Harvard Data Science Review • Issue 3.2, Spring 2021

A Conversation With Michael V. Drake and Jennifer Chayes

Michael V. Drake¹ Jennifer Chayes² Xiao-Li Meng^{3,4}

¹Office of the President, University of California, Oakland, California, United States of America,

²Division of Computing, Data Science, and Society, University of California Berkeley, Berkeley, California, United States of America,

³Department of Statistics, Faculty of Arts and Sciences, Harvard University, Cambridge, Massachusetts, United States of America,

⁴Harvard Data Science Review, Harvard Data Science Initiative, Harvard University, Cambridge, Massachusetts, United States of America

Published on: Apr 30, 2021

DOI: <https://doi.org/10.1162/99608f92.11cb886e>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

Harvard Data Science Review's Editor-in-Chief, Xiao-Li Meng, recently met with Dr. Michael Drake, the 21st President of the University of California and Professor Jennifer Chayes, Associate Provost of the Division of Computing, Data Science, and Society and Dean of the School of Information, UC Berkeley. Chayes served a dual role as a guest and a co-host, being one of the Co-Editors of *Harvard Data Science Review*. The trio discussed data science's impact on society and higher education, the importance of data science education and data literacy and how data science fits into University of California's broader mission of education, research and public service. The interview was virtually conducted on January, 28, 2021.

This interview is part of HDSR's Conversations with Leaders series.

HDSR includes both an audio recording and written transcript of the interview below. The transcript that appears below has been edited for purposes of grammar and clarity.

▶ 0:00 / 34:07

Audio recording of the interview

Xiao-Li Meng (XLM): President Drake and Associate Provost Chayes, thank you so much for joining this Conversation with Leaders, featured by Harvard Data Science Review. Before I became Editor-in-Chief of the *Harvard Data Science Review*, I served as a dean for five years. So I had a little taste about the crazy-busy schedule of being a university administrator. And I knew that has no comparison to what, Professor Drake, you do, because you lead ten universities. It's also no comparison to what, Jennifer you are doing because you are

building a new school. So I truly appreciate the time you take to talk to us. So, let me just get to the questions that we have for you that our readers and listeners will truly appreciate. Since we're talking about data science, the biggest question everybody's wondering about, and I want to ask President Drake first, how do you see data science impacting and changing the world we live in today?

President Michael V. Drake (MVD): Well, thank you very much, and very nice to have the opportunity to speak with you. Data science changes everything we do—even our communications through our Zooming and being so connected to technology in the way that we've been connected through this last very challenging year with the pandemic—shows that the place of data and the place of gathering information and sharing it electronically in our lives is something that's ubiquitous and required for us on a daily basis. Just speaking of the pandemic, for instance, we receive data every day on the pandemic, on the numbers of patients. We're seeing the numbers that are hospitalized, the status of patients in our ICUs—and using those data every day to change our plans and our procedures as we move forward. I could go on and on and on, but I'd say that information and data really inform everything that we're doing as our world unfolds and allows us to have a better world.

XLM: Oh, thank you. Jennifer, you have served in so many different roles, both inside and outside academia. So, what's your view, from your very broad perspective?

Jennifer Chayes (JC): I think there are huge positives and potential pitfalls with data science that are playing out in this world. Data science as an academic discipline—and as a discipline that affects the way we do science and technology and medicine—is absolutely wonderful. As a platform for some of our public systems, it can be the great equalizer and it can help us to distribute resources more equitably, but it can also be misused in those respects. Right now, in COVID times, as President Drake said, we are experiencing some of the advantages of being able to be connected through a medium that really has its basis in data science. But I think we also experience some of the risks of this medium of data and information, in things like the usurping of that power that we've witnessed as destabilization in our government and in governments around the world. So, I think there's great promise and also threats. Therefore, it is our responsibility to make sure that data science is a force for good in this world.

