5.1 An Iterative Process, Part 1

So I'd like to tell a little bit of a detailed story about one of the great iterative research designs of the modern era. I think there's a lot we can take away from this story, and besides, it's just a fun one about how it is that the dinosaurs became extinct. By the way, it's not what that cartoon says. So how did the dinosaurs become extinct? Today we have a very strong view of this, but before about 1980, there were actually bunches of different theories and hypotheses floating around, people sort of making up stories. Some people thought the dinosaurs had just over-expanded. There were too many of them. Other people thought maybe they had gone through a mutation or they had fought each other off, just like in the movie. Many of these were deduced or they were just simple analogies from other extinction theories, and there really wasn't a lot of evidence to support any of these arguments.

Now, from the outside of the dinosaur world, or the world where people were worrying about dinosaurs, a new fact emerges. It's actually not yet a piece of evidence. It's just a fact. Two physicists here at Berkeley just happened to be related-- Lewis and Walter Alvarez-- along with two chemists that they were working with discover what they call a geophysical anomaly-- something that is just weird in the layers of the Earth. And it turns out that all over the world in sedimentary rock layers at a particular time in history, what's called the K-T boundary, there were really big concentrations of what is a very rare element called iridium, which shouldn't be there. In fact, the concentration of iridium in this layer is hundreds of times greater than anybody had expected, and these guys were wondering, what's that iridium doing there?

The geochemists point out, well, actually, it's sort of interesting. Certain kinds of meteors and asteroids contain really high levels of iridium. And actually, if you analyze what that iridium in the meteor looks like, it has an isotopic ratio that is quite similar to the iridium that they're finding at the K-T boundary. So people are starting to say, hmm, I wonder if there was a meteor or asteroid impact that happened at that time. And so now they look more broadly. Is there any other evidence to suggest that that might be right? It turns out that there were other things in the K-T layer that are usually seen in meteorites that nobody had ever looked for before because they weren't expecting it, like particular isotopes of chromium and quartz granules and these things called tektites, all of which are associated with meteor impacts on Earth. They'd always been there, but nobody had ever paid attention to it before. So now they've got a double-- hmm, what's going on?

The hypothesis-- maybe there was a massive meteor strike, and if there had been a massive meteor strike, this would have caused a massive shift in the Earth's climate at the time-- for example, a huge cloud of dust that would block off the sun, and there'd probably be an enormous amount of acid rain. And that would kill off many plants, and then the animals that eat the plants would die. And the further you go up in size, the harder it would be on animals to survive in what might have been a really long, dark winter, and I'm talking about thousands of years. That might just be the cause of the rapid extinction of the dinosaurs.

So what do we do now? We iterate. We look for an adjacent implication or observable. In plain English, if that is what really happened, what else would have happened along with it? Well, what's the obvious one? If a meteor big enough to do this kind of climate change hit the Earth, well, there'd probably be a really big crater somewhere, and so let's go look for the crater. And sure enough, other scientists, once they're looking, find a massive crater in the Yucatan.

And guess what? It's roughly the size of what the meteor calculated by the Alvarezes to have caused-- or have been sufficiently large to cause the extinction-- would have had to have been. By the way, just for trivia, the force of the calculated impact of that meteor-- roughly 10 to the eighth megatons, which happens to be about 2 million times greater than the most powerful nuclear weapon that's ever been tested on the Earth. So this is a big meteor. The Alvarez story is a real-world example of how this works in an iterative, pragmatic, and scientific setting, and it's a great story. You have this existing theory, or several, about extinction, and none of them really have any great evidence. A new piece of information comes into the debate. It's seemingly unconnected. Someone's imagination establishes a possible connection between that new piece of information and the phenomenon that someone is trying to explain. And then we do more investigation, and we find the crater. It's about the right size.

This then, of course, develops into a theoretical story about what that data could mean, which then, in turn, goes back into a detailed deductive analysis. Could the story possibly be right? Do the numbers actually add up? And then we go back to imagination. If the story is right and the numbers add up, what other things would have happened outside of the particular phenomenon we're trying to explain? And what traces would the other things have left on the Earth? And so now, let's go out and look for them. That's the reality of the iterative research design-- back and forth, back and forth, back and forth.

5.2 An Iterative Process, Part 2

So the Alvarez story which we've just talked about is a really good real-world example of how to do iterative research design in a pragmatic scientific setting, how things really happen. Let's just think about what this story really entails. People are trying to figure out, why did those dinosaurs go extinct? There are a couple of theories floating around there about how that extinction happens, but there's actually very scant evidence for any of them. Then this really new interesting piece of information shows up, which is seemingly unconnected. Imagination establishes a possible connection. Why is that iridium in that layer? That then develops into a theoretical story about what this could possibly mean, which in turn develops into a detailed deductive analysis. Could that story possibly be right? Do the numbers add up? Then there's more imagination. If that story is right, what other things would have happened and what traces might they have left that are completely unrelated? So let's go look for them. Then there's more investigation.

We found the crater. And oh, guess what? It's about the right size. And then the story iterates again. What other kinds of things are happening? How much acid rain would have fallen after that meteorite struck the Earth? How would the frogs have survived that acid rain and not the dinosaurs, et cetera, et cetera, et cetera? The point of this-- the Alvarez story is so compelling not only because we want to know why the dinosaurs aren't here anymore but because it's a story about a constant back and forth between theory, induction, deduction, and imagination. And that's the way it really works. Today, in a related set of developments, interestingly, that's more or less outside the scientific world, people are sort of grokking on to this way of thinking about a problem. And the rough analogy is they call it "design thinking."

5.3 An Iterative Process, Part 3

5.3.1 Research Often Generates More Questions Than Answers

This element addresses the following learning objective of this course:

- LO2: Design and apply research questions.

Part of the iterative nature of research is that we often generate more questions than answers. This may happen at several stages. You will likely develop multiple questions at the ideation stage as you try to narrow down the problem space. Let's say you work for a company that delivers groceries to users' homes. Your first attempt at a research question might be technical.

