Week 2 Transcript--Where Science Meets Organizations (updated June 2021)
If you require closed captioning for accommodation reasons, please contact the instructor and we will make the necessary accommodations.

2.1.1 **The Bigger the Decision, the More Important Data Is**

So we're going to take a quick walk in this segment through, actually, a history of many decades, of sort of what's called the modern schools of thought, that it revolved about a kind of high-level relationship between data and decisions.

Now, you might look at that and say, I don't know. Wait a minute. Isn't that relationship obvious? What do we have to talk about here? Shouldn't all decisions really be driven by data? Well, how else would you do it? And actually, for aspiring data scientists that's a very natural way to think about the world, but let's just remember that it is not a natural way for many others to think about the world.

I mean, just consider your favorite-- I don't know-- fashion designer, an intuitive genius, Donna Karan or whoever. For them, the notion that you would use data to define your decisions about what products to build might actually be incredibly counterintuitive and even counterproductive. Imagine talking to Mick Jagger of the Rolling Stones and showing them a data set on what kind of music people like. It just-- it wouldn't make sense to people like that.

So maybe we need to start with a more precise and conditional statement that would be more agreeable to more people across different dimensions and people who are both data scientists and not data scientists. So let's try it this way-- what if we were to say as a starting point, the bigger the decision you're making, the more important it is that we have and use good data to drive that decision. So let me hit that statement again, because we want to deconstruct it a little bit. The bigger the decision you have to make, the more important it is that we have and use good data to drive that decision. And as I said, what I want to do next is just take a minute or two to deconstruct that statement in a more precise way.

2.1.2. **Introduction to Tactical and Strategic Decisions**

So again, let's deconstruct this simple statement, the bigger the decision, the more important its foundation in data. First of all, ask the question, what is a big decision? Well, what does big mean in that phrase? Well, there's a lot of ways to think about that question. Sometimes actually, you don't know in advance. And so here's a not very subtle but relevant example these days.

Who a major country elects as a president probably is a big decision. What color clothes the candidate wears during a televised debate, probably generally not thought of as a big decision. But part of the beauty of the discipline of data science is that it's very good at collapsing those apparent distinctions, or distinctions that sort of appear intuitive.

So what if our political data scientists, for example, do a massive image processing experiment and discover that there's a significant correlation between, I don't know, yellow shirts and victory? And that's regardless of candidate gender. Then what color shirt you tell your candidate to wear suddenly would seem to become a very big decision, not a little decision.

In fact, I think in retrospect, when you kind of sort of wipe away the value judgments, it's very hard to tell what's big and little. In most cases, those phrases are really signals about our ignorance about correlations and causal inference. I mean, in the example I just gave, it's probably just correlation, that yellow shirt and victory with some other more fundamental causation that we haven't yet discovered.

But maybe that also doesn't matter very much. I mean, if I can win elections on a regular basis by making decision X, whether that be some huge strategic statement or whether it be some small tactical decision about what kind of shirt I want someone to wear, it's probably a big decision regardless of whether I know why I'm making it.

So maybe a more pragmatic way to think about it is to talk about tactical and strategic decisions. In the next segment, my colleague Andy Brooks is going to introduce exactly that kind of distinction, and I'll let him tell you all about it.

But at the simplest level, I want to talk about it and highlight that distinction as actually being not so much in the eyes of the data scientist as in the eyes of the decider. In other words, that distinction between what's tactical and what's strategic is in the eyes of the person who has to make the decision. It's not some big truth claim or epistemological claim about actual importance. It's kind of a perceptual claim about what the person or organization making the decision thinks about that particular decision.

In other words, to be very concrete, in that person or that organization's eyes at that moment, is this a tactical decision or is it a strategic decision? And that distinction is going to matter to us, because it matters to the decider. And it matters to decider because it's going to help to determine how that person views data and data science as an input.

So let's go ahead and listen to Andy for a bit and make sure we come back to this conversation later. He's going to start in this segment by talking about some of the purposes to which we assign data, specifically a distinction he's going to introduce between understanding, deciding, and controlling, and then go into a bit of that history in which those purposes have been met in particular ways.

2.2 Decisions You Will Encounter, Part 2

So far in this lecture, we've talked about the realities and aspirations and possibilities of working with big data. These are essentially just more elaborate versions of the aspirations and possibilities of working with small data, but now paired with reason. When thinking about these realities, aspirations, and possibilities, we like to think of them along a continuum, from understanding to deciding to control. We're going to step through those.

First, with understanding, that's when we use data, whether small or big, to better know and to better understand something. That could be something about ourselves, perhaps health and fitness. Maybe we log our food, we track our blood pressure, or like with myself, I wear a Fitbit, which tracks my steps during the day. It could be using data just trying to understand things about our car, how many miles we drive, which we track by our car's odometer. Those are understanding things.

And the third one is we try and use data, whether small or big, to understand our larger environment. What are the things going on around me? Following with the car example, we track data to be able to figure out, how long is my commute going to be today? How many miles is it to go to a place? In each of these instances, we're using some amount of data, whether small or big, to better know and to understand things about ourselves, products, or the environment around us.

Second, we look to use data to help us decide. Whether small or big data, it's about ourselves, again, our things, or our environment. Such as with ourselves, we gather data to understand how we're feeling, but then we actually use it to help us make decisions.

Based on the data, where I've walked, things like that, what should I eat for breakfast? Should I have oatmeal? Or can I have cookies today? Should I exercise? If so, how much, and where? What should I do to what level?

We make decisions about our car. Based on the number of miles we've driven, is it time to change the oil or perhaps rotate our tires? And we use it to make decisions about our environment, again following the model of traffic. If I'm going to work, should I walk? Should I ride a bike? Should I take the bus? Or maybe should I drive? Or if I'm already in the car, should I change my commute perhaps to another highway, based on the data that I now have available to me?

We also use data, both small and big, to control things. And a lot of this is going into the future, and, say, controlling again things about ourselves, our health or fitness. Imagine that maybe there's a credit card that can't be used for certain types of food. Or maybe you get something available, say a reward, after you meet a goal. Those things are controlling your behavior as well as your interactions.

Or perhaps it can even control a car. Think of anti-lock braking, cruise control, your transmission. All of those things, computer systems are gathering data, analyzing it, and using it to control a physical object in the world.

As well, we can also use small and big data to help control our environment. Imagine traffic, right? We know that folks with Google doing their self-driving car are actually thinking of this. How do we gather data to understand the world, make decisions about where to go? But eventually what they want to do is actually create cars that drive amongst themselves.

When thinking about these things, we need to be really careful. There isn't necessarily a direct relationship between data and each one of these. Now, we can understand tornadoes to some degree, but we can't really decide or control them, so to speak, as much as we can. I mean, we can try to manipulate the environment and things like that to control weather and to make it rain, but we're not really controlling or making decisions for them.

We can decide things without really understanding them or being able to control things, perhaps love or even cats or any animals like that. And last, we can control things without understanding them. And that could be perhaps by design, a normal person, say, using a mobile phone, which is incredibly complex, or like I described earlier, the automatic transmission. You can control the car, but do you really understand how it works? But you can also control things perhaps by ignorance in, say, depression through medication. You just don't know about it.

In this course, we're primarily interested in understanding and using data to decide things. And that's purely pragmatic. We think that over the next decade, data scientists like yourself are going to be called on to principally help organizations make better decisions.

Now, when you first go into an organization, you'll often need to use data to first understand things. And some decisions will be about enhancing control with data. But the primary focus from day to day is using data to make decisions.

We've talked a lot about decisions so far in this class, and we will throughout the entire semester. The big thing now is we really want to look at what kinds of decisions are we talking about. What we're talking about is the really, really big ones. The way we describe it are those once in a blue moon rare opportunities to make a significant strategic difference in an organization or a market or for an individual. They really cause disruptions, and they change the future. And oftentimes they're incredibly risky.

Let's go through a couple of what these examples might be. The first one, imagine you're the automaker Peugeot, who's French. And they're thinking about entering the US market, which is highly competitive. Do they go all hybrids? That's a big decision for a company, let alone just entering the market. What kind of cars do they offer here?

Or imagine your Southwest Airlines. Well, right now Southwest has a fleet of entirely Boeing 737 aircraft. Those aircraft are starting to age. It's a strategic decision then to keep those aircraft or perhaps go to a mixed flight-- or excuse me, a mixed fleet, so using Boeing, Airbus, and different sorts of things.

Or imagine you're Nike, sports apparel, a retail company. Do you go out and offer an electronic fitness device, something similar to, say, maybe what Fitbit is doing? It's very different for a company like Nike. Or if you're Microsoft, which is based up in Redmond, Washington, with employees around the world, do you instead maybe decide you let employees work from home all the time? Perhaps that could be a competitive advantage for Microsoft.

Or you're the US government, particularly now. How does the US intervene overseas? Humanitarian efforts, military efforts, those are big decisions that are facing organizations.

We'll go through a couple more here. Turning now to health care, imagine you're UnitedHealth, and you're expanding, say, into a new region or a new state. Do you offer neighborhood medical clinics, which are smaller? Or do you offer big central hospitals, which is more of the classic model that you have? Has ramifications for your organization.

Or one I like to look at too is professional sports. The New York Yankees have been known for years for signing big long-term contracts, Alex Rodriguez, Derek Jeter. What if you stopped doing that? What would be the implications of that for an organization? Perhaps that's something they're going to do going forward.

Or with recently in the press, Nokia. Nokia for years has been making mobile phones, mobile devices, telecommunications infrastructure. Well, they now sold it. They sold their mobile device business. That changes Nokia's outlook and what they're going to do.

Or imagine Chevron, a local company here to the Berkeley campus, currently based primarily in petroleum exploration and refining. What if they expanded their solar research, taking up what British Petroleum, who was doing a fair amount of solar research, is no longer doing? They publicly announced that they're no longer doing it. What if you're Chevron? Can you capitalize on that opportunity and make a big difference?

In closing things up, imagine, let's say, you're a principal or a municipality. I'm thinking of, say, Detroit. What does the city and county or the area of Detroit do now? What is a big decision facing them? Do they redevelop certain areas? If so, how do they do that? Or perhaps do they abandon things?

There's so many questions that we can look at. And throughout this lecture, it'd be great if you start thinking about some of those things, because we may query you in our live session to talk more about them.

Wrapping up, what we want to remind is that when working with data, there's really a continuum, beginning to use data, whether small or big, to understand things, to make decisions about things, then as well to control things, which is often much more in the future.

For a given task and your role in an organization going forward, we want you to think about where are you on that continuum. Often when starting out, you're going to be using data to understand things. And essentially, through understanding things, you can help make better decisions about it. And then through making better decisions about things, you can better control those things. But remember, some things are really uncontrollable, like cats.

But much of your work is going to be about decisions. And these are really ultimately large, big strategic decisions. And these are rare opportunities to change the future. And it's an incredible responsibility

2.3 Decisions You Will Encounter, Part 3

The idea of using data to aid decision-making isn't new. It goes back centuries. Imagine you're a trader centuries ago in search of goods to buy and sell, let's say at your local market. You have an incredible number of decisions to make. First, what's available? So you ask around. You see in the marketplace. Perhaps you talk to other traders and so forth, and you hear rumor of products available in new places.