MVD: Of course, I agree completely, and we all saw this—that data and information are very, very powerful, but also can be misused. Actually, I had an interesting conversation just over the weekend with someone I've

known from grade school and someone who lives in the Midwest. We were talking about things, and she was asking me a question related to medicine and health and said, “Well, you know, I don't really know. You can't really believe anyone any more. So, I don't really pay attention to things.” It was actually stunning and horrifying for me to hear that there are people who are so overwhelmed by the lack of control and quality of the data that we see and the information that we get that they decide, ‘Great, I'll just withdraw from all of it.’ And if that's the case—if people can stop worrying about what's actually true, what's not true, and back away from that—then I think we're lost. Your points are very, very important, and that I appreciate them.

XLM: I just can't agree more, and this is one of the things that we all lose sleep over during the night as data scientists. If nobody trusts what we do, and when everybody thinks anything's possible, it would be just terrible. I guess one way to change that is through education, and that's the key. So, with UC's being a true leader of public universities and universities in general, I want to ask you what UC is doing to advance data science, particularly how it fits into UC's broader mission of teaching, research, and public service?

MVD: Well, again, information and science are the foundations upon which we rest our enterprise. I say that knowledge is our product, and that's knowledge that we transmit to our students and others through education, or knowledge that we create through our research. Data science really as a way of collecting knowledge and information and then transmitting it to people. So, we work really hard on holding up the highest ideals of doing so in an appropriate and transparent fashion. Clark Kerr said—one of his famous quotes was that we're not in the business of making information safe for students, we're in the business of making students safe for information. He used the word “speech,” not “information,” so he said we're not in the business of making speech safe for students, we're in the business of making students safe for speech. But I think data science fits in that sentence in the same place. We want thoughtful, appropriately skeptical, energetic critics of information to be able to receive and process it appropriately.

JC: I really think of data science education and data literacy as becoming as necessary as reading, writing, and very basic mathematics. We have to inoculate our students against misinformation. So, I think that's precisely what Clark Kerr was talking about, what you were talking about, President Drake—we have to enable the students to think critically about the data they are hearing about and the data-derived consequences they are hearing about, which may not, in fact, actually be data-derived consequences.

MVD: Right.

XLM: Absolutely. Jennifer, let me follow up with you. First, congratulations on your leadership role at Berkeley. I know you have been leading this university-wide initiative to build this structure of data science. I can claim a tiny bit of credit because I was serving on the visiting committee to the statistics department when initially this idea started and we, the visiting committee, strongly endorsed building up something at this top level. Can you say something about why you think it's very important to build this thing university-wide, versus, say, a department or an institute?

JC: Thank you, Xiao-Li, and thank you for supporting the formation of this. Our division is called the Division of Computing, Data Science, and Society. Society is a really important element of this. We're not just building an institute or an overlay. I believe that computing and data science have become part of the fabrics of our lives. This is going to be increasingly true in our interaction with our public systems. Our educational system, our social welfare system, our public health system, and our criminal justice system are getting more and more algorithmically-driven. These are driven by data and interpretations of that data. I think it's really important that we pull statistics and computing out of the sole realm of STEM and into the nexus of STEM and all of the human-centered disciplines, including social sciences, arts and humanities, and also all of our professional schools. Right now we're experiencing a market failure in communication. We didn't understand how search engines or friendship networks would be monetized. We're trying to get the genie back into the bottle with technical, regulatory, and normative solutions. We should be creating the next genie for the public good. And that is where Berkeley is just primed to lead. We have the School of Information, the Department of EECS (Electrical Engineering and Computer Sciences), shared jointly with the College of Engineering, and the Department of Statistics, which you commented on, Xiao-Li, as our core divisions. But, I believe that five years from now we will have failed if we don't have joint appointments with every other college and school on campus, in every division, within every college.

XLM: That's a lot of work, speaking as a former dean, understanding how these joining appointments are made. But I guess that's where you make that kind of headway by resolving these issues, and let people see the common mission of data science. And I think that's another kind of perfect segue way to my next question, because we talk about misinformation when training our students. There are a lot of discussions now about data science ethics, including issues of privacy, equity, safety, and other considerations. So, I want to ask both of you, starting with President Drake, what do you see as a university like UC's responsibility when it comes to this emerging technology, like data science, machine learning, A.I., all those things, in the realm of all these kind of data, ethical considerations?

