Quantifying Forgiveness
MLTALKS
FEBRUARY 16, 2018

JOI ITO:
Hi, everybody welcome to MLTalks. Today our guest is Julia Angwin. I think not a week goes by when I don't hear somebody say, we need more Julia Angwins. But there's only one. And we have her today. And Julia is a data scientist and a journalist and it's a really important combination, as you'll soon find out. And she's a Director's Fellow of the Media Lab and working with us and has been helping us for about a year now I guess. And as usual this is being streamed and so if you're watching this on the Internet you can tweet at #MLTalks and we will towards the end be taking questions from the audience and from Twitter. So comment and feel free to ask questions.

But we'll start with some comments and the presentation from Julia. Thank you.

(Applause).

JULIA ANGWIN:
Hello. It's great to be here. So as you can see, my talk is called "Quantifying Forgiveness," which is sort of a strange title for a talk. So I'm going to start with just actually a little bit of background about who am I, why am I standing here, which I think is always a little bit helpful. And then talk about forgiveness. And then talk about quantifying forgiveness. So I'll start with just me.

I grew up in Palo Alto. And I really probably thought I was going to end up at a place like MIT. I learned to program in 5th Grade because Steve Jobs was teaching everyone in the public schools in Palo Alto. This is my first computer. I worked my summers at Hewlett Packard. I was really ready to go into the personal computer industry, which is what it was at that time. And I took a wrong turn. I fell in love with my college newspaper and decided to go into journalism. And I thought, well, I'll just try it for a few years. And maybe I'll go back to like the real world of computers. Because when I grew up in Palo Alto, there were really two life choices. Hardware, software.

And I was pretty much a software girl. But I didn't know there were other choices. So journalism was my rebellion. I ended up at the Wall Street Journal. I joined in 2000 during the Dotcom boom, which is I guess now ancient history. But it was hilarious. They are like you know computers? We will hire you to cover the Internet. And I'm like anything in particular about the Internet? And they were like, no, everything. I was like, okay, seems fine. They just couldn't get enough people to write about technology.

I was there for 14 years. Then I went to ProPublica, where I am now, which is a nonprofit journalism startup, which if you don't know about, it's investigative journalism started by the former Managing Editor of the Wall Street Journal. He left when Rupert Murdoch bought the paper.

So I want to tell you about forgiveness in the real world based on my experience as a reporter at the Wall Street Journal. So I joined in 2000. I left in 2014. And I covered technology, Internet, whatever. And during that time, I achieved what journalists sadly consider -- their great dream is to get somebody locked up. You wrote such hard-hitting stories that somebody went to jail. And during this time two people did go to jail because of my reporting.

Strangely they were both black men.

Now, how many black men are in the technology business. Right? Of all of the executives that I wrote about, it is surprising to me that this is the outcome. So I'm going to tell you the stories.

So first of all, when was this, 2003, spam was a really big deal. And so I was like, I'm going to find a spammer.
And this is, you know, exciting stuff. So I worked with EarthLink, which was looking for a certain spammer. I tracked this guy down, found him at his home in Buffalo. Knocked on his door. He didn't answer. Yelled at me through the door. But I wrote a big story about the hunt for spammers and how it's very difficult. Eventually he was charged and sent to prison for the maximum sentence for 14 counts of identity theft for 3.5 to 7 years.

So you know, everyone at my office was very excited. I felt it was a little weird honestly. But I was young. I was like, okay. This is journalism.

A couple of years later, I'm writing about AOL. And I heard a tip somebody that inside was embezzling. I investigated. I found out there was a guy. He was the head of HR. Also a black man. He removed his pictures from the Internet so I can't show you a picture of him. And he was caught. And AOL had been trying to cover it up. So I wrote about it. And once I wrote about it, they brought charges. And he was sentenced to 46 months in prison.

Now, neither of these people were doing things -- they were doing illegal things. But think about what I wrote about that was the most illegal thing that I wrote about. The most illegal thing that I wrote about was at AOL, the round-trip deals they used to inflate their revenue so that they could increase their stock price. So they did crazy deals during the Dotcom boom where they would, instead of contracting for the cafeteria vendor, just to pay them to deliver food to the employees, they would say, actually we're going to have you -- overpay you and then you're going to buy ads. Because these companies were only being measured on ad revenue, not on net income, not on profits. So this was actually a scheme that inflated their revenue by billions of dollars. They paid $300 million in fines.

And they are all doing completely fine. Okay?

So Steve Case is worth 1.36 billion. He invests in all sorts of good causes. Dave Colburn, who lead all those deals, actually is bringing lots of investment to Israel. And Bob Pittman, who actually was the architecture of it all, is the Chairman of Clear Channel, which is a huge outdoor billboard company. Right? So like you guys know this story. In your guts, we all know this story. This story is the story we all know, which is some people are forgiven for their crimes and some people are not.

And they kind of have similar traits. Some of them are white. Some of them are black.

And that's just my anecdotal experience. But there's an enormous amount of data that supports that. And so that's my personal story of forgiveness, which is, I feel bad about it, for participating in this. And I feel sad when my fellow journalists want to get together and crow about who they got put away. I don't want to participate in that.

So I started investigating forgiveness in the digital world. Because actually, the weird thing about automation and technology is it is auditable. So we can see systemic bias in a way that we can't really see in human minds.

So I'm going to tell you about two different investigations I've done that have led me to some conclusions about algorithmic forgiveness.

The first is -- oh, first of all, just I forgot, algorithms are very important. You know about them. They are in your lives all the time. This is the Facebook Blue Feed, Red Feed, which if you haven't seen, is a really great project by the Wall Street Journal. It just shows you what your Newsfeed would look like in a blue state or red state basically in terms of your political leanings and how different your news looks.

So an algorithm that I looked at is this one this predicts the risk of recidivism. It's used across the nation in criminal justice to decide whether you're likely to go on to commit a future crime.

And it asks you a whole bunch of questions. And they are input into some software, /it spits out a score 1 through 10. Are you risky or not and then it's used for pretrial, whether you're going to get out on bail. In many states it's used for sentencing. And it's often used for parole. And in some places in California it's used within the prison system to sort you into medium or high-risk prisons. It's one of the
most popular. There are different risk assessment tools in use in the criminal justice system. But this is one of the most popular ones. It's a proprietary software. Not Open Source. Not inspectable.