Where do you go get those things? Well, you go get maps. And through those maps, maybe you go sail to the New World, explore new places, but also choose what's actually going to sell, what's available, where to find it. But what's going to sell? What has sold? How much money have you made? You track those goods through ledgers, through notes, or even way back when, like in an abacus.

You're also interested in how do people actually use your goods, something you may not be familiar with. The data here-- essentially what people want, where you can get it, where you can sell it, and for how much-- help you make decisions about your business as a trader.

Years ago, that this was done by hand, and it was often just kept in your head. With computers though, now we call this business intelligence. It's a term of art for systems that help transform data into insights which enable better decisions. And this transformation comes through information technology.

First, we have sensors that, say, gather data. We have places to then store that data. We have ways to query the data, and then ways to visualize the results of those queries, perhaps in a report, or increasingly through like a data visualization.

Coming with this term of business intelligence, it's been around for well over 50 years. Hans Luhn was a researcher at IBM back in the '50s who was looking at how businesses could use technology. IBM had been around for years, primarily in the business technology space. And he came up with this concept of a business intelligence system and wrote an incredibly influential paper about it in 1958.

Let's go through a couple of the terms and what he means by business and intelligence. First, business is a collection of activities carried on for whatever purpose, be it science, technology, commerce, industry, , law, government, and defense, and others. Pretty broad definition of what business is.

But then second he looked at, what is intelligence? It's the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal. Sounds like decision-making.

He continues in his paper, "The objective of the system is to supply suitable information to support specific activities carried out by individuals, groups, departments, divisions, or even larger units." So this was incredibly radical at the time. Now when we read this well into the 21st century, it sounds really basic. But this was incredibly radical because the technology did not exist. Like the concepts for this did exist for years, often through paper notebooks and things like that, very rudimentary computer databases, and so forth. But it just didn't exist, and it really changed the future of computer technology, and particularly information technology.

But it took a really, really long time to do that. It took nearly 40 years, still today, to realize this vision in a practical form. Let's trace through just some of that history of business intelligence systems from about 1958 and on. Essentially in the early 1960s, the US government and big Fortune 100 type companies were really the only ones doing business intelligence as Luhn defined at that time.

That primarily was technology was incredibly expensive. So you had to plan ahead. It wasn't just you could go buy a cheap computer and go run these sorts of things. That was a big strategic decision alone for an organization to buy a computer, to have access to computer technology. Often what companies would do then in the early '60s is start with the question itself and then find the data they needed-- which, again, was incredibly difficult for organizations to do-- that then could be used to answer that question.

The process that they would do is go out and gather that data and then feed it into a mainframe computer. They'd write a query with punch cards. They'd run the query. And then they get the results later, oftentimes the next day or several days later.

They'd then have a report, such as maybe items sold last month, which then they would use to help make decisions. But remember, they were using data from then, perhaps days ago, weeks ago, months ago, to make decisions about today and the future.

Key thing to look at with those is, like I mentioned, data was not anywhere near real time. There was incredible latency from gathering that data into reporting it. And there were very few to no models running on the data itself. Those had to be created by the organization and then converted into algorithms, which they then had to try and program into the computer.

Essentially then, insights that you could glean from the data were limited by the complexity of what was asked. Pretty rudimentary due to the data and technical limits and the costs of just running these sorts of queries.

What were the type of questions that they were asking, say, in the late '50s, early 1960s? Well, let's imagine we're Ford Motor Company, right? The questions that they might be trying to ask in a business intelligence space would be, how are Ford Falcon sales changing from month to month and will they increase? Which combinations of options are most often purchased on our cars? Is it these newfangled seat belts? Is it new radios, different colors, convertibles, those sorts of things?

Based on that, how should we schedule production? Should we schedule production of a lot of convertibles, a lot of hardtops, lot of trucks? As well as how do we optimize our supply chain to meet that demand, all the parts that go into our car, doors, windows, tires, carburetors, things like that.

These are the types of questions that we're looking at, and it's essentially because that's what they actually had data about. So it was production planning, supply chain, inventory, and sales forecasting. You note that very little to none of it was actually about users themselves, let's say, how are people using Ford Falcon cars or anything. That's primarily because they didn't really have any data about it itself. So they were limited just by what data was available, as well as the technology.

Continuing then into the 1970s and the 1980s, we start to move from mainframe-oriented technology to more relational databases. And again, IBM was a key player in this, as well as Oracle and a whole bunch of other companies that no longer exist, because it was highly dynamic and changing.

As organizations, they might take in new data and load it every day. They would store it, say, in a data warehouse, which could be then created to hold lots and lots of data, well more than just the memory that could be stored in the computer at the time.

As well as now users, say business analysts, could design their own queries. But then again, they'd still have to often be run by someone who had direct access to the system itself. And again, the queries would often take a couple hours, maybe not overnight or a couple of days, but nowhere near real time in their analysis.

And this continued on well into the 1980s, things always kind of speeding up. Then in the 1990s and the 2000s, technology caught up, but it was still really, really expensive. Everything was bigger. Everything was faster. So the ability to gather data, to analyze it, and report, all was in much closer to real time, near real time, nowhere near actual time.

We've seen an evolution now into decision-making with less of a focus on tech, because the technology costs have fallen. So organizations now can go back to, what are we ultimately trying to do? Part of the reason why tech costs have fallen is there's an entire community of organizations helping companies do this work. Essentially they can outsource the technology maintaining these systems and so forth. So organizations can now just think about what are the critical strategic decisions that they're facing, not worry about all the operations of making it work.

But essentially, the organizations have the same goals as earlier. A lot of the questions we're focusing on, what kind of insights can help me with my supply chain, inventory, and production planning? Those sorts of things.

But now we fast forward to today. And really the technology limit is fading away. Technology is almost a commodity, in that if you want to do this sort of work, work with large amounts of data, run queries, you can use Amazon Web Services. You can go to these other sorts of places.

And now you have data about all sorts of things, not necessarily just what doors should I put in my car, how many window parts should I order, those sorts of things. But you actually have data about people and your users. The mobile devices that we're wearing, the cars that we have, and so forth are gathering data that now companies are just beginning to start analyzing to get insights.

So that's now part of business intelligence as well. And with that, they're starting to be able to look at really much bigger questions, so beyond supply chain and operations to what users and products and services they should assemble, what they should offer, now as well as well into the future.

And business intelligence, like I mentioned, has now been reframed from just really looking at operations and product side of things to looking at how individuals interact with products and services, and how does that fit into the entire, say, lifecycle or ecosystem of an organization? And, of course, things are much now closer to real time. We can gather data off of our devices, off of our vehicles, and things like that, run queries, analyze it, and get the results much closer to real time.

Also with today, things are increasingly automated with models and algorithms. So we can create a model of something, design an algorithm. Let's say, if x is seen in the data, do y automatically. You can do that with pricing. You can do that with inventory. Lots of logistics-oriented sorts of things. Not so much with users quite yet, but they're getting there.

Good example of this is, say, at amazon.com, right? So if you watch prices fluctuate for the top books on amazon.com, that's pretty much done by an algorithm. There's not really one individual pricing person going in there and raising the price of whatever hot book it is up and down by a few cents. That's an algorithm doing those decisions for you.

Also with business intelligence today, we're starting to see the rise of much more self-service-oriented technologies. Key phrase with this is really dashboards, just like on a vehicle, right? So it presents information to you as the decision-maker. So you can go out and create your own queries, run your own reports, and do your own visualizations. And the companies that offer that are ones like NetSuite and Salesforce.

Looking back on the past 50 years, some of the strengths of business intelligence has been this steady progression from questions to technology, from technology back to questions, and now back to questions entirely. It's much more closer to real-time analysis of deep, nuanced data.

But also there's been some weaknesses in business intelligence. There's always been a lot of promise. And only until recently has technology caught up, such that we can do these things. And business intelligence as a phrase, kind of like Web 2.0 or any of those kinds of phrases, becomes a catchall, and it means a lot of things to a lot of different people. But now we have the opportunity to move on to the really harder stuff, the complex real-world modeling algorithms and decisions.

To now wrap up, the fundamental needs of business haven't really changed. Recall the trader that I mentioned earlier. They're interested in what's going on the environment. What should I do? I need data. I need information to help make decisions. That's still true, if not even more so.

Business intelligence for the past 50 years has promised to do this. And now with increasing volumes of data and models, it's closer to being able to do that. Fortunately, the technology challenges are now fading away, such that companies can now work on the questions that they have and shift their focus from just operations to more broader user-side questions.

But still the ultimate focus of business intelligence is what's the decision? What is an organization trying to do? And we're much now closer to Luhn's original idea, but we're not quite there. And we'll possibly get that in our lifetimes.

2.4.1 Business Intelligence, Data Science, and the Strategic Nature of Humans

OK. So I want to wrap up this part of the week with a couple of key takeaways and reflections that at least I take away from Andy's run through the history. I really have four points.

So the first point that I want to hit is that, look. If you kind of think about the way in which the sort of bases of data and decision making have run through the history of business intelligence, there's been a kind of stability. What's really changed is the technology. And as Andy, I think, pointed to, the technology has changed enough in the last decade or so to really make a difference in what is possible. And likely, that's going to continue going forward. So that's the first point I wanted to make.

Second point is that I think it's the case that intelligence in business intelligence is still the interrelationship of facts that guide action toward a desired goal. In other words, facts that guide action towards a goal that we've already put our fingers on. But increasingly, intelligence is also becoming about the identification of the goal itself. What is it that we should be aiming to do, or what problem should we be aiming to solve?

That's actually a really big change. And I think if there's any characteristic that robustly distinguishes big strategy from small tactics, that's probably it. In other words, we're actually focusing not just on how do we get to the goal, but what is the right goal we should be getting to? And I'll have a little bit more to say about that later on in the week.

The third point I want to hit is that questions, the questions that get posed have really long been limited by people's imagination as much as it's been limited by the data that was available to them. And so it feels like it was often the case that the questions people ask were sometimes a function of what data was becoming available.

Now cynics sometimes refer to this-- you've heard the phrase, the drunk looking under the lamppost for her keys, because that's where the light is. That's a cynical view. A pragmatist might look at that a little bit more gently and recognize that, you know, it's just really hard for people to get serious about focusing their attention on questions that they can't yet see a way to answer.

But here's the trick. Keeping that vector open is going to become more and more important as data science progresses. As that technology moves forward and as new questions become answerable, you want people looking like a little bit ahead of where the light is, not where the light was yesterday. And that's going to be an important trick as well.

And then finally, look. I think thinking over this history, it's fair to say that data about people is where the biggest changes have probably come from and probably will continue to come in the future. And in saying that, I don't mean to downplay things like the industrial internet or the proliferation of physical sensors on, I don't know, roads, airplane engines, machine tools.

What I do mean to do is to highlight the interesting and multidimensional complexities of human behavior, you know, whether those humans have to be your teammates or your employees or your customers or your boss or whatever. And most importantly, perhaps, they're a strategic nature. Because there's one element-- look, people are complex, and they're just as complex as a fatigue function for a piece of high tensile steel in an airplane engine, or anything like that.