Perhaps your question is how do we most efficiently pair shoppers with incoming orders? In this early stage of question formation, you might realize that customers don't see the delivery windows until they fill up their cart. And this may lead to frustration because customers don't know until pretty late in the process that their food will be delivered later than they would like.

Now, this is a related problem because it affects the number of orders that come in, but you will likely need to separate this into two distinct projects. Because frankly, you can't solve everything at once. You might be able to focus on, one, how do we most efficiently pair shoppers with incoming orders and, two, what's the ideal time in the process to show customers the delivery windows?

You're also likely to develop questions toward the final stages of your project. Let's think about the grocery delivery example again. As you research how to best pair shoppers with incoming orders, you might find that in addition to the matching, you realize there is another issue.Your company must be able to predict how many shoppers will be available on a given day.

This is a completely different question. Can we improve how we predict the number of shoppers on a given day? The headline takeaway is that we should be open to uncovering additional questions throughout the research process.

5.4 Design Thinking, Part 1

Design thinking is about asking the right questions, and one of the key challenges before you is that there's never a shortage of questions to ask. For a given situation, you can come up with all sorts of questions. So how do you know if those questions might be right? Well, the right question, of course, addresses a real need or interest of either yourself or those that you're working with. Oftentimes, it engages others, and some might say it's provocative or even controversial. The right question is always specific enough so that you actually can answer it. The right question often doesn't start with what but why or how, and those are the sorts of questions that then lead to further questions from which you can glean insights. Now, asking the right questions, you'll find, is also a mix of both art and science. And like many things that are both a mix of art and science, you learn them and get good through continual practice, practice, practice. Why do you like certain questions? Find a question out there, whether in your company or organization or in the media, and say, why do you like that question? What makes it great? How does it apply to your situation? But then also, of course, look at the questions you don't like. What are the dumb questions that people ask? You can learn from doing this, and that's how this is a mix of art and science is through continuous practice. And that's how you learn.

5.5 Design Thinking, Part 2

Interview with Lisa Solomon

5.5.1 Design Thinking in Practice

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO4: Justify an analytic approach that informs decision making.

Welcome. In this unit, we've been unpacking the concept of iterative research design and how people do this stuff inthe real world. And we talked a little bit about how that intersects with the current contemporary idea that usually goes under the label of design thinking. I'm here with Lisa Solomon, who has spent the last several years thinking about design thinking. So Lisa, thank you for coming in.

Great, thank you. Hi. Great to be here.

Lisa, give us your sense of what design thinking actually means outside of the world of designers.

It's a great question, Steve. A lot of people are excited about design thinking and its hope for how it helps companies reimagine future solutions to help solve their customer problems. And I think that folks are looking to define design thinking in a box. And if you ask 10 people what design thinking is, they'll have 10 different answers.

My answer is that design thinking is an approach to solve some of our toughest challenges fundamentally with the user in mind. So there is a human element to it that you often don't get in more traditional analytical processes.

So we put the customer really, or the user, really, at the center of the discussion. Lots of people say they do that, but the discipline of design thinking is really trying to get disciplined about doing that.

It is. Thank you for saying that. A lot of people talk about design thinking as a creative approach to problem-solving. I think that's a little flawed for two reasons. One is that the creative side often takes away from the fact that it's highly analytical and very disciplined. The other side around calling it an approach to problem-solving actually jump starts the process, where design thinking often starts not with a problem, but with an unmet need.

Interesting, interesting. So the unmet need defines the problem, but not directly.

Correct. So when we start with a problem, that's already making a lot of assumptions about the thing that is causing pain or is not being met from a user perspective. But if you take time to actually watch users in their natural habitat or environments and see what's happening that either helps them get the job that they need to get done done or causes them a lot of frustration or pain in the process, those are things that you have to actually go and watch and observe them to see what's happening. You can't understand that from pure data alone.

Interesting. So in some sense, you're taking that classic question that people ask, what's the pain point, and pushing it back a few steps in the conversation to, where does the pain point come from?

That's right, that's right. And the most important part is actually going to the source to do that, not looking at a bunch of market research and segmenting customers by definitions that marketers came up with.

Now, sometimes this concept might seem a little uncomfortable for people who work with numbers and hard data. Ismost of that work now done in a qualitative fashion, focus groups, observation, ethnography? Or do you see people starting to integrate disciplined data collection into that, say, early phase of the conversation?

I think that's a huge opportunity, Steve. Right now I think there is a bit of divide around what I imagine some scientists say is the softer side, of going deep around a few customer segments and spending a lot of time watching them and seeing what they do. May not prove to be statistically significant, and therefore defendable in a boardroom. But what happens when you spend that kind of time with users is that you start to understand their stories, and you start to understand them from the whole person perspective, not just in the small snippets that, say, a single product serves.

Could you tell us a story where you think moving back in the design process to, again, what lies behind the pain point has really helped and made a difference in people's problem-solving?

Absolutely. One of the exciting things about design thinking is that it's now getting taken out of just the pure products and services realm and into much broader implications, even for society. So I'll tell you a very recent case that's just in my neighborhood, down in Menlo Park, California, where the administration of our local public school was trying to address the problem of middle school students who were feeling very stressed out by their class load.

And so it's very easy to look at that from the analytical perspective to say, how many minutes in class are they spending? How many subjects do they need to take? And instead, what they did was teachers paired up with students and actually followed them around the entire day to experience what it felt like from the moment they got to school to the end of the day, of exactly what it felt like to be a middle school student.

And as a result, they've dramatically changed the class schedule to be much more forgiving about how middle schoolers could actually take in the information and opportunities for them to rest throughout the day in order to be more productive in their learning.

Really interesting, the thought of being an adult and having to go back to be a middle schooler again is not exactly an attractive proposition, but if it works to solve the problem, maybe we ought to try things like that.

