AMS American Mathematical Society
201 Charles Street | Providence, RI 02904-2213 USA

Public Awareness Office
paoffice@ams.org
www.ams.org

Transcript of Mathematical Moment #150—Algorithms with Cathy O'Neil

This video is available on YouTube at https://www.youtube.com/watch?v=2dnl2vXGQxk. Although we have made every effort to ensure that this transcript is as accurate as possible, we cannot guarantee 100 percent accuracy.

Mike Breen: I'm Mike Breen, Public Awareness Officer for the American Mathematical Society, and I'm talking with Cathy O'Neil, who's the founder of ORCAA and the author of "Weapons of Math Destruction," a very popular book, and we're talking about algorithms, or really their unfairness. So, Cathy, you know, a lot of us know algorithms from high school or grade school really, and sometimes we think the only way to do multiplication is the algorithm we're told, but but really that's not true, but algorithms nowadays figure in much more to people's lives than just long division, for example. So can you tell us, really but just to read your book, you know, how prevalent they are and what their flaws?

Cathy O'Neil: Sure. I mean, and by the way, when I say "algorithm," I don't mean the division algorithm. I mean, sort of, the big data predictive algorithms. So they are predicting the future based on historical data and some definition of success. And so, basically, they all come down to a certain type of form, which are, like, given the initial conditions of this in the past, we saw that any either it succeeded or it didn't, justify relative to a very specific definition of success.

And then we … you sort of train an algorithm—the predictive algorithm—to see patterns—pattern match—so that when they are presented with a set of initial conditions, they decide basically is this likely to succeed in the future? So there's a time concept to it. So there's historical data that said, "Here's a bunch of initial conditions, here's the guess." Did it lead to success or not? And most predictive algorithms are…essentially set up a scoring system, what's it … which you can think of as kind of like a probability of success.

So how are those used? The answer is: in pretty much every bureaucratic process, like, especially the ones where people don't like to be blamed when something goes wrong. So some things like messy, or complicated, or has a lot of judgment—if it was just purely rules-based wouldn't need predicted algorithms. If it's a simple rule, like everyone with the following, you know, everyone who applied gets to do this—that's not an algorithm, you don't have to predict anything, like you just…everyone who wants to vote should be able to vote—that's the kind of thing that this should be a rule (not to say that it is a rule in this country). But it's much more when it's a value judgement so it's like what I want … who should get a loan, if they…if there's loan applicants. Who's, like, who "deserves" a loan? And that's the kind of thing we're like…we're going to predict who is likely to pay it back and then we're going to say that's the definition of somebody who deserves it. Who deserves insurance and how much should they pay for insurance? Who deserves a job and how much should they get paid? Who deserves a raise? Who deserves to get fired? It's that kind of thing that predictive algorithms are being used for.

So as I just mentioned a few things, it's like any kind of…any touchstone that you have…any interaction you have with enormous bureaucracy, used to … used to be a human, so that, you know, looked over your application or record, or whatever it was, and decided whether you deserved something. Now it's computers, and, and it's problematic because we are problematic. I mean the idea, sort of initially and maybe 10 years ago, was that "Oh, computers are not biased, so they'll do so much better a job." But, of course, what they're doing is they're training on our historical data, so they're going to repeat whatever we did. I mean it's one of the problems…that's not the only problem…the other problems are, like, the data's really bad sometimes—not just biased, it's missing—and that's certainly true for any kind of, what we call "crime data." I don't like to call it crime data because it's not, it's not a good enough proxy for crime, to call it crime data, it's really arrest data. So it really just is a record, an artifact, if you will, of how

police record things in the official record. But there's…so there's missing data problems, they're just bad choices for what definition of success is.

Like Facebook's definition of success is like what keeps you on Facebook. It's not, you now, there's nothing to do with what we would define as our…as what success looks to like to us on social media, because it is, as we learned, all too…you know, all too recently, what it does is, it privileges upsetting and shaming and outrageous misinformation and fights, because fights keep people on Facebook. So, it's not…it's not that surprising when you think about it, but anyway the point is that it's not just biased data and historical bias, which is a huge problem, it's also all kinds of other problems. I can't even list all the problems that this happens. But I guess the high-level thing I wanted to explain in my book was that there's nothing…there's, there's no reason for us to essentially trust these algorithms. That, you know, even though they are touted as mathematically sophisticated, well there might be some mathematical sophistication in that pattern-matching underlying mathematical structure, if the data is bad, the definition of success is bad, or just like the process has no oversight—no appeal system and mistakes are not caught and not fixed—then we shouldn't trust it. And yet we're being asked as a society to hand over these very, very important—critically important—decisions to the Machine and say … and to expect them to be somehow perfect.