But they're complex at another level which is that, of course, they're strategic. They may be trying to fool you into thinking they are something they aren't, or are doing something they're not doing. And a piece of steel is never going to try to fool you in that way. It's never going to try to outwit its sensors, and people will always do that.

Of course, you know, that's just something to be aware of. It's nothing to be afraid of. And in fact, it's the beauty of a human-centered data science discipline. But it's also a reason why human-centered data science and business intelligence have a very long way to go. Again, because ultimately, they're going to have to contend with the strategic behavior of people.

2.4.2 Ethics of Different Types of Decisions

Ethics are not something that we only explore at the end of the project. It's not just is this ethical or not? Let's move on. In fact, ethics come into play really early in the design process. We should consider the ethics of the questions we ask and the types of decisions we ultimately want to influence.

So first, let's look at ethics in terms of the magnitude of decisions we make, that is, tactical and strategic decisions. And then next, we'll review how we can use data science to understand, decide, and control things in the universe around us.

So let's start off with a review of tactical versus strategic decisions. Tactical decisions are short-term specific activities that want to achieve a particular objective. So for example, let's find a better way to organize data on our company servers. A strategic decision is more a long-term, future-oriented, high-level, really organizational level goals. An example could be should we move away from physical servers to the cloud.

And the ethics vary depending on the type of decision tactical versus strategic in part because strategic decisions become more of a core part of an organization's identity. And the process of reversing course after a strategic decision has been made takes a lot longer than the process of reversing course after a tactical decision. So let's go through an example.

On the one hand, based on the analytics of labor needs and analysis of the labor needs in the industry, a company might make a tactical decision to reduce employee hours during the slow season. This may temporarily place strain on employees, but the hours may be made up during peak seasons.

On the other hand, if a company wants to make a strategic decision to shift how it hires employees and rely almost exclusively on temporary or contract workers, the impact on workers and on the industry may be far reaching.

So now let's shift to think about how we can use data science to help us understand, decide, or control things around us. Let's put on the table the kind of null hypothesis that the ethical stakes go up and potentially get higher as you move along the continuum from understand to decide to control.

Now, it'd be nice if an organization can make decisions to understand that were value neutral. And maybe you didn't have to worry too much about ethics. And it'd be nice if decisions to understand were always less ethically charged than attempts to control. But I think you'll find that this is very rarely going to be true. Let's work through an example.

Let's imagine that you work for a company that created an application to monitor sleep. Let's say you first want to understand how something works. And you find out that the more hours a teenage kid sleeps, the happier they are. Next, you want to engage in decision making. Based on the data you collect, should we encourage parents to move dinner earlier and impose an earlier curfew?

Finally, we may want to control behavior. Should the company monitor sleep and reward and punish the user based on how much sleep they're getting? I think this example illustrates that as you move from understand to decide to control, the ethical stakes do in fact increase.

2.6. **Problem Types Overview**

OK, I want to start this segment by reminding everyone of a famous speech that former Secretary of Defense Donald Rumsfeld gave famously around 2003, I think it was, in which he talked about a phrase that now everybody knows, the known knowns, the known unknowns, and the unknown unknowns.

And he was given a lot of, shall we say, feedback around that speech, much of which was very critical. At the same point, that distinction got a lot of attention. And it's a good one. It's actually really important.

But for the moment, I want to focus our attention on what I think might be a more important typology of problems for data science. Because I'm going to submit for the moment that most of the work that we're going to be doing is in the area of the known unknowns.

The typology that we're going to talk about is actually more about depth of meaning. And interestingly, I think it provides more freedom of action. So we're going to introduce a simple typology of data science problems that appear in the world, and we're going to call them type one, type two, and type three problems.

And it's kind of a conceptual scheme that we're going to use or experiment with using to sort of sort through problems and match with the right level of thinking about solutions. But I think most importantly and most practically, it's a scheme that will differentiate for you the kinds of demands that the world is going to place on you as a data scientist.

Now I should just acknowledge the relationship not only between the kind of known unknowns distinction, but also to a phrase that lots of people use these days, the phrase wicked problems, which many of you have probably heard. It actually goes back to 1973. It was created by environmental scientists to describe complex sociotechnical systems where problems have no definitive formulation. And the literature on wicked problems basically adds up to a statement that the formulation of the problem is itself the problem.

And many of us have to deal with wicked problems. Many of us want to deal with wicked problems, because in some respects, they're the most interesting ones in the world. The proposition behind this next distinction, this sort one, two, three, or type one, two, three problem set is that this typology can often be used to help manage how we think about wicked problems and what we do with them. So let's take a look at that more closely, and then in synchronous discussion, we can discuss whether or not it actually helps, and to what degree it helps.

2.7. Type 1, 2, and 3 Problems

OK, we're talking about Type 1, Type 2, and Type 3 problems. Actually what we're really talking about here are the kinds of questions that data scientists are going to be asked to deal with in their day to day work. And we need some kind of schema for that because questions come in all different shapes, colors, textures, and flavors. And so let's sort of figure out a schema that's simple that we can work with that has only three levels.

And actually, there's a really good schema we can borrow from people who analyze English literature texts, funnily enough. They have a really simple way of thinking about it. There are Level 1 questions, where the answer is found directly in the text that we're dealing with. There are Level 2 questions, where the answer is inferred from the text. You can infer it from what's there, but it's not directly there. And then there are Level 3 questions, which go beyond the text and are actually about abstract concepts that are embedded in the text.

So, you know, let's just do a concrete example. Everybody has read Moby Dick. And if you haven't read Moby Dick, go read it. So what's a Level 1, Level 2, and Level 3 question here? Well Level 1 is very simple in the text. What color is the whale? The whale is white. Simple. The Level 2 question, what's going on? What is Ahab's real problem? What is Captain Ahab so obsessed with? Why is he so worried about the whale? That's what the book's really about, and the meaning can be inferred from the text.

But it's the Level 3 question that is really interesting, and the one that calls upon people to work hard. A Level 3 question might be, what's the nature of obsession as a human trait? How prevalent is obsession? How many people are obsessed? How much of history is explained by obsession? What else does obsession do? Those are the kind of questions that get really, really interesting.

Now let's take that over to the business context in which data science people are normally going to be working. Type 1 questions are questions where typically both the question that's being asked and the answer are going to be relatively clear. You might not have all the information you need, you might have to look for more data, but you know what you're asking and once you got the answer, you know how to explain it to people.

A Type 2 to question is often when the question might be clear, but the answer is going to be unclear or a little bit abstract. And no matter how much data you have, no matter how much information is on the plate, it's not necessarily going to be possible to create a really crisp, clear answer. And then there are the third level questions, where both the answer and the question are both unclear and abstract.

So let's do a concrete example just to make it really simple. Let's say I work for a supermarket. A typical Type 1 question. What should be the price of a pound of peaches in the supermarket today? The answer is really clear, it's going to be a number. And the question is clear. What's the price of peaches? We'll be able to figure that out if we have the right information.

A Type 2 question might be, so what's our strategy for selling fruit in the supermarket? That can involve lots of different things like where we place the fruit, how much we charge, how we light it, how much we polish it, do we have organic and non-organic? Lots of different answers could come to that question. A little bit unclear where we're going to go with that.

But then there's the Type 3 question, which is, in some sense, the most interesting and the most challenging. How about this? What are our customers actually doing when they buy fruit? What needs are they fulfilling? Maybe they're just looking for something sweet, but maybe they're looking to fulfill some hole in their hearts that was left when they were kids from not having enough sweets. That's where the really interesting things start to happen.

And here's where the rubber meets the road. If you think about how business strategy has sort of evolved over the last 40 years or so with the use of information and data. There was a time in the 1960s where the core of business strategy was really about figuring out Type 1 problems, helping a company to decide, what's the right price for a pound of peaches today? But by the 1970s, 1980s, that wasn't really a differentiating feature for companies anymore. What became the more interesting question for strategists was the sort of Type 2 question, what's our fruit strategy? And that's a kind of a question that enormous number of consultancies made lots and lots of money from over the years helping businesses figuring out what's their fruit strategy. And they used information, but fundamentally it wasn't all about information.

Today the really interesting questions that people are struggling with are the Type 3 ones. What are people doing when they buy fruit? What does that mean to them? What value can we create around that? Now all those questions still exist in business strategy and they're all moving forward. So Type 1 problems have moved from static to dynamic pricing, like maybe we need to change the price of peaches every single minute over the course of the day like airlines change the price of their tickets. Type 2 problems have moved to the issue of disruption all along the value chain. So our fruit strategy could easily be undermined by maybe a single product fruit seller. But the Type 3 is where we really want to go. What's the deeper meaning embedded in the story of buying fruit?

And here's our proposition for data scientists. If you look at what data science is doing right now today, data science is making a huge impact on Type 1 problems. Redefining what it means, for example, to price in real time. Data science is starting to make an impact on some interesting Type 2 problems, and obviously we want to grow all that. But the most ambitious objective, the thing that we're really shooting for in this course, is to bring data science to bear on the Type 3 problems. To help people inside your business or your organization figure out not just what the pricing strategy is and not just what the fruit strategy is, but to really understand the deeper meaning embedded in fruit for the customers that you're serving. And when we can bring data to bear on that, then we're making a huge impact on corporate strategy.

2.7 Problem Type Example: Bay Area Basketball Team

So one of the beautiful things about type three problems is that they are so rich in context and so interesting. And so I would ask you to think about considering working through an example of a type three problem that is of particular interest to you.

I'll just give you an example of one that's particularly interesting to me, since I live in the Bay Area, and it involves the Golden State Warriors, as you see on the slide before you. And I'm going to imagine, if it should only happen, that I have just been hired as the Chief Data Scientist for the Oakland Coliseum, and my job is to try to keep the Golden State Warriors in Oakland. Some people think that battle's already been lost here in San Francisco, but let's imagine it hasn't.

And now, I'm confronted with what is the data product I'm going to create to try to actually keep the Golden State Warriors here. And I'm using this type one, type two, type three distinction, so I'm going to think about it this way. Really the type one problem, what's the pricing algorithm for tickets? I mean, maybe we can actually get the ticket pricing sort of more optimized against demand for particular games or particular adversaries or whoever's playing that day. And that would be an interesting type one problem and a great project to do.

But there's also a type two problem that I can offer. So type two problem might be what do our fans like and not like about the experience when they come to the game? And that would involve not just, of course, the price of the ticket but all sorts of things that happen as they're in the Coliseum, maybe involving their experience getting to the Coliseum parking, BART, et cetera, et cetera.

But look, here's the really interesting question-- it's the type three question. Where does loyalty to the Warriors really lie? What does the Warriors presence at the Oakland Coliseum mean for the City of Oakland? How much should the City of Oakland spend or invest to try to keep the Warriors here? How important is that to the way people think about themselves in their identity as citizens of a city? Now, that may seem like a really big abstract question, but ultimately, think about it. If my job is to try to help you keep the Golden State Warriors here in Oakland, which of those questions do you think the answer to which would be most valuable? You're going to want the type three question, and it's also, by the way, the most fascinating and interesting one for a next generation data scientist to try to confront.