MVD: As you mentioned, there are these new technologies that are coming to us and being used to help advance us in so many ways, with the threats that Jennifer was mentioning on the edge. We have issues of social justice and equality and equity as well. In this last year, we've seen the beginnings of—and we fear the ultimate impacts of—the digital divide in being able to gather information and then process and share it with our people broadly. In our universities, we see this with students who don't have access to broadband when they have to do all of their education online, but I really am concerned about it in our K-12 education, where there's just a tremendous difference in ability, now that whole cohorts of our fellow citizens are really left behind. As we move forward with data science and developing increasing capacity, those who are not on the train essentially are left further and further behind. That's something that we have to think of very actively. In the past, we'd tend to get things that are new and terrific and then we wait until the social inequities are well out of the bottle, as Jennifer was saying, they're really things that happened years before we addressed them. As we're moving forward in this new world where computing science and data science and digital media—all of those things—are so critical to us, we have to do it in a way that keeps us cognizant of our broad group of fellow citizens to make sure that they are able to move forward with us.

XLM: Yeah, that's incredibly important. Jennifer, do you have anything to add on that?

JC: I have a couple of things, the first one speaking to what President Drake said. Our division formed because of the data science major. This major went from zero to over a thousand majors in two-and-a-half years. We have 6000 students a year taking a data science class. As you said, many other universities and even two-year colleges are using our materials, but it's interesting because this draws a more disciplinarily- and demographically-diverse group of students. Computing sometimes tends to attract students who are almost preordained. They have gotten a better education in computing. They've heard about it in their families. So, how do we create an opportunity for a cross-section of California to come to college, without having realized that this might be something of interest to them, and discover their aptitude and affinity for data science? We're now an uncapped major, and we are just drawing a much more demographically-diverse group. So, that is on the educational side.

On the research side, there is so much that we should be doing around equity. We have two young professors—a professor of law, Rebecca Wexler, and a brand-new professor of EECS—that's Electrical Engineering and Computer Sciences—Rediet Abebe. Rebecca is trying to get someone released from death row, someone who was convicted based on the output of an algorithm. The output of an algorithm was used to convict them. It's stunning! Furthermore, it was used beyond the level at which it had been validated and tested, as Rediet

demonstrated. So, Rediet is working with Rebecca to challenge this. We're also working with the Innocence Project. We are creating platforms so that the average public defender will be able to challenge unfair biases in the software, and find exonerating evidence. This is such an important area. It is very much in the Berkeley and in the UC ethos. It is the kind of area where we really should lead.

XLM: In your building on those kinds of issues, are you thinking about having courses and research workshops? What's the systematic way of thinking about moving forward on this area?

JC: Oh, first of all, going back to our students because they are our most important asset, and they are the people who are going to be the future leaders. In the data science major, we have computing, statistics, human context and ethics, and a disciplinary emphasis. That disciplinary emphasis can be anything—there are 25 different disciplinary emphases. We are also setting up joint masters with many of the other schools, and we're actually even setting up a joint Ph.D. with UCSF in Computational Precision Health. So, absolutely, these connections have to form and, in fact, we need to have equity at the foundation of all of them, even the ones that sound more like science. How the science is used can be unequal if we don't think about it in advance.

XLM: And I think that kind of consideration is particularly important for those who will become future leaders of our society. Speaking of leaders, I want to ask you, President Drake, what role does data and data science play in your decision-making as the president of this gigantic UC system?