That's right. I think what it does is that it heightens sensitivity to the points in between. So we're getting better at looking at these customer touch points, but we actually need to look at users for the full journey, what happens from the moment that they start to engage with, let's say, a school to what happens when they leave, and then even looking beyond that to say, if really the question is how do we promote health and wellness in our kids, it's a much bigger, broader question than, say, how do we make sure they learn math concepts.

So let's play with this story a little bit. Let's imagine that one of our scientists that was really interested in mobile applications and wearable computing came and said, I'd like to add to that experiment. I'd like to put a Fitbit or something equivalent on the kids and the teachers to test their skin temperature and their galvanic skin resistance over the course of the day to see when they're sweating or what their heart rate is as another piece of data.

Do you think that interaction would be productive? Do you think it would get in the way? How do you think that's going to work? Because people are going to want to do things like that.

Absolutely. I mean, privacy issues aside and all those things, I think that's going to be tremendously helpful. Because right now still, the methods we have to see how people are feeling are crude. We often ask them, how are you feeling right now? And what we know is that people lie.

Or they don't know themselves.

Or they don't know, right, exactly, that they say, oh, I'm feeling great, because that feels like the socially acceptable answer. But their physicality may be telling a totally different story.

Interesting. So we might be able to instrument some of that and then put those pieces of information together in away that makes the whole greater than the sum of the parts.

Absolutely. I think that part of the important discipline of design thinking is to actually be clear about what hypothesis you're testing. And so if you can integrate the more open-minded questioning about peeling back questions to say how might we understand what a day in the life is like of students or how might we improve their overall learning capacity, and then marry that with instruments that could collect data that might confirm or disprove our hypotheses, Ithink that would be very powerful.

Really interesting. So you spend a lot of time teaching students who are trying to become experts in design thinking.Tell us a little bit about the teaching techniques. How does one learn design thinking?

That's a great question, Steve. Well, I'll tell you how they don't learn it, which is in textbooks. They learn it by going out into the world and actually applying the process and the skills against a real-world problem.

So learn by doing.

Learn by doing. So last year, for example, the focus of our class was on the future of work. And you might assume that the initial hypothesis is work-life balance is an issue. So we need to figure out how to marry work-life balance in ways that's more palatable. But when students actually started to spend time with professionals to find out what they did every day, what they realized was that it wasn't really a work-life balance issue as much as it was about what kind of work they did when.

So in this example, what they found was that there was an inconsistency between what they called flexible work and inflexible work. So inflexible work were the meetings you have to show up at, and that that inflexible work was crowding out time in their calendar to allow for the real thinking that they then had to do late at night.

Late at night.

So ultimately, by using the process of design thinking, they came up with an idea of how to engage that person's bestthinking during the hours that they had the energy to put towards their thinking without compromising their responsibilities.

Really, it sounds like we're just at the beginning, really, of the interaction between design thinking and data scientists.So let me end by asking you, if you had a room full of design thinking people who are in that world, and you had a room full of data scientists, people who were in that world, and you wanted to bring those rooms together in some productive way, today what's the one piece of advice you would do to make that interaction work?

Great question. I think finding a live project to work on together would be really important. And I think the other big thing that would help these two groups come together is to create a common vocabulary. One of the things that I see happening with design thinking is that people are talking about the same kind of idea around human-centered problem-solving, using empathy, iterating with a bias towards prototyping and getting live feedback. But they're talking about it in very different ways.

So for example, in the entrepreneurial world right now, a lot of people are talking about the lean startup methodology around customer development. But if you peel that back, there's actually a lot of similarities with design and design thinking. But they're not necessarily using the words empathy. They're talking about it from the standpoint of understanding customer needs and the jobs to be done.

So I think it's about figuring out a common language that will meet the needs of both groups to feel like that there's something productive happening, but also allowing them to engage in making progress forward towards really what is a discovery-oriented process.

Well, we really hope that this discussion will have contributed to people thinking along those lines. Thank you so much for the practical insight about how to do that.

Great, Steve, thanks so much for having me.

5.5.2 Design Thinking in Practice: IDEO

View the video from IDEO Design and Consulting. The following text is quoted from the video description:

"Set the stage for a successful brainstorming session by sharing these rules with your team:

- defer judgment
- encourage wild ideas
- build on the ideas of others
- stay focused on the topic
- one conversation at a time
- be visual
- go for quantity"

This element addresses the following learning objectives of this course:

- LO4: Justify an analytic approach that informs decision making.
- LO7: Imagine, plan, and design a data science project

IDEO Brainstorming Video from IDEO U

5.6 Understand the Context Behind the Ask, Part 1

Interview with **Michael Dalby**

5.6.1 Understand the Context Behind the Ask, Part 1

This element addresses the following learning objectives of this course:

- LO1: Describe the role of data science in organizations.
- LO4: Justify an analytic approach that informs decision making.
- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

Welcome. In this unit, we've spent a lot of time talking about how to reframe questions, how to ask better questions, and how to calibrate research agendas that are often decided by other people so that we can do a better job as data scientists answering the questions that we really can answer and having an impact.

And to reflect on some of the experiences and rules of thumb and ways of doing that in practice, we're going to talk with Michael Dalby, former management consultant who has had an enormous amount of experience having to deal with this in really practical and sometimes very trying situations.

So Michael, thanks a lot for talking about this with us. You've told me more than once that the ability to reframe people's questions to be more effective has a lot of intuition in it but not just intuition. There's actually a kind of a method to the madness, as it were, that you've extracted overtime. Could you tell us a little bit about that?

Sure. It's an excellent question, by the way, and one that data scientists I think in particular need to take account of in understanding what their opportunities and their job is. A data scientist's one premise might be to have the assumption of objectivity. And the work that they do about data ought to be objective and received as disinterested by--

That's why we call them "scientists."

That's why we call them "scientists." They operate in one world, and chief executive officers operate in a different world, which is a world partly made up of facts but also jammed with uncertainties, likelihoods, unlikelihoods, questions, and problems.