But I wanted to look at it. So we went and fought a FOIA battle in Florida and got the records of 18,000 people who had been scored by this program over a 2-year period. In Broward County when you're arrested, every person who comes in for booking gets scored. And that's entered into the system.

Interestingly -- and then the pretrial judge looks at it when he is making the decision about whether to release you out on bail. Interestingly, everyone in Broward County that I talked to had no idea they were being scored. So they were just asked questions at intake. But they didn't know it was going into a scoring system. And the score is not described or discussed in the pretrial hearing. The judge just gets it as information to be used.

So the first thing that we did after fighting a five-month legal battle to get this data was just to look at it. What does it look by race, for instance? Since we know race is a big issue in the criminal justice system. This is what it looks like.

Basically black defendant scores on the left are -- were steady, 1 through 10 pretty evenly distributed. And white defendant scores were strangely clustered at the low end. Right?

So we thought, okay, if we were lazy, we could write a story right now saying this score is biased. But the truth is, who knows. Maybe every one of those people in the low-risk category is actually Mother Teresa. They were picked up for littering. And they are the greatest people on earth.

So we had to do a very sad thing, which was we had to look up the criminal records of 18,000 people. And their criminal outcomes. So basically what we did was we found everyone's criminal history and we also found their actual recidivism outcomes. So we had to drop a lot of people from the sample because not everyone had been out for two years. But essentially we got down to a sample of 7,000 people for whom we had full records, meaning we had their criminal history and then we also had two years worth of days that they were free. Because we took out the time that they were incarcerated for jail or for prison and added up, do we have a two year stretch?

So then we had this very nice sample, which by the way required an enormous amount of blood, sweat and tears. Terrible amounts of blood, sweat and tears. Joining databases on name and birth date is a task I would wish on no one.

There were typos. There were terrible aliases. All sorts of terrible issues.

But Broward County was very helpful because they had wanted to join these databases forever to see if their score was working. But they didn't have the time or interest in doing it. So they actually handchecked for us 1500 records of missed names and birth dates.

So in the end, we had I think 9 months, 10 months after starting, we could run our 5-minute long logistic regression. Which is the fun part. And what we found is that if you controlled for all of the factors, so you -- basically if you don't know what a regression is, it's just a way mathematically to try to create like a balanced pair to see what would the equivalent people -- when you control for -- you remove all of these other factors, what would these people look like if they were similar.

Okay. That's a terrible description of regression. But anyways, close enough.

So we basically controlled for prior crimes, for your future recidivism, your age and gender. Meaning if you had two people who had those same exact things, the same prior criminal records, same outcomes, same age, same gender, what was the difference in scores? You have a difference that was stark. 45% -- black defendants were 45% more likely to be assigned a higher risk score with the same set of facts.

Now, the problem is it's really hard to write a news article that says 45% more likely. Editors don't like that. Readers don't like it. It's very hard to comprehend. What does 45% more likely mean? So the way to really describe this is in false positives and false negatives. So a false positive is somebody who was deemed to be positive, a high risk, but actually was not. So they were falsely accused of being high risk
of future criminality. And false negative is obviously somebody who is falsely accused of being lower risk but turns out high risk.

Then you see when you look at the false negative and false positive rates is that there's this huge disparity. African American defendants are twice as likely to be given a false positive than a white defendant.

And similarly, the white defendants are twice as likely to be a false negative than the black defendants.

And so what was super weird about this was that that problem with these scores was all in the error rates. The score -- did I forgot to put it in -- I forgot to put the slide. But anyways the score is 60% accurate for both races. So that's a pretty crappy record, to be honest. I would be fired if my stories were 60% accurate. But in the criminal justice system this was considered an okay finding. So we found it was 60% accurate. But all of the bias was in the 40% error rate so that one group was getting overscored and one group was getting dramatically underscored.

And what that looks like in real life -- and this is how I tell the story -- is I found people who had a similar crime and described their situation.

So here is a guy who is low risk, Vernon Prater. Got a 3. And Brisha Borden got an 8.

Now, let's look at their facts. So Vernon had briefly been -- first of all, they were both arrested for petty theft. Vernon had previously had two armed robberies and had already served a five-year sentence for armed robbery. The arrest he had for this score was he had shoplifted $80 worth of stuff from a CVS. And after this score, he went on to break into an electronics warehouse and steal thousands of dollars of goods and he's serving a 10-year sentence right now.

Brisha was also picked up for petty theft. Brisha was 18. And she was walking down the street with her friend. And they saw a kids bicycle in the front yard of a house. They grabbed it and tried to ride it. Down the street. The mom came out and yelled, hey, that's my kid's bicycle. She came back and gave it back. However, in the meantime, a very nosey neighborhood had called the police.

And so her -- she was arrested for petty theft. Actually they charged her for burglary also. But later I believe dropped it.

She was scored high risk. Now, her previous offenses I don't know. They are juvenile misdemeanors so the juvenile records are sealed. But I do know that misdemeanors are not usually armed robbery. So I'm guessing they were less than Vernon's.

And her subsequent offenses were none. Right? So this is exactly what a false positive and a false negative look like. She was a false positive. She was considered way more high risk than she turned out to be. And he was considered way more low risk than he turned out to be. And the thing that's weird about it is, in your mind, if somebody had said to you, what do you think these people are likely to do? You probably wouldn't have made that mistake. But the computer made the mistake because of the way its inputs are scored. Now, we don't know how they generate their scores. It's a secret algorithm. So they don't tell you.

I will tell you this, though, the night before we published, the company was very upset obviously about this story. And they said, okay, our secret equation is trade secret. You can't share it with anyone. But Julia, you can look at it. So they sent it to me. It was a linear equation with like KD for constants for the weights for the variables.

Well, I don't know -- how am I supposed to know if this is biased or not, right? And I would defy you even if you had those to prove the disparate impact. The thing is you have to analyze the outcomes to really figure out how this is behaving.

So really what was interesting -- there's a lot of interesting things about this. And there's been many papers on this work because we put out the data and the code for people to analyze. I encourage you all
to look at it if you haven't. But I think it really speaks to the idea that we think about bias. But what this was was unjustified forgiveness. Right? Actually this data, our intuition, was correct.

It's not the only part of the story. But it was actually a big part of the story was that these people were getting a massive break. And it wasn't justified.

So I think it's interesting to frame it around forgiveness. Because I think also that's intuitively what we understand to be going on. That's what I understand to be going on based on my own experience of covering the criminality of the tech industry, which is basically those three examples that I know about.