MB: Yeah, and I think, I remember, I mean, you probably know a lot of the problems with them, but one thing I remember is that—I think you called them proxies—there are things that sneak in that don't…aren't really relevant, but are taken as very important to the decision.

CO: Yeah! I mean, the example I already mentioned is this … we talked about arrest data as proxies for crime. I mean one of the … such a … it's so bad. There's like, that I have, like six different thought experiments to explain how bad that is, but, like one of them is my… just imagining what our society would actually look like if we…if we had a digital artifact, i.e. data, for every crime that actually occurred. You know, so like sometimes when I give a… give a talk, I'll ask the audience to raise their hand if they know anyone who's ever smoked pot but not been arrested after that. You know, I mean, if…yeah, everybody knows that.

People commit crimes all the time and don't get caught, so there's an enormous amount of missing data, and I don't just mean pot smoking. Like, most, like only a small bit over half of murders lead to an arrest, and less than half, if the victim is Black. Very, very few rapes are reported, relative to how many occur, and of the reported rapes, 7% lead to an arrest. And those are the crime categories I can name off the top of my head, like, they're just…it's just…like, the "crime data" doesn't exist. We have arrest data, and the secondary issue, which is not very secondary right now in the…in the context of the George Floyd protests and, and, the police practices under scrutiny, is that the missing-ness isn't equally distributed, the missing-ness is much more missing for white people, for white crimes, than it is for Black crimes, and so you get to these kinds of conversations with people who will defend profiling of Black people by saying, "Oh, Black people commit more crimes," and they're getting it wrong, because they're confused by what we are doing … what the police are doing, which is swarming Black neighborhoods looking to arrest people, where they just simply would not get arrested doing exactly the same thing in a white neighborhood. So that's…that's the worst possible proxy.

Now, don't get me wrong, there's a lot of bad proxies, like another chapter I spent an entire chapter talking about they used the *US News and World Report's* college-ranking system, and they're sort of like the original, you know, big data model to be honest. Like, especially the most influential … it's an early example of an influential model. Influential is simply because people care about these rankings, like we don't … they don't…it doesn't have any official power, it just is the power that we bestow on it as parents and as students, mostly as parents, but it's like the quality…it's supposed to be measuring the quality of a college, and it just uses terrible proxies for the quality of college, like how many people…when I say terrible they're terrible in large part because they're game-able, and they are gamed, they're widely gamed, and the more they're gamed, the less power they have as quality proxies, right, proxies for quality of college. So, things like, you know, how many people get accepted who apply. Well, just get more people to apply, especially people who won't be accepted, and you can game that, that's what we see. Or, how many people who apply who get accepted actually show up, or how many freshmen, you know, drop out.

All these things have led to unbelievably awful gaming. And that, of course, the one thing that isn't measured as a proxy for quality is cost, like as if we don't care about cost or something, and the consequence of that has been it as the administrators have gamed these models, cost has skyrocketed and it hasn't affected their standing that they're ranking because they're not measured on that, so there's….for the gaming, for the purposes of gaming, like tuition was irrelevant so it became very large. So that's another thing you see, sort of like squeezing a balloon, like they're gaming these metrics, and the other metrics are going nuts, and that's perfectly okay. But it's not okay. I mean, the point I was making in that chapter is like, it's been hell for parents and kids, and I have kids going to college, and it's like, it's unbelievably horrible and it goes…it sort of propagates backwards and forwards, I would say like backwards in the sense that like even in middle school and even…even before middle school, some of my kids' friends were already under pressure to like, look good for college. And for that matter, like you get to college, and these kids are so trained to care about their grades that they…they take classes that are too easy so they can get good grades. It's like, what…at what point do we get to challenge ourselves intellectually and become, like thinkers? You know, it's just a terrible, terrible system.

MB: Now, I know, you…it hasn't been a, you know, a decade since you wrote the book—it's just a few years old—and then…but have things gotten better since you've… you published the book, and, you know, other people are aware of the problems with the algorithms?

CO: Well, I guess I would say "Qualified somewhat," in the sense that I think it's become a lot more prevalent as an issue, in large part because everyone is under so much algorithmic power, like they are… they are more and more realizing how much algorithms control their lives and, in spite of the fact that one of my biggest concerns in writing the book is that the algorithmic harms are hard to measure, examples do ever…every now and then show up, and so those examples have been helpful in getting sort of more of an awareness by the public that these things could be really problematic, especially along the lines of racial bias that has become more obvious. I worry that things like penalizing the poor, which I think is just as prevalent—and I could say why—if not more, mmm, probably just as prevalent as racial bias in algorithms, it doesn't seem to be as scrutinized.