So one of the issues that I've observed over the years both as a management consultant and actually as a member of senior management has been the meeting of those two worlds in the most effective possible way. And as you say, it's not only an issue of intuition. There are some techniques you can use.

One of these is a technique to uncover, if you like, the motivations behind a given request. Let's say, for instance, that the CEO of the company you're working for poses a really difficult question for you, one involving a lot of uncertainties about the future, and expects you to shed light on it. In fact, that's a good thing because at this point, your CEO is not wasting your time on routine matters. She is bringing this to your attention as a real problem.

So let's say that, picking up the example we used in our previous conversation, she's asked you to forecast and understand the market for wearable computing devices. Now, the technique you might want to use here is to understand what the motivation was of the CEO for asking that question so that you can understand the context in which your answers, your data, your insight would be most relevant.

For example, a construct that I often think of in such situations is if-then-by. If we understand that the market for wearable computation could grow, then we might invest to take advantage of that growth by, for example, buying a couple of companies preemptively in order to get ahead of the pack. So that's one motivation, a perfectly plausible one.

But it may not be the one that your CEO has. She may in fact be a skeptic about the whole issue. So it may be worthwhile to take that on advisement and think of alternatives and go back and check because it could be that we understand some basic facts about the market.There's potential, but we don't know how much. There's disposable income, but we don't know the fraction of disposable income that will go towards that market.

We have some notion that there are qualities that the public desires in wearable computing devices. The first one is style, let's say, the second one is convenience, and the third one is functionality in that order. We may have discovered this, as a matter of fact. The question is, what does that do in terms of the motivation?

And it could be that the CEO is thinking, unless or if there's a relatively high degree of certainty that these things can be fashionable, then I Don't want to participate in this whole market as a leader and by, in effect, taking a very deliberate, cautious approach, perhaps joint venturing in some kind of distribution arrangement with competitors for the time being. This makes a huge difference.

It's really interesting. I think those of us who are sort of born as or have the mindset of researchers, when we're given a research problem, we just want to find a positive answer. That's our instinct. But sometimes, actually, it may be the people who are asking us a question want us to find a negative answer. And it's not as exciting to walk in and say, there's really nothing there.

It's a problem for data scientists, partly because of the mindset of people who are scientifically inclined and who do work with facts and their patterns and implications, to achieve that level of objectivity and present the results as such. That may or may not address the problem.

So the problem is not the same as the data. The motivations are not necessarily the same as objectivity. But that does not necessarily mean,either, that one would change one's conclusions in order to fit a given situation. What would change is one's frame of reference and the salience of this aspect or that aspect of the data issues.

And presumably, the entire research plan can be made more efficient if you have more information going in about the motivation of the asker. It Doesn't have to bias the answer in any way, but the plan can be more efficient under those circumstances.

It's absolutely correct, and you want to make sure that to the maximum extent possible as a data scientist, you've heard that context. There's a famous cartoon in which a painter is painting on an easel and has laid down a painting. And the person comes to the door and says, the king would like to know, Andre, how you're doing with St. George and the dragon? And it turns out that the painting has St. George and a little red wagon.

Ah. So thanks, Michael, for these reflections on what it means to reframe questions and be efficient in research design. I think it's really useful insight, and it's something that data scientists and management people are going to have to actually learn about each other as they work together more intensively over time.

There's a great deal to all of this, and I'm glad that we've had the opportunity to talk about it.

Thanks again.

5.7 Understand the Context Behind the Ask, Part 2

In this segment, we've introduced a lot of different ideas about asking better questions, about helping other people to ask better questions-- about formulating research designs that try to give people a leg up at least in trying to develop answers to those good questions. And some of that comes from theoretical derivation, some of it comes from the rules of inference, and we just have to be honest. Some of it is really just experience and accumulated wisdom that people have developed over time about what actually works and what just does not work. And there are people who are just really good at this. And they're really good at it mainly because they've had a lot of experience. They usually have a kind of personal inclination to want to work in this way, and they've had a lot of practice. And so what they tend to do is almost accumulate rules of thumb over time that are really about finding ways to enhance connection to the data, to develop the value of the information that they really can collect, to be really efficient with their time in terms of the research that they want to do, and enabling themselves to do what I've heard people refer to as "deep listening." And you know what I mean by that, the difference between when you're kind of listening to someone and you're really listening deeply to try to penetrate what it is that they're truly saying but at the same time, with all of those kinds of capacities not, as they say, leading the witness too much because you don't want to manipulate what you're hearing. And so when I think about the people that I've met over the years who I think are really, really good at this stuff, I think of some really world-class management consultants. They have a kind of accumulated wisdom about how to do this. I actually think of survey researchers who have spent a lot of time and wasted a lot of effort asking the wrong questions on their surveys and have to go back and redo them. I feel like software engineers, the ones who are really skilled at figuring out what their customers actually want in an interface-- sometimes intelligence analysts, who just, again, have a kind of intuition for what the customers downtown really want them to say. And then finally, there's the really great physician, the doctor who can walk into a room and without even looking at the chart, she knows exactly what question to ask the patient to find out what's really going on-- just sort of feels it as a result of all that accumulated experience. So let's take advantage of some of that accumulated experience by talking to some of those people.

5.8 Understand the Context Behind the Ask, Part 3

Interview with **Jonathan Star, Consultant**

5.8.1 Elegant Questions and the Importance of Understanding the Context

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO7: Imagine, plan, and design a data science project.

Welcome. In this unit, we've talked a lot about research design, which is, in some ways, a fancy way of saying how to ask better and more effective questions. And to delve into some of the practicalities of doing that in the corporate setting, I'm going to talk to Jonathan Star who is a consultant with Monitor Deloitte. Jonathan, thank you for talking with us about this today.

Thanks, Steve, it's really good to be here. Thanks.

Jonathan, we've talked a lot over the course of this unit about some of the general principles of research design and how people design really good research projects over time. But I want to cut to the heart of the matter and go back to the issue of the really powerful question, the elegant question, the beautiful question that reframes a problem for a person or an organization. Tell us about your experience. Surely you've been through that.