So I want to tell another story about another algorithm that we were age to quantify.

So this is an algorithm that predicts the risk of car accidents. It's the one that car insurance companies use to set your premiums. So insurance is supposed to be a risk-based metric where you contribute to the pool based on how much risk you're bringing to the pool.

So we decided to test that. Because in fact, it's been long observed that minority neighborhoods get higher rates. And no one has ever been able to explain why. The car insurance companies, say those neighborhoods are more risky. But no one has been able to measure it.

So we decided, my team, because we just hadn't had enough fun joining the criminal justice databases, that we would try another gigantic data project. So we went and actually worked with Consumer Reports, which bought us a dataset, which was 30 million quotes for car insurance by ZIP code across the U.S.

And we bought different driver profiles. And this we could have obtained by reading every car insurance filing in every state ourselves and calculating. But it was easier to buy.

And then what we did was we filed the public records request in all 50 states for the actual risk of -- actual payouts that insurers have made by ZIP code. Now, tragically only 4 states collect that data. So we could only analyze it in 4 states.

But we still had 4 states. So we looked at California, Illinois, Texas, and Missouri. And we compared premiums versus payouts for a single safe driver.

So essentially controlling the risk of the driver, what do you see in the difference between premiums and payouts? Because car insurance companies have this extra factor that they -- additionally to your safe driving profile, they choose to put a certain surcharge or discount based on your ZIP code. This is something that they are allowed to do. And so they base it on this idea that some ZIP codes are less safe than others.

I don't personally understand this. Because I don't know about you guys, but I do drive outside my ZIP code. That's the whole reason I have a car. But anyways I guess this is their fun times.

So basically we wanted to remove all things other than ZIP code and see what was the difference. And so we did this horrifying chart, which I'm sure if any of you have ever looked at it. We currently need some data visualization help. But we did the average prediction -- the average of the minority premiums over non-minorities. Oh my God. This is the worst. And looked at -- here. Let me go to the next one.

So basically the risk versus -- the risk is the x-axis, which is the actual payouts scaled from least amount to most. So the farthest risk is on the right-hand side. And then the premiums are on the y-axis. So the increase in premiums.

And what you see, the red, the linear line, is minority neighborhoods. They actually track risk. So the premiums go up as risk goes up.

What you see in most -- and this was just company. We did this per company. So this was one company in Missouri. This is GEICO in Missouri. But we saw this same pattern in almost every ZIP code -- in every company. What you see is an unexplained different -- reducing risk for white neighborhoods.
So what this showed was there was an unexplained discount in white neighborhoods that didn't track risk. And that was a very surprising result. Because everyone, again, thinks about bias. Right? But it was an unexplained discount.

And this is what it looks like in real life. So Otis Nash pays $190 a month for GEICO car insurance. He has had no accidents. He works two jobs. He's a really diligent father and a really lovely person, who I hung out with in Chicago. He lives in East Garfield Park, which I don't know if you guys know Chicago, but it's one of those really kind of rundown west neighborhoods that is filled with graffiti and trying to emerge but, you know, what we call the inner city.

Now, this is Ryan. Ryan lives in Wrigley Park. And he is -- it's a classic bars and yuppie people neighborhood. And he pays $55 a month for his GEICO car insurance, even though his spouse recently had an accident.

And the thing is that the difference really, a lot of it, was this base rate. So these insurers have set a base rate for property damage in East Garfield Park of $753 a year and in Wrigley Park of $376 a year. So literally twice as much in East Garfield Park. And when we looked at the payouts, they are actually lower in East Garfield Park than in Wrigley Park. Right?

So this is not explained by risk. And this difference in their prices is largely driven by this crazy difference in property damage base rates. And the reason is because Chicago weirdly tried to get rid of red lining in the car insurance market so they said no one can ever change the non-property damage rates by ZIP code. So they lump all of their changes into the property damage part of it. Despite the fact that the risk doesn't support it.

So once again, this is a question of forgiveness. Right? We have this gap. And this is a gap where we have chosen to give one kind of group of people a pass. So I guess I would like to challenge all of you when we talk about bias to also think about forgiveness. Because the data suggests -- not always but in these two particular cases -- surprisingly that it was an unexplained discount not based on risk. That was really the problem.

And so I think we should think about that as a society, that that's one way to think about the challenges we're facing. And I guess I would just want to leave it with the fact that I really am thankful in a weird way, though, that we're choosing to automate some of these biases. Because I think we need to collectively see them. And the ability to audit them is really powerful. Right?

And we have made change through these. California has forced several companies to change their rates as a result. And there's bills pending in other states as a result of it on car insurance. The criminal justice field is debating heavily the use of these risk assessment scores. So I'm hopeful that these kind of data can help change the debate.

Thank you.

(Applause).

JOI ITO:

Thanks, Julia. So I wanted to sort of just start with the last thing that you presented, which was forgiveness in the Chicago premiums. Well, first of all, what -- like who did it? Is it the data? Is it somebody going in there and being racist and changing the premiums?

JULIA ANGWIN:

Oh, yeah, sure. Well, what we found, and really we have the best evidence for this in California. Because the companies have to give more information there. But what we found in California, which I think is likely to be true in the other states, as well, is that actually the real problem was that in -- a lot of these whiter neighborhoods were rural. And there wasn't a lot of data.
And so they didn't have enough data to really make a true risk calculation. So they guessed. So in California what they did was there was a loophole in the law where they could string together a bunch of ZIP codes that were neighboring and use -- so it was like transitive. You could take your neighbor's risk score -- risk -- and put it in yours. So they were just transferring one low risk and assuming that it was spread around.

And so the regulators have stepped in and said they are going to have to work harder to justify their use of those neighboring ZIP codes risk in places where they have sparse data. But I actually think they didn't have enough data so they made a guess. And their guess was, look, these are a bunch of nice white people.

JOI ITO:  
Because the Chicago one is slight different because they probably did have data, right?

JULIA ANGWIN:  
Right. Yeah, so I'm not sure why. Because Chicago there is plenty of data. There's a lot of history in Chicago of red lining and maybe -- you know, one thing that's interesting when I talk to the insurers, because I've talked to them extensively about it, no one has ever said this directly, but there's been a lot of like, you know, Julia it's hard to change peoples' rates. They might leave. So I suspect there may be some like, oh, these people might shop around. So we want to keep it low.

