So this will build on that.

The figure that I show on this first slide concerns, on the right, a method by which one can make improvements in robustness of the system by adaptively experimenting in the parameter space of that system. And the figure on the left is meant to indicate why that approach has worked rather quite a lot better than people expected. And I'll be going into that in some detail, and I hope that as I go through all this, you'll take time to ask questions. Because I think that's better for you and more interesting for me too.

So just to give you a little bit of context, this topic of adaptive experimentation and robust design is really central to my research. It's I think the core of it, but I'm interested in other things as well. Such as we do studies of systems generally and the regularities that exist in those engineered systems. And that gives rise to data that we can use more generally in methodology validation.

So we can look at different ways that people will do engineering and try to assess its effectiveness. And of course, all this is linked in with the work on adaptive experimentation, because we need a way to validate our work. But the main topic today will be what's in the middle of that slide.

That's adaptive experimentation. I think I'll start with some history and use that to explain the motivation. Why is it that we're addressing this topic? Why are we addressing it the way we're addressing it? And then, I'm going to show you the research itself on a depth of experimentation and robust design.

So I start with this quote from Sir Edward John Russell. Now, this fellow was prominent in his day. He was knighted. That's why it says Sir Edward John Russell there under his picture.

And as late as 1926, he was still expressing a view about how to do experimentation. He said that you should seek simplicity in the way that you conduct your experiments, and maybe you should ask just a few questions at a time, or maybe even just one question at a time, as you're conducting your experiments. And this was a prominent view in that they generally accepted, and you can trace it back at least as far as Francis Bacon and maybe father.

And what's interesting about this is just at the time he's still expressing his views about how it is that they conducted experiments there at Rothamsted, he hired another person, R.A. Fisher. Now, I should say that this is an interesting confluence. You have two men, very prominent, very successful, coming together in the same place at the same time. I think the reason for that was that this experimental station, Rothamsted, was addressing I think one of the great problems of their time which was consistency and efficiency of agricultural production.

If you think about the middle of the 19th century, the UK and other parts of the world experienced terrible famines. And so trying to do much better in this area was a pressing problem, probably equivalent today to the way we feel about energy supply or global warming. And now in some ways, we take for granted food supply at least in the developed world.

So here, we have this confluence and R.A. Fisher comes into the lab, and he has a very different view about how to do experimentation. He analyzes this experiment, in which we see the table three in the middle. This is from R.A. Fisher's 1921 paper, and what you see is a plot with different levels of ammonia being applied as fertilizer to the fields. And you see the influence on the yield per acre of oats and the decrement in each year. Because when you apply the fertilizer to the same field year after year, its effects diminish over time.

And apparently, Fisher was fairly discouraged by this experiment. When you read the paper, you see the details, and you see the difficulties they experienced in analyzing the data. They had some pests intruding on parts of the land, and this made it complex to draw inferences from the experiment.

And he was later quoted to say the following thing which expressed that frustration in a way. You have a bunch of scientists plan an experiment based on their knowledge of what to do. Later in, you call statisticians in to analyze the data and to get the most precise inferences you can and precise estimates, and it's really too late.

The time for statistical thinking is before you actually run the experiment. That's when you can get the greatest benefits, and what he talked about is the idea that the most finesse you could apply might improve your analysis of the data by a few percent one way or other. But if you brought in the experimenter earlier, maybe you'd get a factor of 2 or 5 in efficiency and accuracy.

And you can see how thoughtful Fisher was. He seems to be struggling hard to think there, and I think part of it is that machine that's in front of him, which is I guess what you'd use for a computer in that day. I guess it's some mechanical adding device, and I think it's nice that we have so much better machines now for this purpose. And in fact, what we find is that you can do a lot with statistics these days by knowing how to work with computers right. And in fact, a lot of the research results I'll show you today, we wouldn't be able to achieve them with without modern digital computers.

So Fisher begins to plan experiments in a different way, and this is one of the first ones in the literature, this figure one from his 1926 paper on the arrangement of field experiments. And if you've done some work already in this course on design of experiments, you might recognize this. You see somewhat indicated by darker lines that there are eight different subplots in this plot. And if I focus my attention on just say the upper left, what I see is-- and I think the people even remotely can see the arrow. Right?

You see the plot has three different indications in each plot. In this case, it's 2, an M, and early. In the next one, it's 2 and S and late. So the 2 and the 1 indicate an amount of a fertilizer that's been applied. S and M indicate the type of the fertilizer, and early and late describe the time in the season, the planting season, when it's applied.

And we see in this first block, we have four of the 2's and four of the 1's and four of the earlys and four of the lates and so on. And we even see that there is a balance, in the sense that there are two of the two M's and two of the two S's, so pair-wise, there's a balance and so on. And you may recognize that, if I just focus on 2 and M and early and late, and there are 8 of such blocks in here, we have every permutation and combination of the factors in the level. So I would say, you could describe this as a 2 to the 3 full factorial experiment within the block, which is 8.

Now, you might also say that it's a 3 raised to the 1 and 2 raised to the 2, because there are also these four plots in which there are just X's, and no fertilizers apply. You could call that a control, or you could say that this is actually a more complex factorial experiment, where in one of the levels obviates the other two. And once you decide to apply no fertilizer, it doesn't matter whether you apply type 1 or type 2 or apply early or late. And then, to further complicate the issue, we have eight plots, and within those blocks, those subplots, we have randomization in space of the different treatments.

And one of the things that Fisher talks about here is how poorly randomization actually works sometimes. So if I look down in this plot, it's on the lower right. This plot is randomized in spatial distribution of the treatments, but you see that all of the left-hand plots are early, and all of the right-hand plots are late. So sometimes, randomization gives you such problems, and in fact, these days, often people will not use randomization and instead use something like a Latin hypercube design rather than randomization.

But this was one of the early experiments that had been considered in a factorial design, and I think it's interesting that Fischer decided to call it a complex experiment. It's as if he was trying to emphasize as strongly as possible his difference of opinion with Russell, his boss, who had said that you should seek simplicity in your experiments. And he was saying, no, I'm going to do the exact opposite. I'm going to make my experiment as complex as I can.

And here's the motivation for doing that, and this I think is one of the great results of Fisher, is that he was able to demonstrate the specific value that you could attain by making the experiment complex. And what we have here on the upper right of the slide is a cuboidal representation of the 2 to the 3 factorial design. This may not be the first time you've seen it in this course, but you have the three different levels labeled as A, B, and C, and the factors labeled as A, B, and C and the levels indicate as pluses and minuses.

And the way that you would estimate the effect of A is by looking at the observations, where A is at level plus, taking the average of those four. And then taking all of the observations where A is at minus, taking the average of those four and looking at that difference, and that's the equation here. The effect if A is defined as this equation which just implements what I just described in words.