Very much so. In some ways, in the work that I do-- scenario planning-- good questions are essential. We work with clients because they're dealing with some significant uncertainty. They don't quite know which way to turn. They Might have new competitors coming into the market. They might be facing some kind of change to that business model. They might have new regulations in place as well. So any of these things might be in play at any particular time.

Or all of them at once.

Or all of them, that's right. And so the technique of scenario planning is one very much of opening up. It's trying to get clients to explore a range of different situations that they might find themselves in in the future.

Now, because it's about opening up, there is a temptation and a tendency to frame the researcher, frame the study, ina very broad sense. And so, for me, a good question is remarkably important because it gives you the right level of focus. And it's right to think, Steve, about a question that might be beautiful or elegant. But let me bring it down to, Ithink, in some ways, the most important characteristic is, is the question useful? Is the question useful for the group or the client to make progress on? Does it make sense for them given their capabilities to actually answer question Rather than question B.

So you've got this interesting kind of tension where you have people who have really wide cones of uncertainty. The Risk is that they're going to become unfocused as you try to open them up. But what you do is you try to formulate what you call a focal question. Tell me a little bit about how you actually do that, and maybe even an estimate of how much time you spend in a project actually just formulating the focal cluster.

Well, I usually find that I spend-- let's say a project lasts three months. I would say that you probably don't settle on the question until halfway through that time, even though you might be doing other work, you might be doing a lot of interviews, a lot of research, even getting engagement from the clients in terms of workshops.

And yet, maybe the settling of the question isn't a once and done thing. It's something that happens as the project unfolds and as you find out more about the situation at hand, and as the clients find out more about what they're interested in and what they're passionate about. So it takes a long time, and I think if done right, it's very iterative.

It's very interesting because it's sort of an unfamiliar, unconventional cut on research design for most scientifically minded people that you would ask a question and collect some evidence, and then go back and reformulate thequestion, and keep going through that process. And the idea that it's actually halfway through the project that everybody finally realizes what the question actually is. How do when you're there?

I think there's a palpable change--

Ah.

--in the energy of the client. So let me give you-- let me give you an example. So a number of years ago, I was working with an insurance client. And this was an insurance company that was family owned. And it worked in providing insurance for collectors of classic cars. So think '65 Mustangs, '59 Thunderbirds, something like that. Andthey wanted to explore what their future might be like in the next 10 years or so. They'd achieved all their goals up until that point and wanted to set a new course for the organization.

Now, they're an insurance company. So we had our initial question down as, well, what's the future of insurance? Andtherefore, as a result, what should our positioning be, or what should the client's positioning be to take advantage of the changes that are likely to happen in insurance?

And so we conducted a number of pieces of research. We got everyone in the room together. And this was probably five or six weeks into the project. And we then started exploring this question-- what's the future of insurance? And so we got them thinking about, is the insurance industry going to become more consolidated, or is it going to become more fragmented? Are you going to find that there are lots of big players or lots of small players? Are you going to find that it's much more regulated than there has been in the past, or maybe it gets more deregulated? Is it going to be run at a national or international level? Or is it something that's done much more to the state level? So all great uncertainties when you think about insurance.

And so we went through this exercise. And not only did we have the client and the client executives in the room, we had a number of outsiders as well who bring this great kind of alternative perspective. And one of the outsiders on the second day decided to speak up and said, you know what, I'm intrigued. You guys don't sound like an insurance company. You don't ask questions like an insurance company. You don't worry about loss ratios. You don't worry about the kind of things that insurance companies do. The only time that I find that you're bringing a sense of energy and passion into this conversation is when we're talking about the collectors, when we're talking about the cars.

And what our guest did at that point is they moved that conversation and essentially, said, the right question to ask here isn't what's the future of insurance. The right question is, what's the future of collecting?

Interesting.

Once that happened, the energy in the room, as I said, just changed palpably. People got it. People nodded. People Understood that they were in a different business. Insurance might have been their engine. But actually, what they were really about was serving collectors.

It's a very interesting example for people who are-- or think of themselves as data scientists, that the insurance business would be a natural place to revolutionize insurance with very, very large data sets and the ability to extract insight from them. But in this case, that actually would not have been the most effective way to go with the project.And you knew it when it happened. Everybody felt it at that moment.

That's right. That's right. Whereas for other clients, it might well have been the case. And I think, certainly, there's great opportunities in data, particularly to understand the situation as currently given. I think in many ways, a lot of the things we do early on in our scenario work is trying to understand-- trying to predict the present or understand the present, rather than paint stories of the future.

And I think data science and data has a tremendously important role in giving us more insight into the present than we already have because executives, at the moment, rely so much on hunches. They rely on assumptions were prone to biases. Now, those things are never going to go away. But if you can bring data into that understanding of today's situation, that would be tremendously powerful, and I think a great adjunct to the work that we do in scenarios.

Now, I'm imagining a situation where someone who is first and foremost identified by a client or by the organization--or self-identifies-- as a data scientist, being the person who stood up at that moment and said, wait a minute. This isn't really about insurance. That's not where you really are. In a funny way, if it was a data person who said that, made that argument, it would have been even more impactful.

I think that's exactly right. I think that's exactly right.

Well, Jonathan, thank you so much for talking with us about the pragmatics of what it really means to ask better questions. It's been really, really helpful.

You're welcome. Thanks very much.

5.8.2 Ask Better Questions

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO7: Imagine, plan, and design a data science project.

We've mentioned before different ways to answer better questions. Here I want to focus on a few concrete recommendations. One, make sure the outcome of interest or dependent variable is clear. Even if the outcome is amorphous, you can worry about how you will operationalize it later.

For example, how do we predict fraudulent charges? Here the outcome of interest is a fraudulent charge. Or how dowe increase brand loyalty? Here the outcome we care about is brand loyalty.