2.9.1 **The Motivations of Decision Makers**

Some of the decision-making fundamentals I'll talk about will be review to some of you. For others, they may seem pretty intuitive. And for others, they might seem new and challenging. The core takeaway is that the more we know about the motivations and incentives of those who make the decisions, the greater chance we'll have to frame communication in a way that they'll listen. Said differently, know thy audience.

What does that mean? It's important to think about the incentive structure of the decision maker. What do they care about? Do they care about profit? Public relations? The environment? Corporate responsibility? Their legacy?

Also, what's their level of technical expertise? Yes, you'll speak to a variety of audiences, some of whom are technical experts. But it's critical to know how to speak to nontechnical individuals. And it's important to translate technical explanations into substantive meaning. That is, why do we care? Why does this particular nuance matter?

Now, how do you learn all about this? How do you learn all this information about decision makers? In part, you could immerse yourself in the organization or in an industry. This comes with time but sounds like it takes a long time.

Another way is to identify a colleague in your organization who knows the organizational culture. It's important to find people who you can ask blunt questions to. Ask them, OK, I'm presenting to so-and-so in a week. Tell me what's their incentive structure? What do they care about? What makes them tick? And what's their relative level of technical expertise? Now, technical expertise and incentives are quite related.

Once you understand the technical expertise of the decision maker, then you'll have to speak to them in a way that resonates with them. You'll have to speak in a way that's mindful of their level of technical knowledge.

And you'll have to, quote unquote, "speak their language." That could literally mean that you use language and terminology that they're familiar with. But more importantly, it means speaking in a way that makes it clear to them that you understand their motivations and their incentives.

Any kind of jargon and language that requires a technical understanding can muddle the message. I understand that sometimes you want to demonstrate your expertise with technical jargon. But when possible, keep it simple.

Now, I want to admit that the way that you communicate is both industry and organization specific. But I encourage you to remember two following things. You might be tempted to save this epic punch line for the middle or the end of the presentation.

But consider the BLUF approach, B-L-U-F-- Bottom Line Upfront. Its effectiveness will depend on the industry and what kind of narrative structure they're familiar with. But all else equal, we have very little time at the beginning of a presentation to capture the audience's attention. And I recommend you hook the audience from the very beginning.

And I also want you to remember that no audience is truly a captive audience, even if they're stuck in a meeting with you. You and I both know how easy it is to get distracted. Someone might be in a room with you for 60 minutes, but that doesn't mean they have to listen.

2.9.2 **Translate Insight and Level of Certainty for Decision-Maker**

Welcome. Moving on from our discussion of strategy and problem types, we're going to have a conversation with one of my favorite recovering management consultants-- and I should say long-recovering management consultants-- Michael Dalby. Michael, thank you so much for coming in today.

Thanks, Steve. And hello to everybody out there.

Michael, tell us a little bit about your background, and how you came to this addiction and what it's been like to recover from it.

Well, in the early days, I spent a fair amount of time as a university professor. But then, learning that there may have been other places to exercise what talents I have, I actually did become a management consultant-- joined McKinsey, spent a dozen years there. Later was a corporate officer at McKesson in San Francisco. Got to know the world of big business from that perspective. And since that time, I've been slowly recovering in my own firm.

So Michael, over the course of your career as a live management consultant, you've spent a lot of time in C suites, a lot of time with these peculiar organisms called CEOs. Could you tell us a little bit about the highlights of that experience, and what you think it would be like for today's data scientists to really try to interact at that level with the CEO of a big organization? What's special?

It's a really good question. CEOs do, in fact, occupy a position that's really different from people in other places in the organization. Their tasks really typically divide into three big ones. They spend time managing and coaching people, who do much of the operational and routine activities. They spend a great deal of time communicating what the organization is doing and why it's important. And finally, perhaps 20%, 30% of their time, they work on really difficult questions of strategy and the next decade for the company.

So what this means for the data scientist, I think, is really quite profound. The data scientist, in effect, needs to earn a place at that table in order to be really effective. In turn, that means that it's not simply bringing new data to bear on a given issue or question. It's also being able to explain what the implications are of those data, and why they matter in the first place.

Some people are under the impression-- and I've certainly experienced in some of my interactions-- that CEOs have short attention spans because there's so much on their plate. How do you penetrate that and bring that really important message into that person's span of attention?

In a way that works. You're right about the CEO's attention span. And by the way, you can count on one fact, which is that it will be interrupted even as we speak. The most important thing to bear in mind is that an insight from new data or from patterns that emerge has to be salient or relevant to the larger concerns of the organization, and therefore of the CEO, in order to be noticed and taken seriously. The corollary of that-- which hardly needs saying, but here it is-- is that technical problems and techniques for their own sake are not particularly of interest to the CEO. So not doing a shazam or bringing some fun facts to bear is probably a good idea. It has to be relevant.

And we've talked sometimes about how the shazam can sometimes be almost a little bit embarrassing in that situation. Could you say a little bit more about that?

Yes, it can be, because it decreases the credibility of the people involved if they are dwelling too much on details that are not addressing the major questions on the table.

One of the things I think you probably never want to do is make the CEO feel inadequate or stupid.

I've done that, and it was a disaster.

You want to tell us a story about that?

Actually, it's a terrible story. I actually spent one time trying to explain the implications of some data to a CEO who was extremely charismatic, but shall we say, mathematically challenged. And so writing an equation in front of him was probably not the way to go. I was reminded of this by my colleagues later.

So tell us a little bit-- one of the things that you've often said to me and help me understand is that sometimes, 80% salient is more important than 100% accurate. Could you say a little bit more about that, and maybe tell us a story about how you learned that way of thinking?

The most important contribution that the new data science can make, in addition to pointing out implications of what the data are saying, is to deal, I think, realistically with probabilities or the likelihoods that one thing may happen as opposed to another, one course of action may be more successful than another, one thing is more attractive than another to either the stakeholders or the customers or both. And so in order to understand that degree of relevance, you need to spend a lot of your attention thinking about probability, likelihood, and-- as you put it-- salience. The accuracy as such is almost never in question. It's so much better than it was 20 years ago that that is not really the issue. But as always, it's the pertinence of the data and the implications for action that make all the difference.

What about overconfidence and underconfidence as it relates to the underlying data? At what point do you think about introducing that into the conversation differently as you move up the ranks of seniority inside the organization?

One way to think about this would be to say, if my result is such and such, and I'm advocating a certain course of action or insight based on that data, what would happen if I cut it in half? If the result would not change the action that would be plausible or advisable under the circumstances, that degree of accuracy is not relevant to the issue under consideration. It may be relevant to the data and the data scientific angle of the question, but it is not, strictly speaking, something that the CEO will care about.

It's really interesting. I think for many scientifically-minded people, the instinct is to double down rather than cut things in half, actually. And it's hard to switch that around when you're talking to management.

That's right. But by the same token, management has to deal all the time with degrees of uncertainty, and will naturally focus on greater rather than lesser likelihoods. They don't enjoy taking decisions and making departures just for the sake of novelty. That is not helpful.

One of the things we've talked about is the desire of the client or the customer or the person to whom we're trying to provide analysis-- their innate desire to want to have a higher degree of certainty or higher degree of confidence in an assessment of a strategic situation in the world than maybe is possible or justified. How do you deal with that innate drive to get more certainty out of the situation than you can possibly get?

You don't lie about what you know.

Ah.

In other words, if your knowledge and your insight are limited to a specific level or a specific area or a specific segment-- let's say, of the customer base-- then say it as such, and allow the natural decision processes with all their uncertainties to rise to the surface, because that's real life. Data are very important, but data are not everything for a CEO.

We took a look at a famous clip from our former Secretary of Defense, Donald Rumsfeld, talking about the known knowns and-- well, I'm not even going to try to repeat what he said, because I couldn't possibly say it with as much flair as he did. But tell us about the serious version of that in the C-suite-- the type 1, type 2, type 3 problem, or how people sort of know where they are in that particular dimensionality. Isn't that almost a prerequisite to knowing how to interact with someone, to understand where they think the question lies in that typology?

Yes. Even though Rumsfeld's use of those terms were primarily politically motivated, there is some utility in thinking in those categories for business as well. Let's take it one at a time. If you think about the known knowns-- that is, problems or situations that are already known to exist, and the question is what to do about them-- this is actually probably the bread and butter, if you like, of data science in corporate or organizational settings. And what this would involve would be, of course, monitoring or validating the ongoing data stream, establishing what the norms are and a certain set of permissible deviations and operations, and so on. This is almost never the realm of a CEO's attention.

Interesting.

The reason is that they have subordinates who take care of a lot of the operational details, and whose accountability and responsibility it is to assure the integrity of that kind of activity. So maybe 80% or something like that of data science will be at that level.

Now, if you move on to the known unknowns-- that is, categories of uncertainty that you've already recognized, such as, let's say, a competitor's behavior. You can't predict a competitor's behavior with absolute certainty. But you have an idea of where they will go given a certain set of industry conditions. You have some set of questions about whether people will, let's say, buy a new product, but you know what the fundamentals are of the growth of disposable income, the population, and so on. So in these areas, what you're searching for, I think creatively, as a data scientist would be patterns of fact, the implications of which would be important to the running of the business or change in direction of the business.

In the third case, where the CEOs real responsibilities ultimately lie, these are the unknown unknowns. And we move from a realm of even relative certainty here about the world and what it might evolve into a new set of territories.

Let's say, for example, you were trying to take a decision about whether to wear-- let's say, for example, you were trying to help a CEO take a decision about whether to invest in, say, wearable computational devices. Now, we have no idea at this point, other than a certain interest, whether this will develop into a market that will matter or not. And it could cost a lot of money if you got it wrong. The only approach to this where data science has a, I think, unique role to play would be in focusing on the likelihoods-- the, what would you have to believe, kinds of questions-- that would tend to focus on greater or lesser plausibility of a certain future taking place. So known and knowns routine; known unknowns, pretty important. unknown unknowns, this is the real frontier for the good data scientist.

Really interesting. So Michael, let's circle back to an earlier part of the discussion and talk about something you've learned from-- I think it was probably a pretty unique combination of capabilities and experiences that you've had. You started off as a historian, and historians tend to be people who really understand and are comfortable with narrative. And later, you moved into a consulting environment where you spent a lot of time working with numbers. And there were probably some instances where bringing those two elements together in a really powerful way gives you something that was a whole lot more than just the sum of 1 and 2. And I wonder if you could maybe tell us a story about where you think bringing those two elements together was most powerful in changing someone's decision.

Another good question. It reminds me a bit of a situation in which the two sides that you're mentioning of an approach to a problem did, in fact, come together-- the data or the numbers and, if you like, the story or the context. In the mid-1980s, when I was first at McKinsey, my teammates and I were actually working for a client in the North American soft drinks business. It was not Coke or Pepsi, but that's all I'm going to say about the client. They had entered Japan, and so we found ourselves in Tokyo trying to help them understand how their one product-- cleverly disguised here-- in a normal American can style, their one product, their one brand was going to fare in the Japanese soft drink market.