And then what Fisher is able to show is that the standard deviation of the effect estimate has the following form. It is a function of the experimental error, if I model the random variations in the observation due to, let's say, soil pH or the pests that I spoke of earlier or uneven distribution of water on the fields. If all of those nuisance factors that I didn't control influence my yields at the various locations, and those contribute to sigma sub-epsilon, those will be reflected in sigma sub A, my estimate of the influence of, say, applying the fertilizer early or late.

But they are reflected in sigma sub A substantially less if I use this factorial design rather than a single-factor experiment. And he could show that for the case of the 2 to the 3 full factorial, the improvement in efficiency is a factor of 2. So it's a large benefit as compared to the benefits that you could derive by a different kind of analysis of the data. The planning of the data made a huge difference.

But the problem is that, if you're using a full factorial experiment, you see 2 to the 3 or 2 to the k, the scaling with the number of factors is terrible. Right? And so you find that you can't run such experiments with more than, say, seven factors. That's about the largest that I've seen published in tables, so that we can analyze them, because then you have 128 data elements to include in the appendix of your paper. So they're rarely run with more than seven factors. And yet in an experimentation, especially in engineering, we often have many more than seven factors we're concerned with. It's not at all uncommon.

And so almost immediately, Fisher begins to consider the possibility of some means to reduce the size of the experiment, even as the number of factors goes up. And this is the way he describes thinking about that. You deliberately sacrifice possibility of obtaining information on some points. So you decide there are certain things that you would be able to estimate in the full factorial, but you just decide ahead of time that they're not worth investigating, that maybe they're unlikely a priori.

And so you design the experiment so that you can't possibly obtain information about them. But actually, it's a little worse than that, I'm going to show you later. Not only do you not have a possibility of obtaining information about those points, those specific points, if they were to be consequential in the end, would actually interfere considerably with your ability to make inferences.

And so this is the simplest fractional factorial I can show you. Again, we have the cuboidal representation, and what we do is just to make observations at half of the vertices. So I indicate those by the black plus's, and these four are selected in a specific arrangement to give us the 2 to the 3 minus 1 half fraction with resolution 3. I'll describe what resolution 3 means in a second.

One of the things that's interesting about the experiment in my mind is its projective properties. So we see that, if we collapse factor A-- that is, if we imagine that there is a flashlight or such, a source of light off to the right, and we get a projection of the matrix of the cube, and now we just get a square. We'd see that if factor A is collapsed, we would have a full factorial experiment in B and C. Have you studied these projective properties in the other parts of the course? OK. And so it goes in the other dimensions as well.

And so in my mind, what's interesting about this is that in a way the design of the experiment allows you to account for a particular kind of uncertainty. If you have a long list of factors, and you're not sure which ones are going to be the important ones, you know that some few will turn out to be important, but you're not sure which ones. You can arrange it so that you will have done a full factorial in the few important factors without knowing exactly which ones you need to do a full factorial experiment in. And that's what the projective properties allow you to do.

But all this comes at a cost, and the cost is more easily described by looking at a larger fractional factorial experiment. This is the 2 to the 7 minus 4 design, and what we see here is that, if we have 7 factors, and we want to stuff them into a small experiment with just say 8 trials, then what we can do is put them in this arrangement. And the risk is that if, for example, there is a two-factor interaction between F and G, then that will emerge in this pattern. That in the first four experiments, F and G are all at the same level, either plus and plus or minus and minus.

And in the last four experiments, they're all at different levels, such as minus and plus or plus and minus. And so if the two are interacting in some important way, the effect will emerge in such a way that it is coming into being exactly when A is at minus, or A is at plus, and so you will confound the two in your analysis. And if you think that an effect of A itself is more likely than a two-factor interaction of F and G, you're likely to make a mistake, to assign that effect to A, whereas, it actually occurred due to F and G. That's the risk that you have to take to make a relatively small experiment that's investigating a lot of potential factors. So these are the trade-offs we make in experimental design.

Now, one of the reasons that this whole topic is important to engineers is that we're concerned with robustness, robust parameter design. And what I'm showing here is the cover of one of the I think probably the best modern compilation of this topic on experimental design and its application to engineering systems, in particular for the pursuit of robustness. And Jeff Wu and Michael [? Hamouda ?] define robust parameter design here as a set of statistical and engineering methodology to reduce performance variation by choosing settings of control factors to make the system less sensitive. So you're trying to find things that you can control in an experiment in the design and to use those to make performance less sensitive to things that you don't control.

Can you think of examples relevant to your work of something you might make robust? Or even simpler, just of a noise factor you might be concerned with, something that influences a system. Well, we'll have more examples later, and in fact, I believe that in the past, this course has had some projects associated with it. And in some cases, people want to do robust parameter design and projects. Yes?

AUDIENCE: [INAUDIBLE]

PROFESSOR: OK. OK. So you might be doing injection molding or a process like that, and you could control temperature and humidity but at some great cost I guess. So you might choose to allow those things to vary to some degree. And if they influence say the geometry of your parts, resulting parts, too much, you would probably be producing less value to your customer. Probably the function of your articles themselves might be degraded by that, and so you might like to choose process parameters, such as pressures and compositions of your plastic and so on, so that that effect is minimized. So you decide not to control the humidity but just to be insensitive to it. There. Good.

So on the cover of this textbook, you see actually something called a crossarray it's hard to pick out, unless you go and zoom in on your own computer. But you see capital A, B, and C over here, and those represent control factors, such as pressure and speed of the extruder and the composition of the plastic, things that you might control nominally. And then there's little a and b, those are temperature and humidity and things you might decide you're not going to control so much.

And you see that fractional factorial designs or factorial designs are being used in each of these arrangements shown to the left and above, and then the crossing, the product of those two designs, is indicated in the middle. And here, Wu actually begins to hint and foreshadow about the idea of what he calls a combiner experiment. So he shows some of these being white and some being dark, and he's suggesting that as a way to address the rather large size of this experiment.

So it turns out that the first person to pioneer this idea of bringing in fractional factorial designs into the plant and using them explicitly for the purpose of robustness improvements, person who seems to most often get credit for this, is Taguchi, in Japan. So he begins to do work, in Japan, after World War II, when their manufacturing base is in disarray, and their economy is in not good shape, and their reputation for manufacturing is very poor at that time too. So if you bought a product, and it said made in Japan on it, people would actually imagine that that meant it was of rather poor quality. And of course, our impression is totally the opposite now, and it's interesting to consider how that happened. It was a lot of different things.

But maybe one of the things that brought that change about in our perception of Japanese manufacture was some of this methodology. And Taguchi correctly pointed out that robustness could be attained, and it could be attained through experimentation. And he suggested that you might, for example, run an experiment like this, an orthogonal array or factorial design and control factors crossed with noise factors.

And in this case, you take your 8-run array of control factors, cross it with your 4-run array of noise factors, and you would have a 32-run crossarray. And so you see with this cross symbol, the implication is that the size of the experiment scales at least like the product of the number of control factors plus 1 and the number of noise factors plus 1. Which when you begin to consider lots of control of noise factors, it gets awfully burdensome. So people are looking for ways to do it more efficiently, and that leads to the cover of this textbook, the idea of just doing half of them, so that you're actually not running exactly the same noise conditions at every setting of the control factors.