MVD: I actually base decisions on two things. I base decisions on information, and so data science is incredibly important to me to be able to draw information from a variety of sources. In my last meeting before this one, we were doing some research on our patient care populations, about where they're getting service, who's getting the service, etc. So, we need data science to come and collect information and share it with us. And then the other thing I use actually to make decisions is values. So, once you have the data, what are you going to do with it? The 'What? So, what? Now what?' is often a place we find ourselves in. So, it's collecting the data—I think that that's very important—determining how it's important and then processing it through. Really, I think one's core values allows you to apply the information in a leadership sense. One of the things about leadership is that you're doing things that haven't been done before. There's nobody to follow. You're doing things that are being done in a new way. You can't know what the outcome is going to be, because it hasn't ever happened. So, you have to base that on what you can gather from data and information, and then you have to apply it through your values and then work to get the best outcomes possible.

XLM: What you said is certainly very dear to heart for me, as a statistician, because we're trying to present the data and do the best analysis to help leaders to make their sound decisions. You already mentioned that, and I understand you're also a leading physician. So how does the data, data science play in your role when you do medical work?

MVD: You know, I'm really happy to be a physician. I love my career, and I watched data become so much more important to the daily practice of medicine during my time. There really was a transformation. When I was in medical school and just thereafter, just beginning my faculty days—some at UCSF, some in Boston—you go and read journals, you get the information, you'd learn things, and then you'd go and apply it to the patient, largely based on what you could remember and kind of put together what you knew and what you could remember. Whoever remembered the most was the one who was able to advance things. Then it changed from 3 X 5 cards in your pocket or a manual in the coat pocket, to actually being able to look things up on rounds. And so, it became who was able to ask the best question and to search the literature or the data for the best information, and you could do that in real time with the entire world of data at your fingertips. So then, it became asking the question, getting the information, and then interpreting and applying the data. The practice of medicine, even down to the patient level, has changed so much. Today, compared to a year ago, we do 50,000 telemedicine visits a week. In our system, it's not only that you have the data from the world literature—your interaction with the patient, in fact, is done through a computer. It really is a big part of our interface.

The other thing that I think is really important, that I saw and witnessed very much—we actually called it 'evidence-based medicine.' I remember giving lectures on this, you know, 20 years ago a little more. Then you had the concept: what was it before that? Because we could collect that data on every patient, and then thousands of patients across hundreds or thousands of hospitals, we could really use big data to determine if the patient outcomes that we were seeing were really appropriate, were really the best, and what was really good. Hospitals began to then evaluate themselves in ways that were never possible before. You had Mrs. Jones and she had her operation, and whatever happened is what happened. Now, you could compare her to thousands and thousands of other people who have been in similar circumstances to determine if her care is getting the best outcomes. Those things have really changed the practice of medicine from the individual patient to the entire hospital system, and really have improved the quality of care that we're able to give to our fellow citizens.

JC: I would like to add one thing to that. This is why we are starting this program in computational precision health between our new division and UCSF, because it is so complex and we are doing things that are much more evidence-based, but we are not drawing the conclusions on a large scale in the ways we would like to do

them. There's a lot that we are leaving on the table. I was talking with Matt State, who is the chair of the UCSF Department of Psychiatry, and he told me about amazing experiments they're doing in people with mood disorders where they're actually able to put thin film on the brain with 160 electrodes. They are able to monitor these people for a week. Shock therapy is a really crude treatment. With this, there are 160 electrodes, so they can do little things that can actually dramatically improve people's moods. But Matt said to me, "We're leaving 95% of the information on the table because we don't have the computing that we need to do this." Mood disorders are so difficult for individuals, families, and society. We begin to look at how can we actually treat them in ways that are really tailored to the needs of that individual. The world of medicine is so exciting. The data science possibilities are so exciting.

XML: I also want to mention to both of you that we actually have a special issue coming on individualized medicine, which is incredibly hard data science, because now you're talking about data science for individuals. I wanted to follow up on another question, because, Jennifer, you talked quite a bit about what Berkeley is doing, and I want to ask President Drake in terms of what are the other data-driven initiatives in the UC system?