JOI ITO:  
So the way they were just obscuring -- so it's hard to tell whether it's crappy data, bad algorithms, or just somebody hiding behind sort of this veil of data and fiddling and being corrupt, you can't really tell?

JULIA ANGWIN:  
I actually think it could be the lost liter.

JOI ITO:  
Just marketing.

JULIA ANGWIN:  
Yeah, marketing. You have to get those white people in because they are going to bring more people or something.

JOI ITO:  
Yeah, and in insurance as you're starting to poke into this, I was reading the paper by the use of FICA scores, the credit scores, for really shady things like targeting predatory product sales based on households and stuff like that. But it's really at the edge of the law. And in the stuff that you showed in the insurance, does that -- is any of that illegal? Is it regulated? Are they doing the right thing?

JULIA ANGWIN:  
Yeah, it's interesting. So what I learned about insurance is I don't know if you know this, they have an exemption from antitrust law. So Congress gave them an exemption. So they are only regulated by the states. And I think it's fair to say that a lot of states are not really heavily regulating them. California is the most aggressive regulator. Illinois has chosen -- I'm sure it has nothing to do that State Farm and AllState are based there -- to entirely not regulate. They don't check anything. I could start an insurance company tomorrow there.
JOI ITO:
They don't pay tax there, right?

JULIA ANGWIN:
Well nobody pays taxes.

JOI ITO:
Well, rich people don't pay taxes.
(Chuckles).

JOI ITO:
So I guess one of the things that's interesting is as you start to shine the light on these things, are the regulators responding saying, oh, wow, we didn't know, maybe we should do something about it?

JULIA ANGWIN:
We have had a really constructive dialogue with California. They are the most well staffed insurance regulatory office. Really they have tons of actuaries. I think they have hundreds of actuaries. So they take their job really seriously. And we have had some really good back and forth about the data and the real technicalities. And they took it to heart. Illinois, Missouri and Texas said, whatever, guys, thanks for your paper.
(Chuckles).

JOI ITO:
They didn't find the modern Untouchables there ready to take down crime?

JULIA ANGWIN:
No. But I think it has given some conversations. There are these groups of the National Association of Insurers and stuff. So I think it's been talked about. But it's a hard industry to make change in.

JOI ITO:
Yeah. And I think you were mentioning before, in a conversation we were having, that people have anecdotal evidence of this. But the data actually gave a lot of energy to the conversation. Do you find that true broadly?

JULIA ANGWIN:
Yeah. I think the one thing that's sort of depressing about a lot of my findings is that people are like, well, obviously.
(Chuckles).

JULIA ANGWIN:
Right? Of course the risk scores will be biased against black people and of course they are red lining. But yet, it's still important to prove it. Because even if you think it's true, we need the data to support it.
And so sometimes these stories can feel a little underwhelming. My editor is like, whatever, we know that. You know. And yet, I think -- the real benefit I think of we always release our data and our code is that I find that is what propels the argument.
Like we're in a world now where you can talk about whatever. But until you lob data over the fence, you don't get a real policy dialogue going.

JOI ITO:
I heard Cathy O'Neil, who is the author of "Weapons of Math Destruction," on the radio the other day using the term mass spleening. When you pick out the data and hit people with it, it's really hard. Because companies have been doing this forever. Even now -- Cathy was telling me -- no it wasn't her it was another friend, Sinda, was telling me that even just unemployment rates, we're measuring right now people who are looking for jobs that can't find them. But we are not including people with disabilities, people who have given up.

So if you look at those, it's actually going crazy. So the other part you were talking about visualization is how you present the data. And I think you as a journalist are presenting the data in a specific way to shine the light on the bad guys. But that's also really interesting and important and partially also where you get criticized, too, right? Because you obviously have a point of view that you're using the data to expose. And the company will come back and say, no, no, no. Her data only shows this and it doesn't show that.

JULIA ANGWIN:
Right. The insurers say, look, Julia, you're using the wrong data. You're using the average losses per ZIP code. So what we got from FOIA was the average of all insurers and all of their losses on average per ZIP code over a 3-year period. And they are like, our individual losses could be a huge outlier. However, they don't share those. That's secret data. So they are like, well, we have secret data that shows that everything is awesome. And I'm like, fine. Let's share that. Let's talk about it. But they don't want to.

But all of these data conversations always come down to that, which is like, you're looking at the wrong pool. That's why I feel so strongly about journalists collecting their own data. We need to know what we're looking for and go get it. Because received datasets, the people who collect them, there's a reason they don't collect it if they don't want to know it. So you have to go get it yourself.

JOI ITO:
Right. It's the don't ask what you don't want to know.

JULIA ANGWIN:
Yes.

JOI ITO:
I want to talk about some of your criticisms, too, because I think first presentation about the risk scores, that was such a huge impact that I think the company responded. And then academics responded. And then you responded. And I sort of want to go through some of these. Because it's actually interesting because the word bias also -- I'm teaching a course with Jonathan Zittrain. And actually we had Cathy O'Neil on last Tuesday about algorithmic bias. And bias can mean so many things. It can mean a point of view. It can mean unfairness. It can mean data that's skewed.

And so you know, I think one of the criticisms was that if you optimize for the thing that you were pointing out, which is the false positives, then the accuracy rate would go down. And you can't optimize for both. And the argument I guess from the company was, we're more concerned with making sure that we get the number of -- the risk of the -- recidivism risks right and the people who end up a little bit longer in jail don't matter as much. Is that roughly --
JULIA ANGWIN:

Yeah. Their basic argument was it's 60% accurate for both black and white defendants. And we have optimized the algorithm so that it's fair in its predictive accuracy. And we don't care -- we don't think your idea of fairness where you think this disparity in the error rates matters. Right?

And that's a point of view. And it's a point of view that's shared by that whole field. All of the risk assessment scores are designed this way. And it comes from a history that they have had.

But you know, if you talk to people in medicine, obviously you're not going to not pay attention to the false positives. Those are the people who died because your medicine is bad. That's a huge part of your decision in diagnostic tests.

And so I think it's like a semantic argument. We are pointing out that they have chosen a definition of fairness that has this disparate impact in the error rate. And they are saying, well, that's not a fair thing because if you change the error rate, you would change this optimization for fairness at predictive accuracy. But like, I feel like in the criminal justice content to say that you're totally fine with false positives when the whole point of due process is actually the default to innocence. And so I find that a really hard argument. But that is the argument that they are making.