And what we've done in our methodology validation work actually is to show in a recent publication in the *Journal of Quality Technology* that this procedure, although theoretically looks promising in many specific cases, it's failing. And yet, if I look at industry practice today, I think the methods of Taguchi and also the refinements proposed more recently on the basis of statistical theory are finding their way into industry practice and are really quite solidly entrenched.

So some of you may have had some experience with these Six Sigma programs by now. Maybe some of you are in General Electric, maybe some Motorola, and I guess many other companies by now have such programs. And a big part of such programs is that entering engineers will take some coursework, even if they didn't get it in their undergraduate degrees. The company feels people broadly need this, and so they get some coursework, and then they do projects. Right? And they demonstrate that they can do some of this kind of work for the company, and at Ford, at least the last time I checked, their design process encoded just this sort of thing, and they have their multi-step process for attaining reliability and robustness.

And in step four, I dig in and look at the details, and the engineers are asked to select appropriate orthogonal arrays. So the idea is that maybe they have some choices about what size of orthogonal array, number of levels, the crossed array versus the combined array, such as Wu talked about. But a departure from orthogonal arrays itself is not really being considered anymore in industry, and so this has become very much entrenched. And in fact, in some of the textbooks, you see an even more extreme statement of the point. In this text, they argue that the idea that Edward John Russell expressed on that earlier slide, that simplicity should be valued, and that single-factor experiments are often advisable, that has seen its final demise. So that's the opinion voiced in this particular textbook.

But we wanted to question this a little bit. We wondered if sometimes there isn't a role for the simpler experiments still, and this decision to spend some time looking into this came out of some observations in industry. And I thought in some ways now, it seems a shock to me that our initial thoughts along these lines came out of the farm again. So that all these ideas of orthogonal arrays and so on came from agriculture in the first place, and now our wanting to question them came off of the farm, because we were working with a manufacturer of agricultural equipment.

And it seems to me actually that farm equipment is one of the most significant robustness challenges in the world. Because what you have in a tractor or a windrower or a combine is really a factory. It needs to process something, sometimes in not a very simple way. Sometimes, it needs to take a piece of cotton off a plant and then strip away some part of it. It's not a simple thing.

And yet here it is, and it's out in the weather. It's out in a field. Whatever conditions prevail, it has to deal with. So it's a very difficult reliability challenge.

And this particular company had relatively recently brought in a lot of robust design methodology, and they were looking forward to the benefits that would accrue. And they got exactly the opposite, that this most recent launch of a new tractor was one of the worst they'd ever had, and it was causing them real problems. Their customers were quite angry with them.

And so we wanted to look into that, and we decided to, in effect, do an audit, look at their design process for this tractor. What had happened? Were there some mistakes in choices of design or priority or management, whatever?

So one of the things that we asked for is an accounting of all the different robust design experiments they had done, and they gave us a list of experiments they planned. It was a long list, and then we said, well, yeah, but we need more detail in this. Show us the written reports from all these experiments. Show us the data. Show us the decisions made basis of the data.

And they gave us a much smaller list of reports, and we said, OK, you had 30, and now we have 5. What about the other 25? And they said, well, those were never finished, and that was interesting to us. Why is it that the majority weren't finished? And we began to ask questions of the individuals involved, and they always have pretty good reasons.

In some cases, they would say that a piece of test equipment had broken down at some point along the way and was going to take a certain amount of time to repair, so we had to move forward with a partial data set. In fact, a large fraction of the experiments aren't finished, and this prompted us to wonder, what would you do if you knew this from the outset? Let's say that you put a 60% probability on canceling any individual experiment you might run somewhere at a state of partial completion. Let's say that were the case, but you didn't know which 60% we're going to be canceled, because you probably don't know that.

What would you do? Well, we thought, perhaps, you would do a one at a time experiment, because then at least, no matter what you did, you would have stopped, and the data would be relatively simple to analyze, such as what Russell talked about. And we found that actually some of the most prominent statisticians in the past always felt that there was a role for one factor at a time experiments. So for example, you may have heard of Milton Friedman in a different context. What do you know him for?

AUDIENCE: [INAUDIBLE]

PROFESSOR: Yes. Yeah. Nobel Prize in economics, market view of economics and Leonard Savage who had written an excellent text on foundations of statistics. They found that an efficient design for the present purposes-- that was for maximization-- ought to adjust experimental program at each stage in light of the prior stages, and actually, what they described was a one factor at a time experiment. And then Cuthbert Daniel talked about a more social or psychological value of simpler experiments, just reemphasizing the points that Edward John Russell had made. That somehow you feel that you can react to data more rapidly or learn something from each rung.

And he talked about some criteria by which you might make that choice, that for example, he had a demarcation. Maybe it's OK to do it, as long as the effects are at least three or four times the average random error per trail. So he had, based on his experience, come up with a demarcation-- effects three or four times random error.

And we wanted to understand whether that demarcation was about right, so we began to study an adaptive variant of one factor of time experimentation. And in the beginning, we were interested primarily in knowing, if you were to do this adaptive experimentation, what level of probability of canceling the experiments would make this style look good? And the funny thing about it is we started with some maybe 50% probability of canceling, and we looked at the two experimentation styles, adaptive and factorial, and on the cases we were looking at, adaptive looked better. And so we dialed down the probability of canceling the 10%, and they still looked better. And we dialed it down to 0%, and the adaptive experiment still looked better.

And we thought, well, now we know we've made a mistake, because we saw it in the textbooks that this kind of experimentation has seen its final demise. So either we made a mistake in the way we set up the study, or we chose a strange case to start with. We were perplexed.

So we did eventually a big empirical evaluation, and we took 66 different systems of a broad range of kinds-- chemical systems, mechanical systems, electrical systems, and we made our comparison in this case without canceling. Now, we were interested in that issue. Imagine that you're going to do this adaptive, one factor at a time experimentation. You're going to compare that with a fractional factorial experiment and see which one does better on average. And this is what we found is that, as long as experimental error was about 25% of combined factor effect, or if interactions were large enough, then you'd prefer the adaptive experiments.

And so here is that same result shown in a tabular form. So we have on the column headings the degree of experimental error expressed as a fraction of the factor effects, and we have on the row headings the strength of interactions-- mild, moderate, strong, dominant. Strong interactions would indicate that the interactions are counting for more than a quarter of effects in the system, quite a big fraction, and moderate is maybe 10% to 25%.