Now the facts of the matter were, after a certain amount of investigation, reasonably obvious. We in the team spent months working on every available fact about Japanese soft drinks, and it turns out that there were 1,500 products at market in the Japanese soft drink industry at that point. There was a very complicated distribution system. There were segments that did not exist in the normal sphere of competition as defined by North American competitors. And yet the motivation of the CEOs was in all segments of the industry the same-- persist and let brand power do the work.

Our team didn't believe it, and we did not want to encourage the CEO to make a mistake. In those days, the numbers mattered. And it was, of course, nothing like the data dashboard era of today. So when we presented the numbers, it was a big deal, and it had to be reams of paper with a lot of numbers on them. And that was meant to convince people. But in this case, the right answer-- which was to withdraw from the Japanese market-- was not something that any of the senior management wanted to hear. So we had to tell the story in a way that gave the cultural context enough room to do its work.

So get on the plane, go back to the United States, have the final progress review with the CEO. Before the CEO entered the room, we set up a table-- two tables, in fact. One table had this can, and another table had something like 200 cans that looked like this that featured items like Pocari Sweat, that introduced the Beer Shuttle.

The Beer Shuttle? This is before the Space Shuttle.

No. Here's the Space Shuttle. It was the cap on top. These products were vended from vending machines on the streets of Tokyo, and everywhere else in Japan-- including whiskey. Imagine vending whiskey in 1985 in New York. Not likely. One table with this can, another table with several hundred cans like this. CEO walks in, looks, and that was that. Progress review done. Story told. Withdrawal from the Japanese market began next week. Eventually, the brand was sold to a more capable competitor, and a lot of money was not lost.

Successful project. So Michael, thanks so much for sharing these reflections on some of these general principles and rules of thumb of working at the C-suite level. Sounds like over time, it's more likely that these things are going to remain steady, and that the data science discipline is going to have to adjust to the C-suite just as much as the C-suite is going to have to adjust to the data science discipline.

I couldn't agree more. And I think the motto of the story for the data scientists of the future is, think up a level.

Think up a level.

Think up a level.

Great. Thank you so much.

Have a beer, Steve.

[LAUGHTER]

Next Card

2.9.3 **Reframe Questions and Embrace an Iterative Design Approach**

Thank you very much, Craig Denny, former government official, now a global foresight professional in private practice in Washington, DC. Craig, you've had enormous experience working with government officials, decision makers at all different levels, on all different kinds of problems, over a very long period of time. We've been talking a little bit about problem sorting, the type one, and type two, type three, problem distinction, as that comes down from the decision maker. Can you tell us a little bit about how that manifested from your position, trying to actually serve people or answer questions?

Yeah, I'll take a couple of different stabs at it. I mean, sometimes you were working more of a, you know, a topical account or maybe a country account, and the issue might have been something about what's going to go on in the next, you know, month? I'll give you an example, you know, from the late '80s, when I was working in Latin America.

You know, how is the situation surrounding Pinochet and regime stability in Chile going to unfold, surrounding a particular event, which was a plebiscite that Pinochet had called? A yes/no, up/down vote on his tenure, if you will. And you know, a lot of complexity around that, you know, as to whether or not he would actually, you know, allow a fair vote and that type of thing. But those types of questions, you know, notwithstanding the complexity of the actors involved and having to think scenarically or about different, you know, hypothesized outcomes, it's still a little bit more containable than something like, you know, where is Japan going with respect to its grand strategy? Which is something we were dealing with when I was working in East Asia in the mid-'90s, and there was a lot of drift in the US/Japan alliance. And when Joe Nye was at the National Intelligence Council-- Joe Nye from Harvard and the famous soft power guru-- you know, there was a lot of concern at the policy level of a country that seemed to be adrift from the United States and its security relationship, a lot of economic friction and competition back in that time.

And that's a much, you know-- that's a much different animal in terms of, you know, Japan, with respect to its security policy, has a lot of different dimensions as you peel back the onion, you know, from the standpoint of economic competitiveness, you know, sort of alliance dynamics surrounding concerns about abandonment versus entrapment. You know, is the US going to abandon us at some point in time at a strategic inflection point in the future, particularly vis-a-vis China? Which was always there a concern. Or is it going to entrap us in a crisis that could result in Japanese blood being spilled and completely disrupt our economic model from the standpoint of Tokyo if the US gets into a dust up on the Korean peninsula or even vis-a-vis, you know, China over Taiwan.

You worked in a bureaucracy that had many levels to it, a bureaucracy that in many ways probably is more strict than, at least, some corporate hierarchies, maybe not inside really large companies. There must have been times where you got a request or an ask that came down from a customer, where you got that piece of paper on your desk or whatever, that electronic communication, and said, I really want to go back to that customer and try to explain to him or her that this is the wrong question. And I suspect that many of our students, as they move into data science positions in large companies, are going to encounter exactly that experience. Any advice on how you deal with a situation like that, from your perspective, from all your experience?

Yeah, it's a tough one. I mean, in the corporate environment, and even in, you know, an environment like, you know, the Singaporean or the Australian context that I talked about, which I know somewhat, about, in terms of their bureaucracies and their organizations. You know, smaller, more nimble, and fusing of the intelligence side with the decision making and strategy side-- the daily decision making, tactical, and the strategy side. There, it's obviously easier, probably, to have those types of conversations than it would be in the US Government.

I mean, the ideal scenario is you can get in a room with them or you're in a room with them, and you can have an open conversation about those very things. Oftentimes, though, you know, it's the old, you know, do the analysis and throw it over the transom model, where it's then curried and brought to decision makers. They digest it, and then you know, you might get requirements back, or you're hit with the requirements, and then you have to assess it.

But I will say that, you know, what you try to do is make sure that even as you strive to answer the original question, there are ways to actually broaden the question or deepen the question in a more meaningful way and then play that back in your own assessment. So I'll give you an example where, you know, I was advising-- this is more in a consulting role-- but I was advising one of the government agencies that was responsible for looking at, like, open source, not open source from the Steve Weber open source perspective but open source meaning open media, all types of, you know, traditional media and social media.

They were asked from a consumer somewhere, you know, will the internet bring down the Chinese government? That was literally the question. And you know, they asked us, like, what are we supposed to do with that? And their response was, we're just not going to answer this, which can get you in some deep, deep trouble in the government if you choose not to respond to a policymaker. I'm not even sure how that works, because I've never done that myself.

You know, and I advised him the way to do that is to look at the question and ask yourself, is there something they're actually getting at there that's actually meaningful and that has some degrees of plausibility within it that I can then build upon and then reframe the question? Even if I don't tell them I'm reframing it, but I'm actually reframing the question, that still gets at what they want to get at but allows me to conduct an assessment with integrity. In my own mind, it has analytic honesty and integrity because it's a legitimate question.

It's a great example of a kind of iterative, you know, back-and-forth research design methodology, which, you know, is really very visible sometimes when you do it with other people. Sometimes it's really visible and interesting when you do it with yourself as well, and you know, that's a skill that everybody has to learn because it's actually not taught. It's just something that you kind of learn through the experience of being frustrated with not being able to answer the questions you'd really love to be able to answer.

Right, and I mean, at the end of day, too, the best decision makers that I've ever interacted with are the ones that actually want to get you in the room to see if they can hammer out what is the right question and bring you into that, you know, bring you into that process. But a lot of times, you know, there's still a big wall there that, you know, you just sort of rely on the requirements to come down the chain. And then the feedback-- hopefully, you'll get feedback based on your assessments. You don't always get it, and sometimes, you know, it's crazy.

2.9.4 **Organizational Characteristics and Innovation**

How can we think about innovation in different types of companies? Our primary reaction may be to tackle every problem or opportunity that we confront from the kind of startup or data first mentality. Let's say there's a new technology. And we think that if we adopt it, we could differentiate our organization from others. On the one hand, an organization that already has a data driven or a data first approach might be able to move quickly into that space because it already has the culture and the infrastructure to make the necessary changes.

On the other hand, how do we think about innovation in a large organization that does not fit the description above? In addition to stakeholder analysis, our ability to innovate depends on the organizational maturity, culture, and agility of our work environment. Does the organization have the people with the right skills? Does it have the resources, the money, and the infrastructure? How does this innovation fit into current processes?

Does the timeline needed to roll out this idea in order to be competitive in the space jive with the way that products and decisions are made in that organization? Does the organization have the leadership and culture that will embrace the new idea? The more we are aware of how decisions are made in an organization, the greater our ability to influence decisions.

Think about any kind of standard operating procedures or internal politics that might make it easier or more difficult to ask for approval of a particular project. We could think of this as a stakeholder analysis. When you encountered pushback from the org because whatever you propose is not in line with the grand vision or strategy of the organization, we can think of this as kind of a strategic analysis.

Even if a company sees value in the idea, it might not have the IT systems, the governance, the culture to support the new idea. And here I'm talking about kind of organizational capacity, organizational maturity, culture, and agility. Remember our ideas don't occur in vacuums. As awesome as they are, you have to convince someone or a group of people to buy into the idea. And so you may encounter these kind of broad challenges of implementing data driven solutions in a company that may not have a data first culture.

So what can we do? One thing we could do is identify the stakeholders early on and ensure a kind of co-creation or commitment. Engage them in the approach that solves problems with data. Tell them all the reasons you want to solve this with data, such as transparency or replication. And perhaps most importantly, if you get them involved in the process.

This will help develop a sense of ownership. That's one way to kind of get individuals and organizations that don't have this data first culture to kind of buy in. Another way to do it is to make sure that we don't pit the data against the unique insight of the decision maker. You do want to tell them that the data will have a larger recall than the decision maker. And it will be able to identify edge cases.

The decision maker, for example, might have a really good sense of the market share for your company. But the data might be able to identify those edge cases. And you can make major impact on those edge cases. And so the key here is not to pit the person against the data. The data will help the decision maker do their job. It will complement their existing insight.

2.11 High-Stakes Decision Making Case Study: Cuban Missile Crisis, Part 1

So now we're going to tell a story about an event that happened what feels like a long time ago, the Cuban Missile Crisis, 1962. And it's a story about asymmetric information, it's a story about deception, it's a story about misreading the adversary's backbone intentions and how strong they are, and it's a story about an outrageous attempt to disrupt a competitive environment. And all of us should sound very familiar.

So let's get into it. How did this actually occur? We've got to set the stage. In the early 1960s, you have basically a duopoly. Think of it as a duopolistic market, two big superpowers.

They're competitors. They're trying to establish dominance over each other and over the world, just like two big companies in a marketplace. And it's been a really intense competition. This is a period of the Cold War, and sometimes these competitors get very close to getting into major fights with each other. There's a series of crises in the 1950s. There are crises over Berlin. There are crises in the Taiwan Straits. There are crises in the Middle East.

The point is there is very low trust between these competitors, there's very high levels of tension, and they're not cooperating very effectively. And there is this important sense that just one mistake, maybe even a small one, could lead to the end of the world. The stakes are pretty high.

Let's get a little background to see what's happening here and focus on the asymmetric information, the fact that the parties don't know what the other one knows. 1957-- everybody remembers the thing called Sputnik, the Sputnik moment. The Russians are the first to put a satellite in low Earth orbit, and this just freaks out the Americans because it looks to them like they're way ahead in the space race.