JOI ITO:

Then there's the other argument that even with that, they are not as bad as those judges, is that --

JULIA ANGWIN:

Yes. And people say that to me all the time. You know what, judges are so much worse, Julia. You've got no idea. And I'm like, that is true. I have no idea. Please present me the data. And I will analyze it.

(Chuckles).

JULIA ANGWIN:

I mean, I think that is probably true of some judges and probably some are better. And it's a question of how do you do that controlled study. And I'm not necessarily the one to do it. I couldn't -- in the jurisdiction that I was looking at at Broward, they used the assessment. There was no controlled judge who wasn't using the assessment that I could compare outcomes with. And I think that's important academic work. But it doesn't in any way take away from the fact itself is biased.

JOI ITO:

And also it's kind of like fair who? Right?

JULIA ANGWIN:

Right.

JOI ITO:

And I think that's the weird thing about the word fair. Everybody uses -- like they want it to be fair for them. And it's sort of a weird question like just unfixing bias, this is slightly a philosophic question, like is your goal to eliminate -- like what are you solving for? Okay. You're a journalist so you're trying to be neutral and shine the light.

JULIA ANGWIN:

I'm objective, Joi.
JOI ITO:
But I wonder, you know, like if you have -- I think Cathy O'Neil was talking about child abuse. So you have these predictors now that try to figure out which families are beating their children. And so a false positive where you take a child away from a perfectly fine family. Or a false negative where you don't intervene. You have very different outcomes. Both horrible. I mean, obviously accuracy is important.

But assuming you're going to have -- you're going to lean one way or the other. How should we be deciding? What -- and then again, I think their view would be similar to the criminal justice. But our tools are still better than what we have now, which is we can't predict anything and we just wait for a phone call. Right?

JULIA ANGWIN:
Right. I think that's right. I don't know if you have read Virginia Eubank's book "Automating Inequality," but she's good on this point. She talks about the fact that these child abuse algorithms are themselves -- it's too small a lens on the problem. Child abuse and neglect is usually a symptom of poverty. And so if you were to bring some resources to bear to help the family, maybe -- that would probably be better. But instead, it's all about predicting this tiny narrow thing, which is actually really, really hard to predict. So predicting human violence is extremely difficult. One thing I didn't talk about is there was a score that Compas had for violence, they predicted violent recidivism. It was only 20% accurate. It had the same exact --

JOI ITO:
What does 20% accurate mean? That means 80% inaccurate.

JULIA ANGWIN:
That would be right.

JOI ITO:
So that means worse than a coin toss.

JULIA ANGWIN:
Yes. So predictive accuracy is when you predict it will happen, 20% of the time you're right.

JOI ITO:
So a little better. You can't compare it to a coin toss.

JULIA ANGWIN:
It's not the same as a coin toss. But it's not a good number.

JOI ITO:
Yeah. And 60% for the recidivism --

JULIA ANGWIN:
Is also not a good number. The industry gold standard by the way for criminal risk score is 70%. They think they are winning when they get to that point. But I would like to say that previously in -- I looked back in the literature in the '70s, psychologists used to make these violent predictions. They would bring in. Is this person going to be violent? They would interview them. And they were judged
to be only 53% accurate. So that was decided to be not good enough. And now we come up with this automation of only 20%.

JOI ITO:
Interesting. I think you mentioned it -- and Chelsea in the audience, she's gone around interviewing people. Your article actually spurred the creation of I think we're calling it HAUL Humanizing AI in Law. But they have been running around talking to jurisdictions and doing interviews. One of the things I think she found was the data is just crappy. Underpaid clerks entering data. And it's -- how much of it do you think is just that?

JULIA ANGWIN:
Oh, for sure. I mean, first of all, the data is crappy in the sense that -- the really large sense, which is like even what they are trying to obtain, which is the questions on the risk score are, do you live in a neighborhood where there are a lot of crimes? Anyone in your family ever been convicted of a crime? So already it's like anyone -- plenty of people had written before we did this analysis like this is obviously going to be very biased against poor minority neighborhoods.

Secondly, the outcome of what they say recidivism is is actually a new arrest. Now, arrest is not the same as a new crime. A new arrest is -- obviously people get arrested for all sorts of things. Chelsea and I were outside were joking that we could stand on the street corner and smoke marijuana and we would probably never get arrested no matter how hard we try because we are two white ladies.

JOI ITO:
I hope you didn't try.

JULIA ANGWIN:
No, we didn't try.
(Chuckles).

JULIA ANGWIN:
But we know there's overpolicing and overarrests in some communities.

JOI ITO:
I guess I have to be a little bit more technical in my terms because crappy can mean several things. It can mean just noisy. Or it can mean socially crappy, which is sort of what you're saying.

So I guess the question is, and this is something actually Karthik is working on, which is let's assume you had completely accurate data and that you were predicting 100%. Would it be fair?

JULIA ANGWIN:
Right. So that's a question I think that it's hard for me to answer. I personally feel very uncomfortable with -- I think that we should all really question the use of a future crime in the sentencing of a current crime. Like just on a philosophical level. Whether or not it's true. So I think that's -- that's a barrier we all have to cross as a society together and be okay with.

I'm queasy about that. I believe in human change and redemption. So I guess I'm -- I'm not really on board with that. But I think we have to make these decisions as a society. We have made a lot of really terrible and really good decisions together as a society.

JOI ITO:
Yeah. And the work that Karthik is working on is to try to stop focusing so much on prediction but to focus on causal inference and trying to understand the underlying causes and address them. And try to maybe lower overall crime or reduce income disparity, rather than just more accurately throwing criminals in jail. And the other thing -- and this is, again, something that Cathy O'Neil talked a lot about.

But if you look at -- first of all, there are two slides that were pretty amazing, which was the relationship between arrests and crimes. And I think she was saying something like homicides, only half the people are caught. And that most of the people who are committing crimes aren't arrested. And most of the people who are arrested aren't actually committing bad crimes /and the relationship between bad crimes and arrests are not correlated. But arrest records are what you're using for predictive policing. Obviously if police are being guided to neighborhoods where there are lots of arrests, they are going to make a lot of arrests. So they will find their share of criminals and won't catch you guys smoking pot on the corner. It's actually going to a self-fulfilling prophecy that you'll be a recidivism statistic because you're much more likely to get arrested if you live in a neighborhood with a high recidivism score. And then it becomes a sort of self-renforcing positive feedback loop that makes you into a criminal. So even if it were accurate, I think when you think about systems dynamics what happens is it just creates reenforcements, which I think confound the social reinforcements we already have, which is poor people don't get the opportunities so on and so forth.