And the gray boxes indicate that the adaptive one factor at a time experiments were giving at least as good or better results than the fractional factorial, and so you might prefer them, especially given their flexibility. And so we were surprised at how much of that territory turns out to be gray, and that was the main result we wanted to get across in this paper. And we thought, a strength of the experimental error as large as 0.4 looks like a lot of experimental error to us, to engineers. When you see a dependence, such as this blue line here, on the figure and the scatter being indicated by the red, that's about typical of the strength of experimental error at 0.4 that we expressed, and to us that looks like a pretty noisy experiment. And so we thought it was interesting how much of the time you might recommend these experiments.

So we wanted to understand how this could be, since it seems so counterintuitive to us. And so we delved into it a little more in terms of the mathematics, and we tried to develop a theory that would account for this. So we started to write a model of the procedure of adaptive one factor at a time experimentation, and so there's just a little notation here. You imagine that you make an initial observation.

Some of the symbols are coming up a little wrong. This is an x with a tilde over it is an initial starting point level for factor one. And x2 with a tilde is the same, and that k should be an ellipsis, three dots.

And then you change one of the factors. You toggle it, change it to the opposite of the starting value, and that's your first observation. And then you make a decision about the level of the first factor based on the sign of the differences in the observations.

So if the improvement seems to obtain, we adopt the change, and then we go through that for every one of the different factors, putting stars on our x's, and we continue. And what we're interested in knowing is how does the expected value of improvement behave as a function of a number of parameters in a model? And the parameters we put in our model are few. We talked about the size of main effects in general, the size of interactions, and the size of experimental error, because our empirical investigation had shown that those were interesting to look at. And here, we show we wanted to normalize those with respect to the maximum, the best that you could do, so that our results could be easily interpreted.

And another concept we introduced in this paper, we felt we needed to explain the results, was the idea of exploiting an effect. So we had to introduce the idea that, when you do an experiment, if your system has a number of effects in it, main effects or interactions, they can either be exploited or not, and we'll say that they're exploited, just in case. For example, if you're trying to get a higher response out of the system-- let's say yield out of a field [? of oats, ?] or quality of a manufactured article. Then, you exploit an effect, if the coefficient in the model times the level of the factor is such that it's contributing positively to your outcome.

And you exploit an interaction in the system basically under the same condition. That if the two factors involved are such that you're gaining benefit from the coefficient in the model, either increasing or decreasing the response depending on what you prefer, we call that one exploited. So we wanted to understand the probabilities and conditional probabilities associated with this adaptive experiment.

So we begin to do the theorem proving based on our model, and we find, first of all, that if you change just the first factor and compute the expected value of improvement, you have two different kinds of contributions-- 1 due to a main effect and n minus 1 due to interactions. So you change factor 1, and this figure here is meant to indicate the condition with n equals 7. You've got 7 factors you're interested in, and this stack of squares indicates the contribution of main effects.

And none of these other factors are contributing anything, but now the main effect of the first factor is contributing quite a bit on average, after you've done the adaptation. And all of the 6 potential two-factor interactions are on average contributing a little to improvement, because it may be that in some cases there'll be a large two-factor interaction. And you'll see it when you toggle the factors, and you'll go chase that, and so on average, it's contributing something.

And we could write closed form expressions for the contribution, and the contribution due to main effects is larger when the main effects are larger. It tends to be smaller when interactions are large and when error is large. The contributions to the interactions are large when interactions are large in the model.

And interesting to us, if you write out the expected value, normalize by the maximum, so that the improvements are on this scale either 0 or 1 or scaled in between. The contribution you would get in the case that experimental error is low-- so here I put experimental error at 10% of a main effect. The contribution you get after running one 1 of 7 factors is actually about 1/7. So you get 1/7 of the way toward the maximum by changing 1/7 of the factors. That's a good start.

And the influence of error is-- it's there. As you go to error being as large as main effects, it drops a little, and of course, if you make the error really huge, it drops quite substantially. But that's a hard test to pass, but we're just trying to understand what the influences are.

OK. Now, let's talk about probability of exploiting the first main effect. Just when interactions are small, the probability of getting the benefit of the main effect is large. Then, as interactions increase, it's less likely that you'll get the benefit of the main effect.

Now, it becomes interesting at the second step. Now, you've toggled two factors, and then what we think about it is this. Now, you've got the benefit of two main effects, and you get first n minus 1 and then n minus 2 interactions potentially contributing.

And at this point, I hash it in, because this is the first interaction which you've actually locked in. In the sense that any further changes you make through the experimentation process will not reverse whatever benefits you got. Because you've already changed one and changed two, and you've made your decision about that.

And now, we see that there's a contribution due to the interaction, and the interaction is the same functional form as I showed before. That the contribution is large when the interaction is large. Again, we see that the expected value has gone from about 1/7 to about 2/7. So we're continuing to step forward at a good rate, and we want to understand why.

Well, OK. We see the probability of exploiting the interaction. We can write it's closed form expression. It's strictly better than 50-50. So it's in any case better than random, and it goes up with interaction strength.

Now, the interesting thing about it, it's never very high, because actually, any particular interaction is competing with n choose 2 other interactions or n choose 2 minus 1. And so any particular one is unlikely to be exploited by this procedure, and so maybe we get a 60% chance. But if you condition that probability, if you say, well, let's say that this particular interaction that we just locked in is somehow the largest one in the system. Let's say that that happened. We conditioned the probability, we get a new expression, and the probability could be as high as 80%.

The reason that's interesting to us is that we know that the adaptive one-factor experiments are just not capable of allowing you to estimate interactions. You take a factor, you change it, you look at the influence, and what you're seeing is a combination of the main effective the factor and a bunch of conditional contributions due to a bunch of different interactions. And so you can't sort out which one is which. You would need more experiments to do that. Nevertheless, if there's a big effect there, there's a good chance you'll get benefit from it, when you do this adaptive experiment.

AUDIENCE: [INAUDIBLE]

PROFESSOR: Yeah. It's an interesting question, and this is something that we were concerned with. All the results I'm showing you today assume that you've randomized the ordering of the factors and that you've randomized the starting point. And so far, what our investigations show is that the ordering of the factors doesn't matter much but the starting points do.

So I would say, whether or not you order the factors from the most important one to the next to the next, it's a minor improvement, not big. But if you select a very promising starting point, that is go with something that is likely to be very robust, high performance, that makes a big difference. And it wasn't obvious to us at the start. I thought probably what you said was true. It turned out to be the other way around.

So the dynamics of this process of experimentation are such that it turns out the probability of exploiting interactions as you go gets higher and higher throughout the course of the experiment. So it's true that, if you knew where the interactions lie, you'd like to toggle them for the last time as late as possible. But I guess I don't really believe that people would be able to identify which two interactions are going to be involved, so I ruled that out and didn't study it much.

But the point is that, even if you didn't know where they lie, the benefits continue to accrue almost linearly. Even though you're exploiting fewer and fewer new interactions as you go, you see the white ones left to benefit are diminishing over time. But because their likelihood of coming in your favor is increasing, you go up almost linearly, and you get almost to the end, 80%. You can almost to one, well, pretty close,

And what's interesting to us about this is take the case of 7 factors. The adaptive one factor at a time, assuming two levels, that takes 8 experiments. The full factorial takes 128. So you've looked at a very small fraction of the overall space, and you've gotten 80% of the way toward the best possible outcome in the space. That to us is interesting.