Number two, be cautious about sneaking the answer to your question into your research question. Maybe your question is how do alerts influence operator fatigue? In this case, you specifically care about the influence of alerts on fatigue.

But maybe instead, you really care about overall fatigue. And you have a hunch that it's the quantity of alerts one receives that impacts or that has an undue impact on fatigue. And so you want to look at that as one plausible explanation.

But in this second case, you unintentionally snuck the answer to your question into your research question. You'vealready assumed that alerts have a major impact on fatigue. So maybe it'd be better to broaden your question simply to what leads to operator fatigue?

Idea number three, it's important to uncover the motivation behind the question or request. If you understand the context around the question, you may be able to redirect the project in a meaningful way. You may ask, what's the reason or the focus on this particular metric? Only if you immerse yourself into the problem can you know enough to redirect the client in a more productive direction. Let's go through an example.

Think about a client who insures classic cars. This may sound familiar. Their initial question is what's the future of insurance? But after spending some time with the group, you realize that the question they should be asking is what's the future of collecting classic cars?

Four, questions are specific enough-- ensure that questions are specific enough so you can answer them. This might mean breaking down a big question into smaller pieces because we want to make sure that the question is narrow enough so we can answer it within the time and financial constraints.

And so let me step back a little bit on that final point. In the early stages, think big. In the ideation stages, don't worry about money, policy, et cetera. Think about the best question. And think about what you really care about and what would be the ideal design to tackle that question. If you aim really high, then you could always adjust your design based on what's realistic given the timeframe and the constraints.

Now, why am I recommending this approach? Because if you start with realistic expectations, you will inevitably have to pull things back a little bit anyways. So start big and then pull back. Finally, don't let the perfect get in the way of the good. Yes, ask yourself, can I make the question better? But remember, it doesn't have to be perfect. In fact it rarely will be.

5.8.3 Mirror Back the Ask

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO7: Imagine, plan, and design a data science project.

In order to get meaningful, actionable insight, we need to understand what's being asked. One way to do this is to mirror back. After someone describes the ask to you, mirror back and say to them, what I hear you say is, and then articulate what you understood. This will help eliminate any misunderstandings and will also help protect the relationship because it shifts the onus on you to understand the question.

If the client or manager indicates that what you heard is correct, then you could proceed more confidently with the next steps. If instead, the person says, no, that's not what I mean, then you could work back and forth until both parties understand each other.

Another way you could use this mirroring technique is to help guide the client towards a better question if you think the client is missing the mark. If someone says, I want to prove that we're better than our competitor, you could mirrorback back what I hear you say is that you want us to identify your company's competitive advantage compared to others.

Here you've stepped back a bit from the word prove and have added a little room for a broader analysis that allows for an identification of where the company is better than the competitor and gives you room to discuss areas where the company frankly isn't.

Here's one example of how mirroring back can work. The initial ask from the engineers is can you reduce the number of alerts they receive. They're getting too many alerts. And you're trying to figure out, OK, well, what's the problem here? So the data science team asks the set of engineers the following clarifying questions. Are you annoyed that the signal-to-noise ratio of the alerts you receive make it difficult to identify the severity?

If you're annoyed that you can't tell right away if the notice is significant, then we could send fewer alerts and only send alerts with a higher level of severity. But if instead the problem is just the overall number of alerts, we could shift the alerts to other teams that have higher resource levels, or we could simply aggregate and reduce the overall number of alerts.

5.9 Refine the Question

5.9.1 Refine the Question

Spend five minutes on the following prompt:
Describe a situation where you were presented with a question that was ill conceived. Were you able to reframe or reformulate the question? Describe what happened.

5.10 What's the Problem? Question Formation

5.10.1 Problem Statement

This element addresses the following learning objectives of this course:

- **LO2: Design and apply research questions.**
- **LO7: Imagine, plan, and design a data science project.**

What's the problem? This is related to question formation, but here I want to talk about defining the problem statement. You can think of this as pretty similar to articulation of the business problem. It's our responsibility to convince our audience that the status quo or current condition is problematic.

We have to convince our audience that this work needs to be done. We have to convince others that this problem merits the time, effort, and money we need to try to come up with a different alternative. What you perceive as a problem, what I perceive as a problem may not seem like a problem to other people at all.

One way to define a problem statement is to think about this issue in terms of deficit or excess. Rather than say,mechanical failures are a problem, you can say, there are too many mechanical failures, or you can say, mechanical failures occur faster than our ability to repair the machines. We need to find a better way to deploy preventative maintenance. And so the tactic that I employed just above is to use the word too, T-O-O, in your problem definition.

Couple more examples. There is too much content that bypasses our content moderators, or our content moderation strategy is growing too slowly to keep up with online content. When possible, quantify or give a sense of magnitude.

You can say, we're using too much water in our processes. But it's more informative to say, in the past two years,we've used 10% more water per unit. You can say, the complaints of our top customers are addressed too slowly. Butit's more informative to say, the complaints of our top customers, who account for 70% of our sales, are addressed too slowly.

5.10.2 Question Development

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO7: Imagine, plan, and design a data science project.

After you've identified the business problem, you need to translate the problem into a question. The central question of a project is the guiding focal point. You could almost think of this as the mission statement for your project. You will continue to come back to this throughout the entire process. If you're ever unsure about a decision somewhere along the project, ask yourself, how does this help me answer the core question of my study.

Let's go through an example. Let's say you work for the California Strawberry Commission. After much iteration, your final question is, how can we better predict strawberry yields? As you develop your design, you decide that you want to interview farmers to get a better sense of how they predict yield and to get a sense of just general best practices.

You begin to construct a list of semi structured questions that you'll eventually ask farmers. But during this process,you're also a little curious about farmers' attitude toward the California Strawberry Commission but you quickly remember that your main question is, how can we better predict strawberry yields. And you're concerned that this additional question, what do farmers think about the commission, does not help you answer your main question. Andit takes up important interview time.