Could these algorithms be actually not only just representing societal bias but amplifying them at some terrible rate?

JULIA ANGWIN:
Yeah, I think they are. And I think -- the thing that's slightly hopeful is in that moment we're in right now where they are just starting to be automated and amplifying, if we can catch them and diagnose the problem and we all can decide together that that's wrong and fix it. So there's an opportunity.

JOI ITO:
Let me poke the phrase you just said, decide together. How does that work?

JULIA ANGWIN:
Well, you know, democracy.

JOI ITO:
Okay. So we decided together on our President. We decided together on a wall.

(Chuckles)

JOI ITO:
I think one of the worries I have if you have a somewhat clunky situation outside and you're pointing out these biases, first of all you actually have to have people think they are bad. If you're some white guy saying, I'm getting lower insurance rates and I get out of jail free cards, what's wrong with this? So I guess the question sort of is do you think of yourself as a left leaning liberal person trying to hack the system towards your own personal agenda of making things less biased against the popular norm of society? You know.

JULIA ANGWIN:
No, I don't. I actually see myself as a data terrorist. I'm just throwing data. I'm like, guys, it's not what you say it is.
JOI ITO:
A data terrorist was said somewhere.

JULIA ANGWIN:
I think it was already. But I do take my role as a watchdog seriously and I see that as my role in life and I really enjoy it. I like to be the thorn in the side.
(Chuckles).

JOI ITO:
And you are.
(Chuckles).

JULIA ANGWIN:
I'm living up.

JOI ITO:
I was on a mailing list. And I won't say who. But somebody was arguing quite eloquently that we should just ban all algorithms and automation in the criminal justice system. Do you think that's too extreme?

JULIA ANGWIN:
I don't know I think the problem with the criminal justice system is -- having spent so much time -- as a technology reporter and then I went a year and a half basically in jails and prisons talking to people and people who had gotten out, the terribleness of it is so much deeper than algorithms. And so indefensible on so many levels that I guess I just don't think algorithms are the only problem. And I think it's a really complex problem. But I was so horrified. There's a trend now that you can't ever have human contact. So all of these jails I visited, you can only Skype with your relatives. You'll never be able to see them in person. Many jails are being built without any natural light and no outdoor space. And you can be in there for two years. It's just shocking.

JOI ITO:
But then there's -- being on a couple of foundation boards, I know that foundations and society likes metrics. And I think one of the things that both the Coke brothers and the left-wing have agreed on is incarceration is bad. That we're trying to lower jail populations. So we have foundations like the Arnold Foundation that are funding a lot of these risk scores. Because they do seem to reduce jail populations. And that that feels good to both the people who want to save money as well as people who don't want to see people in jail.

But we were talking to a judge recently. And I think this is another thing that Chelsea has been working on. But they are being let out with all of these conditions. With GPS ankle bracelets, curfews. One of the judges said these are kids who have gotten stuck with minor infractions because they are not good at following rules. And then the lawyers come in and bargain less jail time but with tons of rules that they are never going to be able to follow. So they are going to get dragged back in again.

So it's sort of interesting to see that as you optimize for a single score, which seems like a good proxy for bad because these jails seem so horrible, you may be just smearing the problem around into other places that aren't being measured. And I think as a data scientist that's also, to me, an interesting question. Because, you know, are you looking at the right numbers? Could you be reducing the problem to something that maybe isn't representing actually the real issues.
JULIA ANGWIN:

Yeah. I think that's a really good point. My basic feeling about these algorithms having looked at them for so long is that the reason they exist is because people need to tell the public that they are only letting low-risk people out.

So it's part of the movement to end mass incarceration. And that is a very good goal. And this is basically the political step that people feel is necessary to accomplish that goal is to tell the public, look, science is here. Don't worry. Science is on it. And science says these are the cool people. They will be out and you'll be safe. So it's a political story more than it is a data story. The data is there just to solve that actual problem.

JOI ITO:

Kate from the ACLU is here but she -- there was -- you're not from Boston. But in Boston there was an algorithm that really actually an MIT team won for scheduling the bussing. And actually the team that won, I read some of the stuff that they had talked about before. They were actually a very thoughtful team who wanted to go out and talk to the community and figure out what they wanted to optimize for and stuff like that. But they created an optimization algorithm for school bussing. But then the outcome of the algorithm was a terrible outcome where you had elementary school kids starting at I think 7:15 in the morning getting dumped out of school at 1:30. And the parents were in an uproar. And the Mayor's office initially said it was trying to optimize for high school learning outcomes. And later they said something like, oh, but we were also optimizing for costs. And maybe they were saying, oh, we would save them money and we would pour it into pedagogy or something. But in any case, a lot of people were blaming the algorithm. And I think the thing that Kate allowed me to write one sentence and co-author an op ed for her that she wrote. But it was -- the point was don't blame the algorithm. It's the political system that created the optimization that the algorithm was set for optimized for money over the convenience of the families.

So I think that's why I was kind of pointing at we should all decide. I think that's the really big metaproblem is we don't really seem to be good at figuring out how to decide. And I think part of what you're doing is you're using math and science -- not math and science -- I guess math and science and algorithms and data to make it so we can see what's going on. And reflect so that we can then inform ourselves and then decide things. And I think the problem, though, is that the deciding part seems to still be somewhat broken.

JULIA ANGWIN:

I agree. I guess the reason I keep like pushing back on that is that essentially I'm just really good at problems and I suck at solutions.

(Chuckles).

JULIA ANGWIN:

Let's just be real. I am really good at diagnosing problems and I guess I just want someone to pick up that ball and run with it. Like I have my skill set. But I do think that correctly diagnosing a problem, until we did this math, people don't know it was the optimization of the algorithm for fairness. So I'm glad we brought that to the table. And I hope that people can thread the needle from there. So I feel like my value and the work that I do and the work I hope more journalists will do like this and more activists is by bringing really quantification to these problems, it makes it addressable. I mean Facebook, everybody is writing articles, like Facebook is so bad. They are so big. That's not an addressable problem. The addressable problem is the thing I showed, which is like you can buy ads targeted to Jew haters on
Facebook because they had an ad category. And then they took that ad category away. So like I'm in the world of addressable problems.