Now, what we now know with history is that the Russians were not very far ahead at all, but it wasn't known at the time. That's the asymmetric information, point one. The fear in the United States, of course, is that maybe the Russians are going to get intercontinental ballistic missiles before we do. And it's really important to recognize that fear is the reigning emotion here. It's not just about hurt pride. It's fear. And when people are afraid, they don't make very good use of information.

Now, Premier Khrushchev, the Russian leader, takes advantage of that opportunity, and he engages in a massive deception. So all of us have seen the pictures of him banging his shoe on the table at the United Nations and threatening the United States with nuclear devastation. Basically, he's trying to bully President Eisenhower into backing off on Berlin and elsewhere. It's all a big lie. But

The US doesn't have the information about what was really going on. We don't actually know what the Soviet ICBM program is doing. And in fact, what we now know is that it was in big trouble. Now, here's the funny part of the story.

In 1960, the first televised presidential debate, Nixon versus Kennedy-- and Kennedy beats up on Nixon, who has been vice president for the last eight years. He says, how could you have permitted the Russians to get ahead in this race? He quotes about what he calls "the missile gap," and he actually believed it. And he beat the Republicans up for letting it happen.

The funny thing is Nixon knew better. The US had launched its first spy satellite that summer in August of 1960. Nixon actually knew that the Russians were making this whole thing up, that they didn't really have very many ICBMs at all and that the real missile gap was in favor of the United States. But because it was such a secret thing, Nixon wasn't allowed to say anything about it.

And if you want to go watch the tape of that video, that debate, you can see it. Nixon is standing there. He's sweating, he's frustrated, he's angry, and boy, you could actually feel some sympathy for him at that moment because he has a piece of information which will change the game here. But he's not allowed to tell anyone. So Kennedy wins the election.

And as soon as he heads into office, he finds out the truth. Then he decides he's going to turn the tables. Now he's got the information advantage. And when the next Berlin crisis happens in October of 1961, Kennedy has his Deputy Secretary of State make a very public speech in which he basically says to Khrushchev, we are calling your bluff.

We now know that you've been lying to us the whole time. We have all the information about your ICBM program. You don't have as many ICBMs as you think you do or as you told us you do, and we actually have the advantage and we're not going to get pushed around.

Now, Khrushchev's absolutely humiliated by this. But again, it's not just emotion. His lie has now been exposed, and his country is now vulnerable to exactly the kind of coercion by the United States that the Russians had tried to use against the US. This time, however, it's potentially for real. So he almost has to do something, and he has to do something fast.

Now, a little bit of personal stuff going on here-- Khrushchev thinks Kennedy is a wimp. He had met him in Vienna at a summit in the summer of '61, and he just was not impressed with what he saw. He thought Kennedy was young. He thought Kennedy was inexperienced. He thought Kennedy was a bit naive. So Khrushchev figures, look, I'm smarter than this guy. And even now that the tables have been balanced and both sides know what's really going on, I can outsmart Kennedy, even when we have the same information.

So Khrushchev has a genius insight. If he had had the advantage of taking a course from Clay Christensen at Harvard Business School or had read The Innovator's Dilemma, he would have known exactly what to do. But he didn't mean to do that. He just figured out the concept of disruptive innovation on his own.

Khrushchev figures, I'm not going to compete with the United States on its own territory. I'm going to come up with something that's cheaper, faster, and easier with a lot of impact on the strategic balance. I don't need fancy ICBMs. What I have are a bunch of medium-range ballistic missiles, and those are pretty easy to build.

And what I need to do is put them really close to the United States. I need to put them in Cuba. Cuba's only 90 miles from Florida, 550 miles from Washington, DC. And so with a single disruptive move, the Soviets are going to jump back out in front in this competition.

Here's the problem and the only problem. Kennedy had explicitly warned Khrushchev in Vienna that the United States would not tolerate the introduction of offensive weapons in Cuba, and he had made that warning public. So his credibility was on the line, both with the Russians and with the Americans.

Khrushchev's synthesis-- he figures it out. Here's what I'm going to try to do. If I place the missiles in Cuba secretly, by the time Kennedy finds out, it will be too late. He's not going to have the information he needs, I'm going to be able to do it undercover, and ultimately, he's not going to have the backbone to stand up to me when it counts. So I'm going to win twice actually in this game.

I'm going to win once by having the missiles there, and I'm going to win again by showing the world that I have guts and Kennedy doesn't. So here's the strategy. Because the information is not shared, because secrets can be held, I am going to create a fait accompli in Kennedy's face and win this game.

2.12 High-Stakes Decision Making Case Study: Cuban Missile Crisis, Part 2

We now know that the Soviet Union, not content with Dr. Castro's oath the fealty, not content with the destruction of Cuban independence, not content with the extinction of Soviet power into the Western hemisphere, not content with a challenge to the inter-American system and to the United Nations charter, has decided to transform Cuba into a base for communist aggression.

Into a base for putting all of the Americas under the nuclear gun. And thereby, to intensify the Soviet diplomacy of blackmail in every part of the world.

The United States answer to what Adlai Stevenson termed "Soviet blackmail" in Cuba was a quarantine of all offensive weapons being shipped from Russia to that island fortress. The US Threw up a steel fence, prepared to stop any vessel carrying materials of war. In Cuba itself, 100,000 men were put under emergency orders, as they had been during past invasion scares.

The waterfront in Havana and along other parts of the coast bristled with gun emplacements as the Cuban regime waited to see what their bosses in the Kremlin were to do. Cuba became the focus of world attention. Here centered the most critical threat of global war since the surrender of Germany 17 years ago.

Castro has put every able-bodied man through military training. He has even armed some as young as 12 years of age, and authorities assembled thousands in cities and villages for patriotic rallies. As in the past, these rallies are designed to whip up hate of what Castro calls Yankee imperialistic warmongers.

To suggestions that a UN team inspect missile sites, Castro said that they had better come ready for combat. He went on to call President Kennedy a pirate for setting up the quarantine. The United States arrived at the decision for an arms blockade after studying reconnaissance photographs made with high powered cameras from planes flying several miles from the Cuban coast. These cameras are described as capable of spotting a golf ball on a putting green from 40,000 feet.

Literally thousands of pictures can be made on each flight by these planes, and they are studied by photo interpreters who are capable of analyzing details that an untrained eye would miss. Here, for example, is a medium range ballistic missile base that has been labeled by these specialists. Suddenly, the veil is torn from the Russian secrets.

Another photo revealed a surface to air missile assembly depot, a base to supply the offensive sites. Russian technicians ripped through heavy jungle growth to carve out airstrips for high performance MIG-21 jets, a plane easily capable of strikes far into the United States. In the greatest display of hemisphere solidarity since World War II, the Organization of American States unanimously endorses the actions of the United States, and many pledge arms and men to the cause. The vote is 20 to nothing, with Cuba absent, commending the US for its efforts to bring about the dismantling of the missile bases. The organization votes the use of armed force to carry out the resolution sponsored by Secretary of State Dean Rusk, thus uniting all of the Americas in a common cause.

Meanwhile the United States continues to reinforce its Cuban base at Guantanamo Bay, the Naval depot that Castro wants the US to give up. These Marines have been assigned the job of protecting the base against any Cuban threats that might arise during the present crisis. They'll be on a 24-hour alert, our first line of defense.

The United States went to the UN Security Council for a resolution calling for a withdrawal of all offensive weapons from Cuba. A delegation from the island heard [INAUDIBLE] call on both sides for a three week freeze, but the Secretary General was told that President Kennedy wants the missiles scrapped first. [INAUDIBLE] boss, Khrushchev proposed that the US withdraw its vessels and he would stop shipments. President Kennedy's missile scrapping demand was his reply. The US resolution was firm and strongly worded.

Mr. President, I am submitting today a resolution to the Security Council designed to find a way out of this calamitous situation. This resolution calls as an interim measure under Article 40 of the charter for the immediate dismantling and withdrawal from Cuba of all missile and other offensive weapons. It further authorizes and requests the acting Secretary General to dispatch to Cuba a United Nations observer corps to assure and report on compliance with this resolution. Upon UN certification of compliance, it calls for the termination of the measures of quarantine against military shipments to Cuba. And in conclusion, it urgently recommends that the United States and the Soviet Union confer promptly on measures to remove the existing threat to the security of the Western hemisphere and the peace of the world, and to report there on to this Security Council.

2.13 High-Stakes Decision Making Case Study: Cuban Missile Crisis, Part 3

So the question we want to ask at this point is, what is the United States going to do and why is it going to do it? Now, before we get into the specific options, let's just talk about two general principles of high pressure decision making that have been observed over time and across domains, because these are going to come back later in the semester as we think about having to advise people who are in situations like this and giving them the information they need in a way they can use it.

Two really simple principles. One is the inverted u-shaped curve which you'll see on the slide here. It's basically trying to track stress versus performance. At a very low stress level of stress we know people actually don't perform that well in tough decision making situations. As stress goes up they start to perform a little bit better, but everybody knows this in their own life, there's a certain point where you get to the inflection point on the curve and more stress leads to reduced performance.

The thing is, nobody knows exactly where that inverted u-shaped curve is for themself. We know some of the reasons why. People who are under a little stress they'll do a little more information search, they'll feel some pressure to close the deal, they'll have vigilance about the environment around them, but all that stuff has to be in balance, and it's one of those things where you sort of know it for yourself when you're on the flip side of that inverted u-shaped curve and your performance is going down, but you may not be able to tell where other people are on that u-shaped curve. So that's one issue to think about.

Second issue to think about is the well-known phenomenon that's called groupthink, and again, we'll talk more about this in a future week. The real interesting work on groupthink was actually done about the Cuban Missile Crisis. It's sort of the iconic case of bad groupthink. And it's really commonly seen in small groups under high pressure trying to make high stakes decisions. They get involved in all sorts of bad errors like what we call premature closure, trying to come to a decision too quickly or drinking the Kool-Aid. Agreeing with everybody else in the room because it feels good to do that.

Or not really taking the time to question basic assumptions, because that might run into actually time pressure. And most importantly, resistance to discrepant information. If everybody around you is saying, we know what to do here and a new piece of data comes in, sometimes it's really hard to get that data into the conversation. So keeping those two really simple ideas in mind, let's talk about what Kennedy's options really are at this moment. His advisers prepare for him basically four options that he needs to consider.

One is, maybe we should do nothing, nothing at all. The second is what's called a surgical air strike. Let's mount a really specific airstrike and just try to attack the very specific missile emplacements that Khrushchev has put in Cuba, not do anything more. The third option is a massive airstrike followed by invasion of the island. So we're going to basically bomb the island, and then we're going to invade the island and take over Cuba. That's the biggest option. And then the fourth option that's kind of sitting out there is a Naval blockade. Let's put the Naval ships around the island and let nothing come in or out.

So Kennedy's got to make a choice. Here's the analysis that he goes through. Do nothing. Well, actually some of his advisers and some historians since have pointed out that maybe this thing Khrushchev did is not really such a big deal. Maybe it isn't really that disruptive. Yeah, it changes the strategic balance in the short term, but in the long term it really isn't going to make very much of a difference. A couple of years from now we know that both countries are going to have ICBMs and the system's going to come back into balance.