MVD: We have programs really at all of our campuses. Jennifer is at our Berkeley campus, but I would say that, broadly, this is something that's extraordinarily important across our system. Just a couple of things that I'm aware of are programs that are taking place at the School of Information and Computer Sciences at UC Irvine, where I was before, to get people interested in data science and interested in computing. When we had a school of computer sciences, yes, the demography of the people who were first involved in the school was something that would have been relatively predictable. In our school of social sciences, we had a great focus on game theory, on the way that people and systems interact, using math to help predict game theory. We had a real strength in that. And we used those two together to look at actually gaming and computer gaming, and ways to use computers for entertainment and for this kind of interaction. We went from a circumstance where we had relatively few majors and were worried about the school, to being among the most popular schools on the campus with the greatest competition for slots. It really changed dramatically a school that was looking for its place in the future. This one particular program was a window to computational science and data science for a whole different demographic of students that we hadn't thought of before, so that's extremely important.

We mentioned medicine and the ways that we are gathering information in medicine and feeding that back to the medical practitioners to help them do a better job. We're also thinking a lot about doing that with education. Medical doctors and professors both teach, but nobody really watches you. It's really difficult to evaluate what happens to the students. Now that we could collect data, we could actually evaluate what happened to patients. And that led to the evidence-based medicine and outcomes-based medical decision-making. We're just starting to do that now with education and higher education. How successful is this course in helping you to move

forward? What happens the year after you took the course? We've all had teaching evaluations. It's a lot like, 'did they like you?' I'd say it's like if you were a bank, it would be a customer satisfaction survey, and it means that you have pretty ferns in the corner. You also want to know that you were writing loans. You want to make sure that you're doing the work of the bank. And so, we want students to enjoy their experience, but we also want them to learn, and so we can use data science to help monitor and support our professors. There are programs at UC Davis, in particular, looking at educational outcomes and wondering what we can do to help our programs work better. UC Riverside is working particularly with students who come from more challenged high schools and looking at interventions that we can have early with them, monitoring their progress, gathering information on that, and then feeding it back to them through advisors. There are multiple programs going on at all the campuses to be able to use data science to gather information and then to interpret and use that information to make us better at what we're doing.

XML: I know that at least two of your provosts are on my board. Other than Jennifer, Hal Stern is from UC Irvine, who I used to take courses from him years ago, and he's a wonderful statistician.

MVD: I appointed Hal Stern as Dean, so I appreciate that, so, very good.

XML: I see. So, he really worked for you before. You mentioned this issue about evaluating the students' performance. When I was Dean, that this one of the problems we worked on. Some faculty reminded me, 'Xiao-Li, all these course evaluations don't tell us what student remember 10 years from now.' What's really important is what they remember 10 years from now, 20 years from now. How do you do that, evaluate those things? I think you are absolutely correct. That's where potential data science comes in and the possibility is just tremendous there.

JC: There are methods to deal with the fact that not everybody learns in the same way and not everyone has the same context all the time. For online teaching, which we're starting to do more and more of, there's something called reinforcement learning, in which the algorithm can contextualize, and it can prompt the particular student at the time with what they need. This isn't following them later, but this is understanding and really gearing education to individual students. Now, I know that most of the students President Drake talked about in K-12 who are really suffering on the wrong side of the digital divide, don't have access to computers that could do this, but there may be a point at which we can intervene with these students and try to help them

get past the gap that this pandemic has created. I think that there are possibilities to use data science and computing in this way as well.