And then if I want to compare the use of resolution 3 designs and adaptive designs, we have expressions for each. And then to make the comparison straight, one against the other, we do the following. Just hold them up against one another, look at-- we see the experimental error is plotted parametrically. So let's look at just the crossing point when experimental error is about the size of main effect.

They cross at the point where interactions are about a quarter of main effect. The reason we think that one is interesting is that I alluded to earlier the idea that we are doing these studies, big studies of complex systems and regularities. And one of the things we found is that interactions being about a quarter of the size of main effects, that's pretty typical.

That's a typical amount. So if you go into a factory, and you get your engineers to list a number of factors they're interested in studying. And then you run a big experiment and study some of the interactions too, you will generally find they're about a quarter of the size of the main effects. It's just a reasonably reliable regularity, and so we think that's typical.

And the way I would interpret the graph is that, if your system is pretty typical in this regard, strength of interactions, you should be doing the adaptive experiments whenever error is less than main effects. So Cuthbert Daniels criterion for demarcation was error should be about a third or a quarter of main effects, and we're saying now, it's actually more like the size of main effects, not a third or a quarter. That's assuming, importantly, that your main goal is just getting improvement, knowing exactly where they came from or estimating effects accurately.

Now, our primary reason for going into all this in the first place was to deal with robust design in the end. And so we attempted lots of different ways of combining the noise factors with the control factors, and so far, the simplest thing turned out to be the best. And that is what you see here in this cuboidal representation are controllable factors A, B, and C and our noise factors, little a, b, and c, are being crossed in effect.

We take the adaptive procedure. We run a factorial design in the noise. Then, change one of the control factors, just one, and run again the exact same design. And then what we do is we select our level, in this case of A, on the basis of our preference for this set of observations versus this set.

If we like this set because, for example, it has lower variation, then we select it. We select A as positive, and then we just continue to wind our way through the space, reversing changes that we don't like and adopting ones that we do. And we study this, this says-- it actually appeared in *ASME* journal. I've got to change that.

So we began to validate this approach by applying it to a number of different systems. One that might be of interest to you is sheet metal spinning. So in this process, you go into your factory, you take sheet metal articles, and you turn them into manufactured goods, surfaces of revolution, cups, and bells and whatever you need. By taking the circular blank, pressing it up against a mandrel repetitively, until you get the deformation that you want.

And you might be interested, for example, in consistency of the geometry, and you might be interested in how a number of different parameters affect that consistency of the geometry. Such as what material you chose, the shape of the mandrel, the shape of the path which you use, the number of times and the force you apply. And this was studied by a group in Germany, and we used their results to simulate the process of running these adaptive experiments and running an alternative across the ray design on these systems. And we found that the Taguchi-style crossed array made indeed improvements in signal to noise ratio of the system, but our adaptive procedures were doing a little better, until you got to some crossing point, an experimental error up here of 2 millimeters squared of this quality measure.

And this is where we found another result, the one that I alluded to earlier. That if you gave the engineers the benefit of the doubt and suggested that they didn't choose their starting points at random, but instead choose informed starting points. And let me define what I mean by that.

Let's say, you have a number of factors. They're all at two levels, a random starting point is just flipping coins. You choose each one.

But let's say that somehow the engineers are able to do better than 50% odds of getting the best setting for each factor, but instead have 75% chance of defining the superior level of each factor. In that case in the end they would get this result, substantially higher than either of the other two alternatives, and they'll never cross, no matter how much error there is. Yes?

AUDIENCE: Why is it saying the smaller the better on the [INAUDIBLE]

PROFESSOR: Yeah. So smaller the better is the parameter of interest was meant to be smaller is better. But then, when you compute a signal to noise ratio, larger it's better. So the underlying parameter, called A20 in this case, smaller was better. But in Taguchi's methodology, you then take that parameter and its variation and transform it to get a signal to noise ratio, and with signal to noise ratio, larger is always better. So I appreciate why that's confusing.

This one I included in the paper, and I include it for you. Because in some cases, you'll feel skeptical about these results, because they still are pretty counterintuitive. And you might need to demonstrate them to yourself, and the paper airplane allows you to do that.

So Steve [? Effinger ?] came up with this template for demonstrating some of the methods here in robust design and design of experiments. And the template allows you to fold a variety of different paper airplanes, in this case, it's 3 to the 4th, or 81 different paper airplanes. And we ran an experiment in which we had created different noise conditions for the flights, and you could replicate our results. Actually throw these airplanes and do it adaptively one factor at a time, and I think you'll find the same results as we did.

That if you do the L9 Taguchi array, or else that's a [? Plackit-Berman ?] design by another name, cross with a 2 to the 3 minus 1 noise array, you get these kind of improvements in signal to noise ratio. But if you run it adaptively with random starting points, you'll get better results. And if you use an informed starting point, such as looking at a plane that has about the right aspect ratio, and maybe you know right ahead that folding up the little winglets out here is probably better for lateral stability. If you make such judgments, you'll get even better results, and so that's what the paper airplane study tells us.

So we did four such studies, and here's a point that I find interesting. OK? First of all, if you're to lower error states, what you find is that, on average, the random approach, starting at random starting points, works better than the fractional factorial. Informed starting points are better than that. If you raise the error high enough, things turn around, and the factorial design is a little better than the adaptive approach, unless we use the informed starting point.

But what I think is further interesting is that if you look at the interquartile-- if you look at the range of results across cases, the adaptive procedure even at high error, to me that's a preferable range, 51 to 87 versus 36 to 88. I think you actually take less risk from case to case with the adaptive method.

The factorial design is optimally suited to minimize the influence of experimental error on your outcomes. But some of the uncertainties you face in experimentation aren't that kind. They're not experimental error. They're actually uncertainties about where the interactions lie, and in fact, fractional factorial designs are among the most sensitive designs to that particular uncertainty. Because when you make this confounding between main effects and interactions, it's actually very harmful to your outcomes.

So those are the principle results that I meant to show you today. There are a couple other things that have emerged more recently that I can talk about. For example, recently, we've been interested in the idea of producing ensembles.

Now, here, the situation is that you might find that you have an adequate budget to run a larger fractional factorial design. Let's say instead of a 2 to the 7 minus 4, you actually have a budget to run a much larger experiment, 2 to the 7 minus 2. So the counterargument from some folks in industry was that they rarely ran resolution 3 designs. They thought that that was too big a risk anyway.

They were tending to run larger experiments, such as 2 to the 7 minus 2 higher resolution fractional factorial designs. And they said, so because of that, I don't think we should do the adaptive OFAT, and we were interested in that condition, that issue. It turns out that, if you take the same budget that it takes to do the 2 to the 7 minus 2, you would be able to run four OFATs.