**JOI ITO:**
I guess in the case of something like Facebook, it's their job to address the problem. So I think that's -- I think they are thanking you probably somewhere for your service.

**JULIA ANGWIN:**
I don't know about that.
(Chuckles).

**JOI ITO:**
But I think when it involves a political system, it's a little bit tricky. We're sort of at the half -- comments part of it. Does anybody have any questions? Maybe Kate.

**JULIA ANGWIN:**
There's the ball. It's a throwing situation. Oh, my God.

**JOI ITO:**
There's actually a little warning underneath that says don't throw at peoples' heads.

>> Is this on?

**JOI ITO:**
Sorry about that.

>> Hey, guys. I just have a couple of thoughts. One is it strikes me if we are producing risk assessment tools to, for example, say there was a risk assessment tool in the criminal justice context, that instead of determining whether or not someone would go to a cage or remain in a cage, would determine whether that person needed maybe direct cash assistance. Do you need help getting to court, for example? Here is a bus pass. So in other words, it seems to me that the risks involved with risk assessments can be substantially lowered, if not eradicated entirely, if the action that is taken at the end is something that it doesn't matter if there's a false positive or false negative. Because it doesn't hurt anyone to give them health care or a babysitter or a bus pass or something like that. And maybe we should start using risk assessment tools in those types of situations because it will help us get more data about how they actually work and stop using them in contexts where a false positive could be really detrimental.

**JULIA ANGWIN:**
Right. That's how they were used in Canada. So they were first developed in Canada. The people who developed them all had this intent is they are actually called risks and needs assessments. So what are your needs? And basically in Canada you -- they actually try to meet your needs.

When they came to the U.S., they still have the needs part of it but the judges that I've spoken to say, look, I only have three treatment beds for drugs right now and so I can't give it to all of the people who need it. So it's nice to have this needs section. But also the needs -- at least with Compas in green and the risk part is these giant red like high risk. You know, so judges are also really scared of being that statistic where they let the person out. So they are just guided by the risk portion.

I will say this, in the California prisons, they are only using the needs right now. And I went and spoke at San Quentin. And everyone knew their risk score immediately. And they were like it's good to get high
risk because you get more services. So they were fine. Although they weren't super happy when I told them about the bias in it. But then they were like, whatever, we all have high risk scores anyways.

>> But then nobody cares because you get to go to the gym for longer or you get special classes or something.

JULIA ANGWIN:

Exactly.

>> So that was one thought. Another thought was just like you said, Julia, the reason why I think we are turning to these tools is because of things like the Willie Horton problem for things who don't remember Massachusetts political history there was a crisis here when Governor Dukakis let somebody out of prison and he killed someone. And as a result of that, I think judges are terrified of the political consequences, especially in places where judges are elected of letting people out of prison so we as a society really have to change the political zeitgeists so that judges aren't relying on tools like this to deflect personal responsibility because they are scared of what may result from a bad decision.

Then I had a question, which is, how did folks in Broward County respond to your work? And did they actually change something about how they are using this system?

JULIA ANGWIN:

They were really happy. Because they were like, look, we were wanting to join these databases for a long time. So thank you for doing that work. And also, by the way, we are not going to change anything. And I was like, okay. They were like, look, the company that built this score says you're using the wrong definition of fairness. I was like, okay.

>> Wow.

JOI ITO:

I wanted to add one thing to what you were saying and I think we work the town of Chelsea. And they have this thing called the hub which is not only the police department but also social services and supports. They are not using the risk scores. But they are trying to address the underlying causes. This is Karthik's causal stuff. And there's really interesting -- like failure to appear could be like all of these different things. So I think when you're looking especially at the pretrial stuff, if you could just get one layer deeper and figure out what the failure was, then you could separate the people who could be helped by a little bit of bus money or maybe they need some medical support or maybe they -- they are out committing a crime. They all turn out as failure (inaudible). It's kind of like when diabetes was one one thing. So I think a lot of it could be address by having more data. But my concern about something like is that if we created a massive database that identified all of the needs and all of the vulnerable people, you could use it to help the people but you could also use it to discriminate against the people and to sell for-profit universities' spam to these people.

So that's I think the other fear that I have about creating massive databases to help people is you can use the same databases to hurt them, right? Do you want to -- sorry.

>> Yeah, thanks. This is pretty interesting. And it's pretty prerogative to think as a problem of forgiveness instead of bias and I think it has quite a bit of value. You were saying you were not good at finding solutions but identifying problems. That's the first step. That's good. But let's see if I can try to help in identifying some solutions. What you're finding -- you need the definition for fairness.

What exactly is fair? Well, if you identify that for a particular problem you have two communities whose experience follow two different distributions, like the case with the insurance companies in which
black neighborhoods had this kind of increase in premiums where the white had some kind of better forgiveness.

Maybe fairness would not be to actually make the white communities go up and start paying as much as the black communities but have as a policy that whenever you have communities having different experiences, just map to the better one. Maybe that's something that can be constituted as a policy.

What do you think?

**JULIA ANGWIN:**

I'm definitely in favor of more forgiveness instead of more punishment. I agree with you. I think when you're talking about company's profit margins, their likelihood of adopting that might be low. But I would do that. That's why nobody is letting me run any kind of profit company.

**JOI ITO:**

I've heard that a number of times that we're willing to try to be more fair as long as it doesn't cost us money.

**JULIA ANGWIN:**

Right. Exactly.

**JOI ITO:**

Which is actually weird to explain.

**JULIA ANGWIN:**

i think you have to throw it.

**JOI ITO:**

I'll go here and then behind.

>> Thank you. It was very interesting. It illustrated structural inequality in a quantifiable way. And the first question that comes to mind is when we do that, who do we serve and what do we serve? Sometimes measuring and rating and demonstrating is really the easier, even though it's very complex, as you said that. But the easier task when we look at constituting fairness. In order to be fair, not necessarily we have to be more specialized in analyzing data but more, as you said, more specialized in inventing new criteria, how can we resolve the issues. For example, can we quantify the economic loss to society for all of the biases that are being done? Can you do that? It's a question.

**JULIA ANGWIN:**

That seems hard. I would like to.

>> Would you?

**JULIA ANGWIN:**

Well, I think it's going to be really hard because there's so many compounding factors. But I would say that by sector, like I can do it for insurance, I can do it for criminal justice in some small ways. But also economic projections are a little different than what I do. So I wouldn't want to try necessarily. Because I'm sort of against the future.

