

## 4.6 The Scientific Method in Practice

Let's take a little deeper dive into the traditional formulation of scientific method as it's generally thought about by people who do this stuff in practice. There are really five key steps and we know them, but I want to delve into them a little bit so we can get a little bit of texture about how they actually work. So the first step always is to formulate the question. And for most people, this is actually the most important part. And when you dig into it, there's a lot of disagreement on what are actually the best kinds of questions that people ought to be trying to answer. Are they the ones that we pretty much know can be answered in the course of a set of experiments, like, for example, which is more common among 30 to 40-year-olds, skin cancer or colon cancer? Or maybe the best questions are a bit more open-ended, like, why has colon cancer become more common in the United States in the last decade? I think what everyone has learned by rough experience as we go through our scientific endeavors is that poorly formulated questions are just really nasty things. They waste enormous amounts of time-- effort. They cause confusion. They cause frustration. Everybody hates them. And yet we've all seen questions that we just know from the beginning can't be answered through scientific reasoning of the kind that we're going to do with data. So I have really simple advice about that. Don't ask those kinds of questions. And don't let people around you ask them. Now, how to actually do that, we'll come back to that a little bit later. So let's assume we've got a decent question formulated. What do we do next? Well, the traditional scientific method says, find out what's already known that bears on this question. That sounds obvious. Why wouldn't you do that? But it's actually also extremely tricky. Think about 20 years ago. The risk for someone like me 20 years ago was that there might be a piece of knowledge out there tested and confirmed perhaps that you really or I really would have wanted to know before I started my experiments or my research, but I just wouldn't know about it because I wouldn't have found it. It was sitting in some obscure library somewhere on a dusty stack that I never would have found. In fact, when I was in grad school, this was everybody's greatest fear. How do I know that there isn't some other graduate student out there that has just finished precisely the dissertation that I'm just starting? This is in the pre-internet era, and actually, it was a real fear. Now, we like to imagine that today, searching for the existing literature or finding out what's already known about a question has become really easy. And there's no doubt that the internet has made a huge difference and made it much easier for us. But let's not forget that there's still sometimes a competitive aspect to this question. There are people out there who don't want you to know what's already known about this question, whatever it is. My grandmother in Brooklyn used to say, does Macy's tell Gimbels? And for any of you who grew up in Brooklyn, you'll know what that means. They're two competitive department stores. They're not going to share data on what their customers are doing. In many respects, that risk is still out there, but there's another risk. And it's even more insidious in some ways, and this is where the internet's made it harder. It's when and where do you stop searching for relevant preexisting knowledge and proceed with your own experiment. The web has made it really easy to just keep searching. Some people call it "analysis paralysis." Keep following the long tendrils of inquiry out into all the areas that might just somehow be relevant to your question. Advice-- on this one, there's no formula for how to do this perfectly, but there are some rules of thumb for knowing when it's time to stop the search. We'll come back to that later, too. Third point-- third thing we do is generate a hypothesis. Everybody knows this one, too. We're going to generate a hypothesis. Think of it as a hunch. It's a hunch about

what the answer to the question might be. Well, there are good hypotheses and bad hypotheses. The best hypotheses are those that you draw for a reason. You have a reason to expect the answer to be x rather than y, maybe because of some other knowledge that you have, maybe because you have confidence in a theory which suggests that, even though it hasn't been proven. But the reason people sometimes call this a hunch is because sometimes hypotheses are actually just more like random guesses or gut feelings. Actually, you can work with both, as we always do. But it's very, very important to know and for others to know which are we really working with because the basis for a hypothesis in practice has a big impact on how quickly or how slowly we're going to be willing to discard it if a bad or discrepant piece of data comes in that suggests it's wrong. So for example, think about purely empirical A/B testing, where we don't really have an ex-- we don't know if the blue button's going to work better or the red button is going to work better on this web page. So that's kind of a hunch, and we should pay very, very close attention to the large randomized clinical trial of the blue button versus the red button. But if our bet is based in some kind of deep knowledge about how human beings choose between red and blue, then we probably want to look a little bit more closely about the surprising results that come from an A/B test which appears to falsify our hypothesis. We might be a little bit more resistant to discarding the hypothesis if we have deep reason to expect that it's right-- could be that the experiment was faulty in some way. So this can be a very tricky situation in practice. That's point four. Here's what we need to do next. We need to do the experiments. And actually, everybody wants to jump to that as soon as possible. Sure, I do, too. I want to get into the lab. I want to get that hypothesis, I want to make predictions about something that could happen, and I want to design an experiment that evaluates whether those predictions are correct. And I want to get to it as quickly as I possibly can, but it's hard. It's much harder than it seems because of all the statistical confounds that creep into everyday life, which we're going to learn about in other classes, because of all the variables that are so hard to control when it comes to human affairs, and-- oh, yeah-- because of gremlins. There are always gremlins in experiments. And again, we'll spend lots of time later on in this course-- other courses. But I want to highlight one important piece of this here. Progress towards shared understandings in real-world decision situations just depends enormously on prior acceptance by the people you're trying to convince that the experiment that you are proposing is a reasonable one to test the hypothesis. You've got to get that agreement upfront. Otherwise, it's just going to be too easy for people who don't like your findings to come back later and criticize the design of the experiment. There are obvious ways to get around this in large clinical trials, double blind studies, and other such things. But actually, in organizational decision making-- much, much harder. There are ways to get around this, there are ways to beat it, there are ways to work with it, but you can't skip the critical step. It is very rare that the findings are going to be so compelling that they're simply going to overcome the intellectual and emotional opposition of people who don't want to believe what you have to tell them. They're just not going to win. So you've got to get that agreement upfront. Finally, you need to analyze the results of the experiment, draw the appropriate and justified conclusions, communicate the results, and then, of course, iterate. It never ends. What matters here? The obvious stuff-- precise analysis, the right level of confidence in the conclusions, and crystal clear communication about both of those things to everybody around you. And this is easy in the easy cases, but let's talk about the hard cases. I said before, it's 2003. Does Iraq have weapons of mass destruction? Or to point to the example we're about to look at, it's 1999. Wen Ho Lee, a Chinese scientist at Los Alamos National Labs, is accused of spying. Is he guilty? This is the background to what we call the "analysis of competing hypotheses methods." 30 years in development for high, hard-- excuse me-- high stakes, hard issues,

ambiguous questions, where it's not just one person trying to figure out what's right but lots and lots of people, many analysts, all of whom are touching a piece of the story and seeing different pieces of evidence in different ways. And so we want you to play with this tool to see how it works. Go to the [competinghypotheses.org](http://competinghypotheses.org) web page, sample projects, and see one way to work with this. Spend some time, play with the tool, and see how it really feels in practice.