Now, if you do that, if you choose four different starting points and four different orderings of the factors and run four OFATs, and then take those results and make an ensemble of them. And for example, a simple thing you could do is just pick the best of the four. Again, the ensemble would produce a better result than the fractional factorial design. And then the interesting thing we find about it is, remember before, when you increase experimental error, you expect the OFAT to eventually degrade in performance and to cross with the fractional factorial. So now, the issue is how much experimental error can I endure before the adaptive experiment is no longer recommended?

With the ensemble method, it's actually the opposite way. As you increase experimental error, eventually, the distance between the ensemble and the fractional factorial increases. So the ensemble method by itself is providing a robustness to experimental error. And so we've been able to address another one of the counterpoints to the adaptive experimentation idea. That in those cases where experimental budgets are higher, you might still want to run these.

The other benefit of the ensembles we've been able to show in another paper. I didn't put some slides in, but I want to tell you a little bit about it. Unless you have a question. Maybe I'll take your question.

AUDIENCE: OK. So what are some practical issues that you face influencing one factor at a time. For example, I can think in my mind, say you can't do the measurement immediately after each experiment. Right?

PROFESSOR: That's right.

AUDIENCE: In that case, you would probably still have to go to some fractional factorial design. Are there other issues that-- like practically, because this is incredible. It seems like it could save a lot of people a lot of time.

PROFESSOR: Yeah. So you make a good point. For example, we know in agriculture, you want to run 64 different treatment conditions, and it takes a season to get your results. So you want to run them in parallel. It's important to run experiments in parallel in agriculture, and in agricultural equipment, it's exactly the opposite.

So what happens is you're trying to develop the next generation of tractor. And what they do is they take the last generation of tractor that they built, and they use that as what they call a mule. In the automotive industry, they use the same term.

You're making the next Taurus. You take the 2007 Taurus, and you start using it, putting additions on it-- new fuel injection, new valve timing, and so on. You use that as a mule, and when you're using articles like that in that style, and you have a limited number of them, actually, your experiments are necessarily sequential. Whereas, in agriculture, they're necessarily parallel. So I actually agree with you, when you're in an engineering scenario, and your experiments are necessarily parallel.

It sometimes happens, in say lithography. You're going to etch it. You're going to have to etch all these things all at the same time in a batch. Yeah. That's a reason to do it the other way, but I think more often than not, we're doing our experiments necessarily sequentially, or maybe it's OK to do it sequentially.

Another factor, since you asked, is the possibility of time trends, and indeed, when you do adaptive experimentation, you are necessarily making restrictions on randomization. Right? So let's say that you're measuring apparatus is drifting over time. Let's say it's drifting toward making all your results look worse. That would give you somewhat of a bias toward your starting points, as opposed to all the leader changes.

As best we can tell, actually, the adaptive procedure is not so hypersensitive to those time trends. They reflect themselves to some degree as if they were additional experimental error. But if you randomize the starting points anyway and the order of factors, I don't think it's worse than experimental error. Because you're not interested in factor effects in the end anyway.

This is probably the biggest concern I think with the method as I'm describing it. The biggest issue is you're making an explicit trade-off to say, what I'm interested in doing in this case is getting some improvement. And I'm willing to make a change in my experimental procedure that will give me somewhat less knowledge about why I got the improvement.

Now, if you're at Ford, and you've got three months to product release, and you need to meet your emission standards. The goal is to reduce carbon monoxide in those three months, and you think the learning that you're going to gain about this particular engine and its configuration is not likely to be reusable on the next one. Then, in that case, I think the adaptive method makes sense.

If instead you think the primary value of your experiment is archival. Right? You're doing the experiment, you're going to take the results, you're going to make them available to all Ford engineers, and you're going to benefit on the next model and the next model. Then, yeah, you probably ought to arrange the experiments, so that you get the highest precision of effect estimation, the highest validity of inference. That's factorial design of experiments.

But we had an interesting learning along the way and that is to do our big meta studies, where we're trying to decide how big are interactions? When do they occur? We went into Ford's databases. They had run lots of experiments in the past, and they had them in some company database. And we were drawing these data sets out to put them into one of our studies.

And very frequently, we'd go through the table, and there was some ambiguity in our minds. Maybe there was a plus in the experimental matrix, where we thought there should have been a minus, or there's some question. And in those cases, we would go track the person down and ask them what they meant. And in almost every case, that person would say, huh, you're the first person who ever called me about my experiment.

They'd put it in the database, and no one was using it. We were the first ones to go back in and use it. So we are not so sanguine about the prospects of people reusing experimental data, at least so far as industrial, day-to-day experimentation is concerned. Most often, those experiments serve their purposes at the time, and then they're gone, for practical purposes.

AUDIENCE: On that note, I was working for a startup company, and we would use a facility that we knew other larger corporations would use. And we had a much smaller budget, and I wouldn't to say that my mind [INAUDIBLE] we didn't go in with a full factorial design, because but we couldn't, but we had certain objectives, certain targets. And so we go in and say, all right, look at the results vary and then change the parameters. I think your method has to be more educated than that, but it almost seems like it's a similar mentality.

PROFESSOR: It is. Yeah, and in fact, I never formally tried to study this. But if you back off a little bit from our structure and apply some different structure, one that suits your circumstances, it's probably still pretty good. It might even be better.

The bigger point is that you're doing adaptation as you go. And I have been asked the same question, in fact, Jeff Wu asked me at the last time I presented a conference. He asked, how much of this result is due to adaptation, and how much is due to the one factor at a time part? Because this is both. Right? And it's a hard question to answer.

My sense is it's mostly adaptation. On the other hand, one factor at a time in some sense optimizes adaptation. You can be adaptive all the time with every new observation as much as possible. So it's not as if I can cleanly decomposed the two contributions to the result, but my sense is that it's mostly adaptation. And if you're doing some other adaptive procedure, especially if it's informed by some prior knowledge or topical area knowledge, it's probably a good approach.

So I'll just tell you about this one last thing. OK? We did an investigation about another kind of uncertainty in large systems. So let's say that you're developing the next tractor or the next car, and you think about applying some robust design to the system.

Now, for any reasonable scale product, there are literally thousands of things which you could potentially apply a robust design experiment to. You can run one on fuel injection. You can run one on braking and one on environmental controls and one on electrical, just keep going, and you're not going to do them all. In fact, I would say less than 5% of all opportunities to do robustness improvement work are actually executed.

And then the question I always ask myself is I wonder if-- let's say, it's 5%-- I wonder if companies are doing the right 5%. I wonder if they know where their biggest quality problems are. Now, we know for a fact that, if you look at retrospectively where there have been big quality problems, if I look at Ford Explorer and the rollover and so on. Right?

I know they didn't do a robust design study on tire inflation and its delamination and rollover. They didn't do it. Now, why didn't they know that that was the one to do robust design on? Well, they just didn't. It was not so likely after all. It wasn't happening to other SUVs.