And depending on where in the survey or where in the interview you ask the question, it may unintentionally impact farmers' response in a way that, frankly, you didn't intend. And it might also cause farmers to lose a little trust in the interviewer because they may suspect that this is some kind of investigative project. They might be a little extra careful about the responses.

You'll have one central question to your project. But you may have several sub questions that are nested under the central question. Let's explore the strawberry example a little bit more. So remember our main question is, can webetter predict strawberry yields. Some sub questions are how accurate have we been with predictions in the past five years. What does precision agriculture or how does precision agriculture affect our ability to predict yields. How do our predictions vary for organic versus conventional fruit?

Depending on the method you use, you may or may not explicitly test hypotheses. Nevertheless, you may think of these sub questions as a foundation for your hypotheses. Keep in mind that you'll most likely iterate in the question development phase. This means that as you talk to others, you might reframe your question. As you develop sub questions, you may realize that one of those sub questions should be the main question.

As you begin to look at potential data sources, you may realize that measurement for one sub question may be particularly difficult and so you reframe or push off that question for investigation at a later time. I encourage you to front end your question development but be open to an iterative approach throughout the entire process.

5.10.3 Drawing It Out

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO7: Imagine, plan, and design a data science project.

Let's talk about how to take stock of what you know and how to define the problem space. You could think of this in terms of ideation, or I like to think about it in terms of drawing it out. And another way to say drawing it out is to say that we're going to use a word web.

Now, this stage might occur before or after you come up with a question. Use a whiteboard or a chalkboard or a large piece of paper to do the following. Put the problem statement or question that you've developed in the middle of the page and circle it. Then draw lines out to smaller circles where you will write all the related concepts you can think about. As you think of concepts, you'll continue to draw lines from the center circle, and you'll probably draw lines from circles on the periphery to circles that are even further out.

And the final result is a really messy page with a bunch of linked concepts. But that's the goal. The goal is to take stock of what you know and to better understand the problem space. If you do this before you define a question, it might help you figure out which slice of the problem you want to tackle. If you do this after you define a question, this may help you figure out what some questions you want to explore and may help you think about how different variables are related.

5.10.4 Operationalization

This element addresses the following learning objectives of this course:

- LO2: Design and apply research questions.
- LO7: Imagine, plan, and design a data science project.

So we've presented these steps as following some kind of temporal order. First you think about the problem, than you think about how to formulate a question, then you think about theory, et cetera, et cetera. But these things don't always occur in order. Elsewhere, we'll talk specifically about operationalization in more detail. But I want to briefly address it here.

In brief, operationalization is how we will measure a concept. And so as you develop a problem statement, as you come up with a problem definition, and as you draw things out, keep your mind open for ways to measure concepts.Now, be careful because you might unintentionally narrow down your thinking because you might say, I don't know how the heck we're going to measure that concept. So if you'd rather wait to think about operationalization a little bit later in the data or measurement phase, that's fine. That totally makes sense. But I encourage you to keep an open mind because you might surprise yourself and think of clever ways to measure concepts as you develop the early stages of your project.

5.11 Emission Credits Example

Spend eight minutes on the following prompt:
You work in the manufacturing industry. Your boss comes to you and says you need to figure out if the company should keep polluting at the same levels and buy emission credits, or pollute less at the risk of reducing output. In other words, in the long term, what is "better"?
You identify the magnitude of the decision and want to encourage your boss to step back and think about the problem systematically.
Use the following framework to think about the problem. Fill in the details:

- Problem statement:
- Question:
- Conceptualization: How might you conceptualize what your boss means when they say "better"?
- Operationalization: How might you operationalize some of the inputs and outputs of this problem?

5.12 Use of Theory

5.12.1 How Concepts Are Related

This element addresses the following learning objective of this course:

- LO4: Justify an analytic approach that informs decision making.

Often we think of theory as something that's exclusively for academics. That is, academics sit in their offices and come up with theories that are abstract and have no place in the real world. That's partially true. But let me try to convince you that theory has a place in our work.

Theory is simply our story about how concepts are related. It's an articulation of your prior expectations before you embark on a project. This can be based on academic literature, best practices in your industry, historical data, or something else. It's important to explicitly set out your expectations because it will guide you in various aspects of your project.

For example, if I work in the airline industry, and I'm on a team that models airline ticket prices, my baseline theory is that during the end-of-the-year holiday season, people's willingness to pay is higher than in other months because of family-related travel. This theory is informed by many years of historical data.

Perhaps my company wants to then implement a new baggage policy. And they asked me if they should implement itin winter. My expectation because of the theory that I've articulated is that people's willingness to pay is higher during the winter months and that, while people's behavior in those later months may not generalize to the rest of the year,customers should be willing to pay more in light of the new policy. Management might ask you, so should we go forit? And hopefully, your reply is something like, let me do a little bit of analysis first.

5.12.2 How Variables Are Related

This element addresses the following learning objective of this course:

- LO4: Justify an analytic approach that informs decision making.

So first we start with a theory about how concepts are related. And then the next step is to move to think about how variables are related. And elsewhere, we'll spend more time on conceptualization and operationalization. But let's go through an example of how to think through the relationship between variables.

Let's imagine you work for an airline. And you want to predict how customers will react to a new change in pricing.You believe that during winter months, people value spending time with family and have a higher willingness to pay.This is your theory part. At the concept level, we can say the following. In winter months, people will be more willing to pay-- excuse me. People will be willing to pay more for airline tickets and accompanying fees than in non winter months..

But let's dig a little bit more narrow. Let's go to the variable level. Let's operationalize winter months as November andDecember. And let's be more precise with how we think about airline ticket prices. And let's focus on just baggage fees. Now we can make the following hypothesis about the variables. In November and December, passengers are willing to pay more for baggage fees than in other months.

So we'll talk more about hypotheses in another video. But I want to point out that in this video, we dipped our feet in conceptualization, operationalization, and hypothesis formation. Well done, team.

5.12.3 Describe vs. Explain

This element addresses the following learning objective of this course:

- LO4: Justify an analytic approach that informs decision making.