So yeah, it was insulting, it was obnoxious, it made Kennedy look like a wimp, but ultimately it isn't really a big enough strategic issue to matter, and maybe it's not worth going to war over. Kennedy won't buy that argument. He can't, he's a politician, even if it makes perfect sense. The second option, the surgical air strike, well, let's go and just take out the missile emplacements and minimize any collateral damage. The problem here is, again, the information isn't good enough. The Air Force can't assure Kennedy that they can get all the missiles. They can't assure Kennedy that some US planes might not be shot down. And oh by the way, they also tell Kennedy that we suspect there are probably some Russian advisers at those missile sites. So those bombs are going to kill Russians and that's going to be very provocative.

The third option starts to come into focus, massive airstrike, and then invasion of the island. And in some ways, to a decisive decision maker, somebody who really wants to solve the problem, this is the most attractive because it basically solves the immediate problem. And at the same point, it solves the longer term problem of Castro's Cuba, which Kennedy was really troubled by. Of course, it's also the riskiest option because credibility is a two way street. Khrushchev is pledged to defend his allies, and Kennedy knows that if he takes that strike at Khrushchev's ally, then Khrushchev may have to strike back at one of his allies for example, West Germany and Berlin.

And then there's the Naval quarantine. Interesting idea. Blockade the island, stop any ships trying to deliver military equipment to Cuba. This is interesting because in some ways it's the least provocative of the choices. Doesn't involve the direct use of force. Blockade is considered an act of war but if you're just ringing ships around an island you're not shooting at anybody. But here's the thing. It's also directly, at that moment, the least effective. Blockading the island doesn't do anything to remove the missiles. At best it kind of freezes the situation where it is. And Kennedy just was not certain that the Cubans needed any additional shipments to complete at least some of the missile deployments.

So what it does is, blockading the island slows things down. Probably a good thing, but it gives the next choice and in some sense it passes the initiative over to the other side.

2.14 Decision-Making Models

So what we want to focus on out of this story is the choice process-- how does the choice get made and why does it get made to what it is. Well, the executive committee that's tasked with making this choice deliberates for several days. And actually, opinions are split about what we should do. There are kind of hawks and doves.

The early consensus that starts to form is around the airstrike option. And you can see why. It's sort of the most obvious way to deal with the problem. We told them-- the US told the Russians, don't do something. They did that thing. We're going to strike specifically at what they did, and we're going to fix the problem. And clearly that puts us in the right.

But that decision starts to fade over the course of the week. And something very important fades it, not because of new information or new data coming in, but because Kennedy engages in empathic thinking. He starts to ask himself, so, let's imagine for a moment that I'm in Khrushchev's shoes. What would I do? And that changes the game.

There's a deadline here. The deadline is October 22nd and Kennedy is going to make a speech to the United States, to the nation, a public speech. It's going to be televised all over the world. And he's going to essentially explain what he's doing and why. And at the last minute, essentially, he decides to reconfigure around the blockade option. And that's what he chooses.

Now, it turns out, historically, that this was a good choice. But we didn't know that at the time. It could have easily gone wrong. And it almost did in many different ways. Partly for that reason, and partly because it was far from the obvious choice, this is probably the most closely studied a high pressure decision of all time. There have been books and books and books written about this. Why did the decision come out the way it did?

But there's one book-- it's an old book, but it's particularly useful in thinking about how this experience can be generalized into better models of decision making. The book is called Essence of Decision by Graham Allison. Worth a read, actually, if you're interested.

But let me summarize, essentially, three models that come out to try to explain the decision. And this matters a lot because, if you want to actually influence decisions effectively, you need a kind of theory of the case about how important decisions actually do get made.

Personal story. I was once involved in an advisory role for a development bank. And I had to prepare a bunch of documents leading up to an important decision. And I was in the room. I wrote the documents. I sat in the room. I listened to what everybody said. I saw the way the decision came out. I walked out of the room and asked myself, do I really know why that decision came out the way it did? And I'm sure to this day that I actually don't know. And the reason is is because I didn't have a sort of implicit model, how is that decision getting it made.

So there's three models we want to consider. They all come out of Graham Allison. Model 1 is going to be the one that's easiest for people to grapple with. It's called the rational actor model. And actually it's kind of the default for many scientifically minded people. It is what it sounds like. The assumption is, look, the government or the company that's making the decision, that's the right unit of analysis, and it's going to make a rational decision. It's going to look at the data it has. It's going to weigh the costs and benefits. It's going to look at the risks and opportunities, maybe do a SWOT analysis or something like that. It's going to develop an expected utility calculus. And it's going to land on the choice with the highest payoff.

In some ways, that could explain what happened here. It's a simple and somewhat cleanest explanation of the crisis. The executive committee evaluates the options. It says that the blockade has the highest subjective expected utility. That's the dominant choice. And it works because Khrushchev's best choice then is to take the missiles out of Cuba. Not entirely satisfying, but it's an explanation.

There's a second model that Allison puts forward. And it's a really important one for anyone who's worked inside a reasonably sized organization. This is called the organizational process model. I like to call it standard operating procedure. What do organizations like this do in situations like that? And this is the default, not for many scientifically minded people, but actually for many organizational behavior kinds of people, and many consultants.

The assumption is, look, you got a complex organization, whether it be a government or a business. And these complex organizations have kind of ritualistic standard operating procedures, ways of doing things, a kind of repertoire of decisions and behaviors. And so that's what they do. And so when they face a hard decision-- you know, leaders don't evaluate it as some kind of rational whole. The kind of break it down into pieces. They assign the pieces to the right department-- you go solve this problem, you solve that problem, you solve that problem-- figure out what I should do, and then try to put it back together.

And because they're under time pressure, and because they're in the resource constraints, the leaders of those departments delegate even more. They delegate down to their subordinates. And then they put the pieces back together to find a solution that is good enough. They don't optimize. They satisfice. They look for something that's good enough. And they often stop searching for a better solution once they find a solution that satisfactory.

The top leaders do exactly the same. They're basically trying to satisfice. They're trying to reduce short term uncertainty. And they pay less attention than they probably should to the longer term consequences. They are basically going to favor options that allow the parts of the organization to do what they already know how to do, rather than do something new and unfamiliar. And so, basically, what they're doing is choosing among preexisting plans.

And in fact, there's a lot of evidence that this is exactly what happened in the Cuban Missile Crisis. There's a moment where the Secretary of the Navy says the Secretary of Defense McNamara-- "Mr. Secretary, the Navy knows how to run a blockade." That's a classic example of organizational behavior and standard operating procedure.

Third model, briefly, what we call bureaucratic politics. That's the default-- again, not for scientifically minded people, but for cynical politicians or political scientists like me, and often for people who have worked for a long time inside a particular company. The assumption here is people do what benefits them individually, or colloquially, where you stand on an issue depends on where you sit inside the organization. It's always politics, from top to bottom, and people make choices based on what they think will benefit their political power and their influence inside the organization.

In fact, for bureaucratic politics models, there's no organization that makes decisions. There's just a bunch of political players who are jockeying with each other to see who can come out on top. And in that world, who is sitting at the decision table is the most important determinant of what gets decided, more important than anything else in the external environment or in the market.

And the outcome that wins is because of the person who advocates it-- the leader who gets charisma, personality, bargaining skill, personal ties to other influential people to get what he wants or she wants. It's not necessarily the best idea. Effective leaders go out there and create a political consensus. And they create, in their minds, a belief that that's really the best thing to do. Right?

The story here is Khrushchev put the missiles in Cuba to save his own ass. Kennedy was under pressure from congressional Republicans who made Cuba huge issue for him. And he had to save his own ass. And both sides were obsessed with saving face.

Now, why does this matter so much? Look, not only because it's interesting and fun to understand the history, but because, fundamentally, any attempt to influence a decision inside an organization demands a theory of the case about how that decision is actually going to be made and how it can be changed. And that is as true for using data to change a decision as it would be for using anything else.

Here's a hint that we're going to follow up on. One of the biggest mistakes that you or anyone else can make is to assume the data always favors the rational choice model, and it will always work that way. It's a seductive assumption for scientifically minded people. And it requires less insider information and much less data about the decision making system. And it's very, very nice if it were true. But none of that is good if the model is downright wrong. So let's role play that to see what it actually looks like when people are playing it out.

2.15. **Separate Outcome From Process**
Andy, thank you for joining us today. It's a little role reversal. You're in the office. I'm at home. But that'll work just fine.

Great.

I can see that map of California there behind you.

Yep.

So what we wanted to talk about today was-- so what is a good decision anyway and how can you set things up or tilt the tables so that good decisions become more likely? And I thought the way we would start is I would outline for you a story that I'll call not atypical decision process. That's probably something that everybody recognizes.

So the first thing is you have a problem that requires a decision, but the problem's sort of semi-defined by an executive, not tightly defined. And then, it's pushed down for somebody else to generate alternatives. That somebody else gets an analytic team together that goes through its own process of outlining alternatives, assessing some facts and logical basis, and developing a recommendation. And that all sounds good, and that recommendation is then encapsulated in a 165-slide PowerPoint deck.

Right.

The third thing that happens is that deck is emailed out to all the participants in the meeting where he's going to make a decision late the night before the meeting, so nobody reads it. OK? Then in the meeting, someone presents a kind of truncated summary of the deck. And now, it gets interesting.

Some people believe the team has not looked at the right data. Some people argue that the criteria for decision making isn't clear, or maybe it's just plain wrong.

Mm-hmm.

Some people imagine a better alternative that the people who created the slide deck don't have on their list of alternatives. And some people don't think that no matter what we choose, we can execute on that decision, so it almost doesn't matter what decision we make. And so on, and so on, and so forth. So that is not atypical decision processes for anyone who's ever-- does that sound familiar to you?

Yeah.

And how familiar is it, and what do you think in terms of adding a data input to that?

From my experience in industry, it's incredibly familiar. It's one of-- you describe that deck where it's almost, let's kill them. Once we get into this meeting where this decision's going to be made, let's kill them and flood them with-- let's just call it-- data or information. You know? And what I've often found in those conversations of being in those rooms where you're sitting around and you have different people debating aspects is you're missing the almost often is the process of how we even got to this point. What was our original question that we were looking at? Why are we facing this decision? What type of question is it? Is it a how-based question? Or is it a should we do A or should we do B, which can impact those sorts of things.

So you have the extreme of having a whole lot of data, a whole lot of information that's out there that people are trying to parse exactly as you described in a very brief period of time. Many of those folks have not been necessarily part of the process and so forth.

You have that or I've actually experienced a number of times the opposite effect, where there isn't a slide deck. There's nothing written down. You know? But it's been collectively happening in various people's minds. And this happened a lot with a lot of product in management, project management decided related decisions and things like that.

It would be bounced around where the data wouldn't necessarily be recorded. Information would be floating around, so you'd have almost too much data, too much information or too little data, too much information for both equally high stakes decisions. And one of the challenges we ran into in both of those is when we would bring in data for making a decision, particularly for product things. Because we would use a mix of qualitative data from-- we spoke to these users or these sales prospects-- with quantitative data, pulling out of the systems that we were designing themselves, saying people said A, we saw the following. Let's compare the two to help guide our decision.