So I ask myself, if you could run a relatively expensive robust design study on 5%, let's say that's the case. You have the budget for that. What if you did something a tenth as expensive and did on 50% of the system?

So we would call this idea streamlining. We know how to take any robust design experiment people are proposing and to do something a little bit loose but to do it at a tenth of the cost, and you'll get about 80% of the same benefit. The trade-off works like that. It's a good old Pareto 80-20 sort of trade-off.

And we ask ourselves how that trade-off in doing relatively many relatively less perfect experiments compares to doing a few very good ones. And we ask ourselves how this trade-off is affected by the likelihood of identifying the one or two biggest quality problems that you face. Because that rollover thing cost Ford a lot. You would pay a lot of money to avoid that one.

And so the big conclusion of our study is you would have to have a pretty high probability of knowing exactly where your problems were in order to do it the way they do it now, which is relatively small number of expensive studies. You'd have to know like 95% certainty where your biggest quality problem is. If you are at least a little uncertain, if you think you're only 70% sure where your problems lie, scale them all back by a factor of 10, and do 10 times as many. You'll be much better off. That's my view, and we didn't have to make so many assumptions to come to this conclusion.

People's uncertainty about where their biggest issues lie are bigger than they think, and therefore, they need to democratize these kind of processes. It's not that you have to do them very, very well in a few cases. What you need to do is make sure every engineer knows how to do this kind of thing, and that they're doing it to almost everything. At least almost everything that, for example, is not proven in the field already.

You don't need to do it on the alternator, because you used that same alternator on the last three models, and it's fine. But for anything that's relatively new, you need to be doing some of this robustness refinement work. And even if you do just a few experiments, exposing your systems to harsh conditions, making changes, making improvements on the basis of what that reveals, that's the key thing, not so much the finesse you apply.

Anyway, that's the conclusion we came to so far due to this research. So our main conclusions are that we can show through empirical work and through theorems that this adaptive procedure gives benefits. You get a long way toward your results that are desired, especially when interactions are not so small. And that you can cross these adaptive experiments with fractional factorial designs to use them for robustness, and it seems to work pretty well. And now, I think I'm pretty much out of time.

Now, if you all have ideas for your case studies, I'm always interested in head-to-head comparisons of experimental designs being done in companies and the adaptive procedures that I'm talking about now. So if you're interested in doing such things for your projects, I'm interested in helping you to do that. And you can see my email there, in case you want to take me up on the offer. OK? All set? All right. Good day.

[APPLAUSE]

--Problems that we run into quite often in microfabrication is that the metrology of the things that we've made takes up quite a lot of our time and budget. It's quite easy to vary parameters and make many samples. But choosing which ones to measure might be the biggest challenge, and I wonder whether you've looked at situations, where the acquisition of the data lags, the doing of the experiments, but it's sort of concurrent, but you can only measure a certain portion of the experiments you do. Are there more complicated situations like that where there's a lag? But it's not the case that you do all the experiments, and then you do all the measurements.

PROFESSOR: It's an interesting question. So you're saying that you can do an experiment. There's a time lag. Then you get an outcome. And in some cases, it's actually more complicated than that. You get some indication right away, but the more preferred measurement comes later.

So for example, sometimes there's a categorical variable that becomes immediately obvious. Either I heard chatter, or I didn't. And then later, I get a nice measurement of surface condition, something like that. Well, any of these delays tend to weigh in favor of the factorial design as compared to the adaptive. You just admit that right away. But then I'll say that sometimes, even with categorical variables, immediately available perceptions, you can get a lot of the benefit of adaptation from that.

PROFESSOR: Yeah, OK, thank you very much. Anybody else? So we've arranged to have an extra half hour today where I can answer questions about problem sets and any questions that you might have in the run-up to the quiz. So I think what we'll do is just go straight into that. If people want to run off at this point, then that's fine.

AUDIENCE: Are you going to [INAUDIBLE] the solution of the problem sets [INAUDIBLE]?

PROFESSOR: Yes. Yes. So there's a few housekeeping things first. Yeah, so problem sets, I will-- I'll put the solutions for 6 and 7 today at some point. And I put the 2006 quiz, too, and solutions on the website. I'm still trying to track down last year's quiz. But hopefully that will be out today.

And yes, for problem set 8, I know some people have been sending me questions, which I'm happy to continue to answer by email. I'll be around this afternoon if you want to come and speak to me in person. And yes, last week, a few people asked me for extensions. That's fine. If you need an extension, you can have an extension. I'm not super strict about these things.

OK, so really, the floor is open for you to ask me things. And oh, yes grades for 6 and 7 will hopefully be done within the next day, as well.

AUDIENCE: [INAUDIBLE]

PROFESSOR: Yes? Go ahead.

AUDIENCE: So we have a question. When you compare the Taguchi method and a surface response, this is for problem 4. But you want to compare the number of experiments it takes for both methods. Are you considering the quadratic terms [INAUDIBLE]? Or are you're not?

PROFESSOR: In question 4 on problems set 8, we're talking about here? Yes?

AUDIENCE: That's right.

PROFESSOR: I think that it's good to not include the quadratic terms in that question, although you might want to consider what would happen if you did.

AUDIENCE: OK, so we're only considering the linear terms and interactions?

PROFESSOR: And interactions, yeah. I think that's right.

AUDIENCE: OK.

PROFESSOR: I think that's the right way forward. Yeah?

AUDIENCE: Only two [INAUDIBLE] interactions or three or four [INAUDIBLE]?

PROFESSOR: Yeah, so with this, we're not really interested in quadratic factors. Just look at the linear factors in the interactions. OK? More questions?

AUDIENCE: Can I ask one last question?

PROFESSOR: Sure.

AUDIENCE: So actually, [INAUDIBLE] we are kind of confused about question 2, problem 2.

PROFESSOR: Yeah, everyone's confused about that. [LAUGHS]

AUDIENCE: So can you give us a little [INAUDIBLE] we are looking for in question 2?

PROFESSOR: Yeah. So I wrote this question in an attempt to stretch everyone, including myself. So it's pretty subtle, I think. But I would say the right way to approach this is to recast that model in terms of differences between the variables and the mean input.

So if you have a set of data, and those data have a mean value for x1 and a mean value for x2, so you might write it with some new coefficients a-- x2 minus x-- sorry, x1 bar, x2 bar. And that will be for some values of the a's that are functions of the b's. That will be equivalent.

And then, if it was just the first two terms, that would be exactly the case that was derived in Mayo & Spanos. And then this term looks very much like this term, except it's for a different variable. So the standard error of that coefficient, you can get at by the same way you got at this one.

And then if you look at this term, well, we have this coefficient here. Well, if you're trying to evaluate the variance of y hat, you've got a product of two quantities. Do you have an expression for the variance of a11, which you're going to get at in a similar way to how you get at the expressions for these variances.