MVD: You know, I'm thinking of two experiences I had, one many, *many* years ago, the other one many years ago. One was, I had a fantasy—I don't know why, but this is what I was a teenager—we had maps, and the maps were so big that you had to unfold them and find where you were going and you never could fold it back up right. What a nuisance. You couldn't have enough. But I enjoyed reading them and thought, wouldn't it be amazing if you could get something in the car where you'd have a map? What I imagined as a teenager was like microfilm, or something that James Bond would use to help you find where you were going even when you didn't know. Let me say that that evolved now, so that in your car today, you touch the steering wheel and say here's where I'd like to go, and then a nice person tells you how to get there and shows you the map just like that. Another conversation I want to mention was a few years beyond that. I bought an old house and there was a guy painting it. I was talking to the painter and he said, 'You know, the most valuable thing in the world is knowledge.' And then he said, 'And what's going to happen in the future is that knowledge is going to be free. This most valuable commodity for everybody is going to be free, I mean, it's going to be that everybody can know almost everything.' When he talked about this 30 years ago, I thought that he needed some of those patches on his head, to bring [him] back to reality—and now we can reach in our pockets and have this incredible wealth of knowledge. One thing that really has spread across digital devices is cell phones and access to the Internet. As we look forward to what the future will be, I think we can expect that we can hold ourselves to the challenge of making sure that we are able to connect people in ways that we almost cannot imagine today, and to allow them to be engaged in our future in ways that maximize the human potential. So great to see things changing and know that we're on a path to continue that.

XML: You are absolutely correct. It used to be like just fantasy, now it's all reality. That actually really brings me to my final question. We have been talking about what UC can do, what Berkeley can do, for data science. Now I want to ask what data science can do for UC, and what data science can do for Berkeley? So, if you have this one wish, like something data science can do to make your job as the president easier for leading UC, what would that be? What would be something that can help you?

MVD: Yeah, I would say that I always look through any advances that we're making through a lens of equity and social justice. To make sure that's not something that we come back and reach for later on and try and bring in to our thinking. But that as we are moving forward, as we're developing AI and using it more and more, that we always have a lens to equity and to social justice.

XLM: So, you're saying that using the advance of data science to monitor situations to make sure we're doing these things in a very equal way for our society?

MVD: I quoted Clark Kerr a little earlier, now I'll quote Lin-Manuel Miranda. You know, there's a line in the play *Hamilton*, a song called "In the Room Where It Happened." The concept is that there are places where decisions are made that change us and society as we go forward, and power in the 18th century was to be sitting around the table where those decisions were made so that you could influence that future, you could be in the room where it happened. When we're all in those rooms and those things are happening, I just want to make sure that social equity is in those rooms as well, that we bring that to those discussions and to the work that we're doing. So that would be something I'd ask for data science.

XLM: Thank you very much. Jennifer, anything you would ask for Berkeley?

JC: There's something I would ask for the world. We've discussed social justice, which is something that we have to make sure that data science helps rather than hurts, just as President Drake said. That it will allow us to always ensure that social justice is at the table. We've talked about biomedicine and health. What we haven't talked about yet is climate and sustainability. I think we are all in a lot of trouble if we don't figure out how to tackle climate change. There are so many aspects of climate change that are data-driven but not yet integrated. There's the chemistry of new materials and there's environmental science, but there are also all the economics of alternative energy sources. There's understanding the human aspect of it. Bill Collins, who is one of our great climate modelers here, and he's been on all the IPCC reports, says the piece of it that we don't understand is human behavior. So, I know this is a lot, but I am hoping that we can integrate over all these disciplines and start to figure out a way to make more intelligent decisions on climate change, to integrate the environmental justice and the economics and the pure science of it and the geopolitical aspects of it, to begin to understand how one part of the system relates to the other. Because, for many of our students, this is the problem of their lifetimes. I hope that data science will help us to find the solution for them.

XLM: I'm quite sure that Berkeley and, more broadly, the University of California system will be taking a great lead in all these efforts. Thank you both again, and I hope that not in the too distant future, that [you] will come back again. We'll hear more about your aspiration and your achievement in leading the data science in higher education. So, thank you so much. I really appreciate it.

JC: Thank you, Xiao-Li, for the opportunity, and thank you, President Drake, for joining us here. This is really wonderful for us.

MVD: Great to see you both. Take care.

Disclosure Statement

Michael V. Drake, Jennifer Chayes, and Xiao-Li Meng have no financial or non-financial disclosures to share for this interview.

©2021 Michael V. Drake, Jennifer Chayes, and Xiao-Li Meng. This interview is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the interview.