As you begin to develop your theory or your story about the relationship between concepts and the project, ask yourself do I want to describe? Or do I want to explain? If you care about what is going on, you often will be engaged in what we call descriptive research. If you care about why it is going on, you're probably engaged in exploratory research.

Now, what does any of this matter? The project you design to describe what's going on, i.e. descriptive research, is probably going to look a little bit different or a lot of bit different than a study that you design to explain why it's going on, i.e. exploratory research. So let's go through an example.

Let's imagine we work in the human resources department of an organization. And they want to improve the exit interview process so we could try to minimize turnover. If we do a descriptive project, we might seek to better understand what the current exit interview process looks like and how it's changed over the years. We might describe some of the information human resources gathers during the interview process.

If instead we do an explanatory project, we might instead design a project to understand why people leave the company. Depending on the format of the information HR gathers, we might try to uncover patterns to explain why people leave.

Now, depending on the level of existing work in a space, even if we want to do work that's only explanatory, that is,we want to explain why things are going on, the reality is that we might first have to do a little bit of descriptive work to understand the context of the space. Now, this example that I just laid out offers a great opportunity to talk a little bit about selection on the dependent variable. So let's go into it.

So one major issue with the study as described is that it suffers from selection on the dependent variable. If someone ends up in the interview process, they've already had enough of the company. And for whatever reason, they plan to exit. Simply put, we cannot determine what causes people to leave the company by only looking at individuals who leave the company.

A different, more appropriate design would be to gather data from both people who leave and people who stay with the company to determine what the difference is between those who stay and those who leave. Said differently, in order to determine what leads to a success, i.e. a long career with the company, and what leads to failure, i.e. a quick departure from the company, we have to look at both successes and failures.

5.13 Hypotheses

5.13.1 Theory and Hypotheses

This element addresses the following learning objective of this course:

- LO2: Design and apply research questions.

A theory is your expectation about how concepts are related. For example, I think that people who have higher salaries will be less likely to leave a company. Now, you can imagine a scenario where the opposite is true. Maybe the highest paid individuals are the most desirable across companies, but just bear with me. A hypothesis is a testable, falsifiable statement about the relationship between variables.

Now, it's really important that to meet the scientific standard, we must ensure that our hypotheses are falsifiable. So a falsifiable hypothesis is people who have higher salaries are less likely to leave a company than those with lower salaries. The core idea of falsifiability is that it is possible to collect evidence to show that the hypothesis is false.

A non falsifiable hypothesis is it is wrong for people to leave a company because of their salary. We could engage in a philosophical or ethical debate here, but we can't demonstrate that this statement is false.

So to make this discussion a little bit more concrete, let's briefly review the difference between independent and dependent variables. The dependent variable is also known as the outcome of interest. It's the outcome you care to explain. For example, if I cared to explain employee turnover, the dependent variable may be the number of years one is with a company.

Independent variables are variables that influence or, in some cases, cause the dependent variable. Again, if you want to explain why people leave a company, one of your independent variables is salary. Now, this is a simplification of the relationship between independent and dependent variables. In other classes, you'll complicate this relationship.

5.13.2 Types of Hypotheses

This element addresses the following learning objective of this course:

- LO2: Design and apply research questions.

The null hypothesis is a statement that there is no relationship between variables. For example, there is no relationship between salary and one's decision to leave a company. The alternative hypothesis is a prediction about the relationship between variables. And you can have either a directional or non-directional hypothesis. A directional hypothesis is the greater one's salary, the less likely one will leave a company. A non-directional hypothesis is those with higher salaries will have a different likelihood of leaving a company compared to those with lower salaries.

The more precise you can be with your hypothesis, the better. If you don't have prior expectations that are relatively strong, you might formulate a non-directional hypothesis. But if you do have strong priors, focus on directional hypothesis.

5.13.3 Introduction to Hypotheses Testing

This element addresses the following learning objective of this course:

- LO2: Design and apply research questions.

Let me step back real quick and acknowledge that what many of you may do is not necessarily hypothesis testing.You might operate in a world of prediction. And maybe you're agnostic about the theory or the mechanism. As long as you can predict an outcome, maybe you don't care why.

And that's totally fine. And I don't mean to place any kind of normative value on hypothesis testing or on prediction.But let's just take a few minutes to think about how we test hypotheses. Let's think about employee satisfaction in a company.

Let's think about the following null hypothesis. There is no relationship between salary and employee satisfaction. An alternative directional hypothesis is employees with higher salaries will be more satisfied than employees with lower salaries.

Now let's take the Cassie Kozyrkov perspective on this. She's a major leader in decision intelligence. And the way she thinks about hypotheses is as follows. Quote, "Does the evidence that we collected make our null hypothesis look ridiculous?" End quote. Now, you'll go into much more detail in other classes about rejecting the null or failing to reject the null hypothesis. But let's think here about hypothesis testing in the following way.

Remember, our null hypothesis is there is no relationship between salary and employee satisfaction. The alternative hypothesis is employees with higher salaries would be more satisfied than employees with lower salaries.

If we survey employees and find that there is a relationship between one's salary and one's satisfaction at the workplace, then we can reject the null because we've collected evidence that in the words of Cassie Kozyrkov made our null hypothesis look ridiculous. As a result, depending on the task we were given, maybe we give people a pay raise to help increase employee morale.

Now, to be very clear, we did not prove the alternative hypothesis. We found evidence in support of the hypothesis.We have to remember that the same set of evidence may be consistent with other hypotheses. For example, it might not be salary that influences employee satisfaction. But it might be that employees that have higher salaries have been with the company longer. And they feel more stable in their career. And they're less stressed about their position in the organization.

It might be a perceived sense of stability one has with a company that experiences-- excuse me-- that influences employee satisfaction, not salary. But importantly, we didn't test that hypothesis in the current setup. And so I encourage you, as you design your projects, think about alternative explanations to the relationships you observe.