And you'll see the expression for the standard error is a function of-- is going to be a function of x1 minus x1 bar. And you know that when you're finding the invariance of some constant times something that you know the variance of, you square that constant, multiply it by the variance of the thing that the constant was multiplied by.

AUDIENCE: Should it x1 star or x1 [INAUDIBLE]

PROFESSOR: Oh, sorry, yes. Yes, these are-- exactly. There you go. So this is-- yes, you're trying to find an expression for the confidence interval as a function of x1 star and x2 star. But at a given at a given combination of x1 star and x2 star, this is the thing that has variance. And this, for a fixed x1 star, is a constant.

So I don't want to give too much away. And I think this is the right way. There may be some subtleties that have escaped me. So if you can think of any, let me know. I think because there aren't any interaction terms in our model, I think that means that covariances don't need to concern you. But if you think I'm wrong on that, let me know. OK?

AUDIENCE: So speaking of the covariance, but x1 and x1 square, these two are correlated. So the covariance-- you have to consider that one, right? How could you ignore that? x1 and x2, they are not correlated-- x1 and x1 square.

PROFESSOR: Yeah. Yeah, maybe I'm wrong. Well, if you want to-- if you can work out how to consider that, then tell me.

AUDIENCE: OK, but in terms of the results, can we just use an x-- the matrix as a general-- we don't need to break down the matrix, right?

PROFESSOR: No, I don't think so, no.

AUDIENCE: OK.

PROFESSOR: Anyone else? Have we run out of coffee? Yes, I expect so. Let's see. What are the things that people are most puzzled by in the material since quiz 1? I mean, if someone asked you to fit a model to a factorial experiment, and you didn't have Minitab what level of confidence would people have in doing that? I mean, is it really something where you rely on the structure that the software provides to know how to calculate some of the squares and so forth? Because I can foresee exam questions-- obviously not going to have Minitab. We're going to be dealing with very small numbers of data, and it's just going to be a case of knowing what to subtract from what, really.

AUDIENCE: [INAUDIBLE] practice.

PROFESSOR: Yeah, absolutely.

AUDIENCE: [INAUDIBLE] with the [INAUDIBLE], you could do it. But we won't [INAUDIBLE] practice doing it versus using JMP or Minitab [INAUDIBLE] get used to the data being [INAUDIBLE].

PROFESSOR: Yes.

AUDIENCE: I think it would be more time-consuming.

PROFESSOR: Fair point.

AUDIENCE: [INAUDIBLE]

PROFESSOR: Yeah, I need to I need to track it down, because they didn't post it on Stellar last year. So I need to get it up [INAUDIBLE]. Yes? Who's that?

AUDIENCE: [INAUDIBLE] question 4-- can you explain more about the noise? You mean that you cannot [INAUDIBLE] during an experiment? [INAUDIBLE]

PROFESSOR: Sure. Sure, yeah, this question's a bit cryptic. But really, all it's asking you to do is confirm your understanding of the last question on problem set 7. You may recall from that question where you made that factor z, that noise factor, you were imagining that you could control it in that situation. But then I think you will have written an expression for y in terms of the two X's and the z factor that showed that the sensitivity of the output to variations in z were a function of x1 and x2. So you could choose a combination of the control parameters at x1 and x2 that would minimize the sensitivity to z.

Now, really parts B and C of problem 4 are just sort of checking that that went in, because if you can't control z, but you know that if you're not controlling it, it's probably going to have a variance, the input, the noise input, is probably going to have a variance that doesn't change with time. So you can see how much of that noise propagates through to the outputs if the noise interacts with the control input variables-- in other words, if there are these product terms of z and x1.

You can't control z. You can control x1. Therefore, you can control how much the noise propagates through to the outputs. And B and C aren't asking you to write down very much. It's just sort of checking that the concept is there, really. Does that answer your question?

AUDIENCE: Oh, yes. Thank you.

PROFESSOR: Thanks. OK, anyone else? Dave?

AUDIENCE: Do you have an example of a potential modeled problem that might be of a smaller data set that might be a little easier to go through some of the calculations of fitting a model to the experiment? I know a lot of our [INAUDIBLE] a little bit bigger, and it may be useful to be able to work something through.

PROFESSOR: That's a very good idea. Yes. Yeah. I'll try to come up with something that's sort of manageable on a sheet of paper-- yes, very, very good thought. We may be able to do with some of the data sets we used to look at ANOVA. But yeah, fine.

AUDIENCE: I had some confusion with problem [INAUDIBLE] the lack of [INAUDIBLE] I was having some trouble with [INAUDIBLE]

PROFESSOR: Right. Right.

AUDIENCE: [INAUDIBLE] if you could do [INAUDIBLE]

PROFESSOR: [INAUDIBLE] points, OK. Fine, fine. So that was on problem set 7. You were at 12-1 or something, but yeah.

AUDIENCE: I don't know [INAUDIBLE]

PROFESSOR: Right. No, that's a good point. A lot of people had difficulties with that. So that I will look at, as well, OK?

AUDIENCE: [INAUDIBLE] problem with something [INAUDIBLE]

PROFESSOR: Sorry? Something [INAUDIBLE]

AUDIENCE: For example, [INAUDIBLE] minus 1 [INAUDIBLE]

PROFESSOR: OK.

AUDIENCE: [INAUDIBLE] some examples. [INAUDIBLE] data versus [INAUDIBLE] so you don't have enough information [INAUDIBLE] I don't know why, but that's--

PROFESSOR: Yeah, I think there is a mistake in Montgomery. I need to track it down.

AUDIENCE: [INAUDIBLE] makes sense [INAUDIBLE] look at that. [INAUDIBLE] you don't really need to add all data and calculate all contrasts. So there is some [INAUDIBLE]

AUDIENCE: [INAUDIBLE]

AUDIENCE: [INAUDIBLE] example you have [INAUDIBLE] data.

AUDIENCE: [INAUDIBLE]

AUDIENCE: [INAUDIBLE]

PROFESSOR: OK, sure.

AUDIENCE: I'm not sure. On the Mayo & Spanos, on the 8.1.3, [INAUDIBLE] duration of [INAUDIBLE] of the parameters. It seems to me that [INAUDIBLE] just jumping and without [INAUDIBLE].

PROFESSOR: Yeah, you're right.

AUDIENCE: 8.1.3 It's a general form of the [INAUDIBLE] regression [INAUDIBLE] using the matrix [INAUDIBLE]

PROFESSOR: Yeah, right. Well, sure, I can try to find a complete derivation. I think it's probably not worth spending a lot of time trying to understand that. I would focus more on the stuff that's been in the problem sets. But sure, I can try to find that.

AUDIENCE: Because if you can get a derivation of that, any kind of regression data just changes form.

PROFESSOR: Yeah.

AUDIENCE: So there is no need to do that kind of thing, try to change the form of it. [INAUDIBLE] just need to fit it.

PROFESSOR: Yes, I'm sure you're right. I'm sure you're right.