So, on the one hand, you have better awareness, so that's a good thing, but the bad thing is like, it hasn't changed much. Look, I haven't seen algorithms get vetted. I haven't seen algorithms like actually stopped being used. But it's a step. I'm not saying…I'm not saying I'm disappointed, like the…when I was writing the book, I did not know anyone else who was thinking about this problem, and now there are large communities of people thinking about these problems, and it's really, really satisfying to see that and it's good. Having said that, like nothing, nothing, nothing has happened at the federal level. My book came out right before the election, the 2016 election, and like, I was all set to get a job at the CFPB, the Consumer Financial Protection Bureau, which is what Elizabeth Warren set up, and like talk about racist credit modeling, you know, and like that didn't happen. Nothing…has… like, those, you know, the federal…at the federal level, these kinds of issues have just been completely ignored.

MB: So, what…from your opinion, what should mathematicians do, or…or students who are interested in maybe trying to fix all this? Are there any courses they should take, who would like to, you know, make things better? … Is that a big question?

CO: Maybe ethics. You know,…my goal isn't, by the way, to be clear, it isn't to get all mathematicians and data scientists to be also ethicists, because ethicists exist and they can they can do that job. But mostly my goal is to make sure that mathematicians and data scientists don't think that they don't have to worry about ethics. You know, that they should be like, oh, wait, what I'm doing is implicitly making…is implicitly solving a trolley problem, like I shouldn't do this by myself. I should get a public conversation around this before I lift another finger towards analyzing this data. And what's happened instead is, it…like with crime risk scores, for example, like somebody will…the people who hire the mathematicians, data scientists, to build a crime risk score will say, "Well, they're the PhD, so they must know what they're doing we're just going to use whatever they make and we're going to trust it, because of course they're so smart." Whereas the person building it is like, "Well, I'm not an ethicist, so I don't get…you know, I don't get to decide how to use it and it's not up to me—I'm just building it. I'm just following the data." And so you, if you follow the logic there, nobody in particular is in charge of really

important decisions that could be deciding whether someone goes to prison or doesn't go to prison, or how long they go to prison, or whether they get parole and no one in particular seems to be in charge. So, like, that's a problem.

I don't think the solution is make mathematicians or data scientists become experts in everything, but I do think they absolutely must say…like, have some kind of…really what I want, by the way, is a society for data scientists with like…with, with an ethical board where they have…where someone who's being asked to do something that is really beyond their expertise, like an ethical choice, can say, "No, that's not my job, like you…you need to follow the following process, like some kind of review board process, and if you're not doing that, I refuse to do this job." That's the kind of thing we need eventually. Right now, we're just sort of, yeah, we're just sort of expected to make problems go away as data scientists, so that nobody has to actually have the hard conversation.

MB: Cathy, is there anything you'd like to add, any big thing that we missed, you think, on this topic?

CO: I mean, I would…I should probably know of like the different programs across the country. I know there's something in Berkeley [University of California, Berkeley] that already offer kind of data science with ethics…with ethical considerations, and there's stuff popping up all over the place, but I want people to be aware of the fake ones popping up as well. Like, so, there's a lot of what I call "ethics washing" going on, and it's, you know, like, where we're instead of talking about like what is happening right now with…with actual algorithms being used with scoring systems for people to decide whether they deserve something or not, right now, there are conversations about, like future human, like machine consciousness and you know like abstract future problems. That's…I'm not saying that those aren't interesting questions, and like, they're almost philosophical questions about the nature of like, our relationship to machines and stuff like that—fine. But they're not…they're not current practical ethics for…for…for working data scientists, which is desperately needed and sometimes not really done. Sometimes we're pretending we're talking about ethics, but we are actually just talking about philosophy. You know what I mean, like abstractly, rather than practically.

MB: Cathy, thank you very much for taking the time to talk with us, I know you talked about that you're moving so that's a lot of fun, and so, a half hour is precious to you probably. Thanks very much! That's … that's Cathy O'Lee… O'Neil, who is the founder of ORCAA, which advises people on algorithms, right?

CO: Yeah I audit algorithms, which is, I…hopefully going to be a new field, hopefully a lot of data scientists can go into that.

MB: And, and I'll say again that Cathy's the author of the "Weapons of Math Destruction." Cathy, thanks very much and good luck with the move.

CO: My pleasure. Thank you very much.