Oftentimes, when we were using the data, the quantitative aspect, it was in those multi-slide deck, hundred slide deck things. There would be so much data in there that people would be trying to parse that it would be just overwhelming trying to analyze and understand where it came from, what it meant, and things like that.

But the flip side is when you have a zero slide deck or a short slide deck. You had that challenge or that battle of, how do allow data to not just be an anecdote.

Right.

But actually convey, no, no, no. This has meaning. This has weight. Here's how we got here. So it's one of those. It's a challenge either way.

OK.

In what are you doing. And I think for all of those, it was always just, how do we be clear about the process of--

Well let me ask you--

Yes.

Let me ask you a question about that. You know we've talked about almost like a cartoonish version of a research design template. What's the question? Where does it come from? What's the answer? How did I get there? So what?

That may be a reasonable template for research design, but one of my advisors from way back once said to me, the logic of discovery is not necessarily the logic of presentation.

Sure.

But maybe, actually, in this case, it is. Have you ever thought about constructing a decision process around that kind of a story or ever lived in a decision process that went through those steps? And how do you think that did work or how do you think that would work?

And actually, we did use a system like that, particularly for product management related decisions. How are we going to change the product? Are we going to put-- we were a startup. We had a finite number of resources.

Are we going to put our six engineers and developers on A or are we going to put them on B? And what we found is often, early on, those decisions were shaped by the founders. You know? Which worked. That was their ideology of starting the company. We think we should do this.

But as we got bigger and had 60, 70, 80 employees and it got bigger, you couldn't just focus on this one person. So we would have to document and really-- not to be heavy handed or regimental behind what we were doing, but I think it's the word deliberate.

Yeah.

OK, we have resources. We can spend them in some ways. What are we trying to do? What is our objective through using engineering? Is it increasing sales? Improving the quality product?

Yeah.

So having some sort of transparent process behind how we would come to those decisions.

So one view of a process like that would be that you need to get closure on each step before you can move to the next step. In other words, sequential processing of each problem. Everybody's got to agree on what the question is before we can move on to why are we asking it, before we can move on to how are we going to answer it. So that there needs to be sequential closure.

Another view would be you actually can never get that level of consensus. And so you need to step through those things with a little bit more alacrity and recognize that you're never going to get closure on any piece of the process before you move onto the next. I think that's a pretty important choice to make intentionally. What do you think about-- again, how have you seen that work? Or what do you think about that kind of step?

Sure. So I can think of a lot of examples of that. You know when you have the regimented sign-off model in the product world, people would actually physically sign documents. Here's the market requirements doc of what it's going to do. Here's the product doc. And people would physically sign it.

The problem with that is it is so lockstep and regimented that it couldn't ebb and float with realities of running a startup in a very dynamic environment with lots of customers and things like that. So people liked that process, but at the same time, what happened then is it didn't match with reality of what people would do.

Yeah.

There was that, but then there was the whole back channel, under-handed decision. Not under-handed. I don't mean it that way. Obfuscated, hidden decision making that was going on shaping those sorts of things. And that took a while to kind of tease out.

I was much more the fan of, we don't necessarily have to have everybody sign off, but we have to agree that we are not running into potential landmines. We're not missing something really, really obvious. And everybody feels confident and that we're moving in the right direction.

Yeah.

Say like 80%. We have confidence that we're going the right way. If we try and go for that last little bit to get that formal sign-off, we're going to spend so much time and so much energy doing that we're going to lose time. People will get disinterested or go to--as I described-- these alternative decision making processes, which were not great.

Right. Right.

So as long as we can collectively move as a group in the right direction-- not in the right direction, but continuing the process forward-- then we can just keep making better decisions and better outcomes that we can learn from over time.

So a decision process that gets too rigid-- I mean, ideally, boy, wouldn't it be great if everybody could agree? Yes, that is exactly the question we're trying to answer. And no, it's not both/and. It's an either/or or something like that. It would be wonderful for a kind of an idealized notion of a decision process to be able to step through each one of those steps on the ladder before you move onto the next. But in reality, we all know that that's never going to happen.

On the same token, having been inside of faculty meetings that try to make decisions, you have the extreme on the other side where, as we often say, a 36 to 1 vote is a tie. And so everything gets hung up. That's not a good decision process either. And so, I think what you're saying is, at least in the best processes you've been involved in, there's at least a kind of consensual awareness among the group of some threshold that they need to get over.

Yeah.

And when you've gotten over that, people know.

Yeah. One of the phrases that we would kind of toss around when we were doing it was more of a checklist.

OK.

We have a process, but we have a checklist of things. And sometimes it was an actual, physical list written on a big dry erase board that everyone publicly could see in the office being like, here is the process of how we're going to get from product, concept, idea into actually pushing the thing out.

Yeah.

But then, more of a checklist of, do we have buy-in from this, how we run it by these folks. That helped in those moments of, we need to make the decision. One of the dynamics is we go, OK, what's on the checklist? What has worked in the past--

Yeah.

--to help us make a-- ideally a great decision, but just a better one than last time.

Yes. Interesting.

How do we keep better and more informed because we can't go for the ideal and the perfect? That will take forever. How do we just learn from what we did and make it better and better and better? And sometimes, that checklist would get really long, but then, we'd realize, well, we don't need some of those things. So it would be a little bit smaller.

Start to fill it up.

And it would kind of ebb and flow over time. But we keep a snapshot of what are the steps that we went through? What are the things that we missed such that we could just make smarter, more intelligent decisions as we went as an organization?

Good. Now let's rock it up for a second to that the essence or the meat of the question, Andy. So from your perspective-- I know what it is in the kind of academic, abstract decision literature, but let's put that aside. From your perspective, what's a good decision? How do I know?

How do you make a good decision or what is a good one?

No. What is a good decision, and how do I know when I've made one?

I think a good decision is one of-- let's decouple it from the outcome itself.

Mm-hmm. OK. Important point.

There's so much variable that can impact whether the outcome was right or good or not. And I think a good decision is one that you as an individual or as a participant in the process can stand behind with confidence on, here's the question or decision that we were facing. Here is the process that we followed. Here is the data that we used to guide the decision. And I mean data in any sort of sense, part quantitative, qualitative, whatever it is.

When you can really confidently stand in front of-- whether it's your users, your customers, your colleagues, or whatever it is and being like, this was a good decision. Here's what we did. We followed these sorts of steps.

Given that there is risk and uncertainty behind what we are doing, this is what we followed, and here's how we got here. That is a good decision. It takes out that luck aspect. Because I can think back in a work experience of like, oh, that was a great decision. And you are like, that was flat luck.

Well, let's call it--

Like everybody recognized that.

Let's call it prob--

That's OK, but we can't bank on that.

Let's call it probabilistic or stochastic. It's like we all know-- and this is the part I want to highlight in this element-- we all know that a good decision process and a good decision may not always lead to a good outcome because we live in a world of risk and stochasticity. But in practice, people always seem to forget that.

They do.

And it requires a high level of self-consciousness to recognize later. You don't want to go back and sort of undermine that. So let me then switch to the second to last question is, how often and how valuable do you think a decision process post mortem is? In other words, going back over how we did this and analyzing, so what did we do well? What did we do poorly?

Yeah.

Do people do it? Should they do it more? How valuable is it?

I think it can be incredibly valuable. And one where I've seen it in the workplace experience, particularly in startups and big companies, too, is often comes in at, I would say, at the wrong time. So say something goes really bad, then people want to come in and being like, oh, what happened? You know? And everybody gets defensive and tries to justify what they're trying to do. And I've seen that definitely.

But more, what I learned more was when managers would come in and being like, wow, that software release went really well. Let's talk about what happened there. So people aren't necessarily on the defensive, and people can be more coy and candid. And oftentimes, in those moments, people will be like, you know what? That was flat out luck. Like it was really luck that that part just kind of worked.

Yeah.

And that's where you learn more. I say it is described as a post-mortem. I immediately think of like, post-mortem is bad! You know? Just the language is bad. You know? But instead of looking at, how did we come to this decision? I think it's valuable to do that, but its one in organization is, start with the little ones. Start with a little, basic decision. It's not high stakes, not risky or anything, where you're fairly certain of the dynamics of what can happen, and learn from that.

Then eventually, get to that point where you can do post-mortem on those more difficult challenging ones. Because I think it takes a while to get the organizational culture up to that point where people then feel comfortable when things go bad, they go south, when go, you know what? We've to this process before.

Right.

No one's going to get called onto the carpet. No one's going to get thrown out of the building. Sometimes that happens, but it's rash. So then, people can feel more comfortable with it at that point.

Good.

Rather than having it come screaming and disrupt things.

All right. So--

And then people just behave in all sorts of ways for security reasons, to protect themselves.

So one last question on this. Let's imagine that you had at your disposal a data science team that you could task at one point but not for the entire decision process.

Yeah.

In other words, it's a limited resource. Lots of people want to use those data scientists for one or another problem. You're the guy who's in charge of a particular decision or owns that decision. You're going through this kind of a process, all the way from framing the choice, to collecting the data, to make recommendation, et cetera, all the way to the post-mortem. But you could really only deploy your data science team at one step in that process.

Yes.

Where would you deploy them?

So I would say frame it, the choice.

Interesting. OK, say more about that.

OK. So why I would do that-- because I feel like if you brought in an educated team of data science folks, they would be looking at this as, here are some choices. Let's keep it simple. A, B, and C. What do we do? A, B, and C.

They can help you think of it at that stage of, how do we capture data?

Right. To help understand A, B, and C, the process of, say, comparing A, B, and C, and so forth. So you can map that out through the rest of the process of what you are doing.

Right.

So that you could, say, at the end of it have enough data recorded and captured that a team could look at it. And I say that just from the point of, well, if you bring them in at the end, they may have nothing to look at.

Yeah.

Or they may have-- particularly in our cases, they would have data to analyze that would be only collected from certain folks, but not from others.

Right.

And then, there's, say, inherent biases in what you're going to collect and so forth. And instead being like, OK, here we are. We're going to frame this choice. Help us figure out what we should instrument to learn from this process.

So that's pretty much--

As we get smarter.

That's very much a product manager's mindset--

It is.

--brought to the world of data science. No, that's not a criticism.

Good.

I actually think that's an important insight. The same kind of discipline of product management applied to the engineering team, in this case, applied to the data science team as well.

Yeah. Yeah.

Good.

Because I think then, you can use the results and feedback of that to tune and tweak those things as you go down the road versus just capturing and analyzing the data at the end. How do you go back then and inform and make the process smarter? Integrate the data scientists from the get-go.

Right.

What can we learn? What do you guys know about data? What are we missing? What are we blind to? How do we capture that? Whether you do or not, that's separate, but you try.

Yeah.

You're at least aware of what are those data elements that you can collect and then analyze at some point to make your decisions better.

Perfect. Listen, Andy. Thank you so much for this.

Sure thing.

Go off and make some good decisions today, will you?

All right. I'll try to.

OK.

All right. Thanks.