Like I don't want to project -- I'm really into the ground truth. Like basically I'm like what, is happening on the ground? Can't I quantify it? And that's my sort of sweet spot. And there are people who are really
good at predicting the future and spinning out a story from the ground truth. But I'm sort of a specialist in the ground.

JULIA ANGWIN:
Don't you know they say you predict it by inventing it?
(Chuckles).
>> The price that we are paying is current for structural inequality.

JULIA ANGWIN:
Yes. I'm doing what I can. I'm just saying I'm doing what I can.
>> Maybe to you.

JOI ITO:
Behind you and then to Judith and then over.
>> Thank you so much. Joi, I would actually like to come back to a point you just made about data being able to be used in either way. Combine it with your point about solving -- thinking about solvable problems. Because one thing I keep thinking about in particular as you make your suggestions is how are data analysts trained nowadays? And I really see that as a key component to actually -- and a potential solution to the problem. Because if we obviously only train them in using data and maybe targeting it and tailoring it to the extent as they can, then we will never give them the opportunity to develop that conscience or the awareness of the wider implications. And to be honest, as sort of a side comment, as you were speaking about the legal system, I would actually argue -- and I have a legal background so it's not that I'm speaking completely off the top of my head. But I would argue that part of the problem is when you try to make the decision based on data.

So when you basically have people in charge who think about how can we come to a solution based on availability of data, then you get these weird outcomes. Well, if you think about it the other way around, in terms of what is independent of the data, availability of data that we want, then you can get different outcomes.

So I think to me, if I may make a suggestion, the solution would be to bring more social science education to the data scientists.

JULIA ANGWIN:
I definitely agree with that point. I mean, I look at it through the lens of journalism. So in journalism there's just -- you know, because the profession is so underfunded and under pressure but also because people don't choose to go into it because of math and literacy. It's only people like me who fell off the train somehow and got there.

So journalists are too happy to -- also to write about the available data. I always joke that there's three clean datasets, baseball, the Fed and polling, and wow, what does FiveThirtyEight write about? I mean, it's like really easy to receive datasets and then make a visualization and write a hot take. It's not easy. But it's something. But I believe we have to be artisinal and we have to basically collect our own data. What I do I think of what question I want to answer and then I think about what data I have to go get. It's a total nightmare. All the time my editors are like why do all of your stories take so freaking long? Well, they are artisinal so it takes a while.

>> I really like the idea of artisinal data but I also think there's an interesting side to this work a little bit less than what was explored. I was interested in hearing about the insurance piece because it seems in a rational world the insurance companies wouldn't be doing anything like this because it seems like it's costing them money to be treating people without equality. And so what I think you have is a very, very
interesting set of data to help us understand the motivations for some of these structural inequalities. And I think that's a really important thing to understand.

Because they are not just mistakes. They are systemic things that people are doing that they intend to be doing. I think it's partly -- certainly why you get so much pushback. Like thank you for providing us with this data. And now let's put it in a jar and go away.

So I think, you know, one thing that comes out of it is you don't have to look at it across all of society. But you can basically say, okay we can now understand how much are insurance companies willing to pay to be able to treat people unequally. And that's something we haven't really thought about in that way. What's its value to them? Why is it so valuable? Is there some reason why it's economically value? Is it something that's not economic that they are doing?

So I think that's an interesting piece of information to understand for itself. And I think it's also essential for understanding how we can change this. Because if we just look at fairness with the assumption that everyone's ultimate goal is to be fair and leave out trying to understand these motivations, we're not going to make much progress.

JULIA ANGWIN:
Right. I think you're right. Although, I suspect what they have done is just raised -- they are not actually going to lose money. So in order to give this discount which I suspect is some sort of marketing cost in their mind, they've raised everybody's higher so that line that looks linear would have been a lower premium to begin with.

But I agree with you. Like the economic incentives are obviously what drive these decisions. At least on the for-profit company side of it. And it's definitely worth exploring. And I'm working on more stuff along that line right now.

>> Yeah, and some of it may not be economic. There's a woman whose name I just spaced on who did a lot of interesting work on who did a lot of interesting work on -- she has a book called "Pedigree: How Elite Companies Hire," which shows they will systematically make a lot of really, really poor hiring decisions because of embedded beliefs of what kind of people they want there. So it may uncover that they are doing things that are actually to their own economic harm. But it has to do with their view of the culture and society. It's not necessarily a good economic decision they are making.

JULIA ANGWIN:
Right. I suspect that's true, also. I haven't yet met anyone in the insurance agency who led me to believe there's a person going, haha I'm going is to figure out how to get these people. I'm going to really screw them. I don't feel like that. I don't feel like that. I feel like it's a bunch of well intentioned people who were shocked.

JOI ITO:
I wouldn't go as far as saying well intentioned.

JULIA ANGWIN:
Some of them seemed really nice.

(Chuckles).

JULIA ANGWIN:
Although I will say, they invited me to speak at their convention. But I was like why are you inviting me? The whole industry trade group, they have a meeting in Texas for all their top lobbyists. And they said, we want you to do a keynote. I was like, are you sure? I asked like six times. And I was like, I'm going to talk about the work. They were like it's fine, it's fine, it's fine. And then I got there. And they
said, send the slides the night before. I fly in. I give them the slides so they could load it up on the 
machine. And they were like, oh, we didn't know you were going to mention names of companies. 
Because it says GEICO. They said, you can only speak if you take the names out. So I didn't.

JOI ITO: 
You didn't speak?

JULIA ANGWIN: 
I withdraw. And I sat in my hotel room above the ballroom and I did a tweet storm during my 
proposed session trolling them about how I wasn't speaking, I was supposed to be on stage. And it was 
called ProPublica. Like this whole talk was to be an interview with me.

(Chuckles)

JULIA ANGWIN: 
But anyways that wasn't your point at all. But . . .

I do think though that -- I don't really believe -- one thing that I think is a fallacy that sometimes is so 
easy and such a narrative that we all want to believe, which is if you find the bad person and root them 
out, the one who is making the bad decision. And I think it's oftentimes not that.

>> I don't think it's necessarily bad decisions. My guess is that it has to do with just fundamental 
understandings of risk. And just societal systemic assumptions about what makes something risky.

JULIA ANGWIN: 
Correct.

>> Hi. So there seems to be an objective of unfairness and then there's the subjective feeling of 
unfairness that may or may not correlate or show up in a visualization. And it seems quite clear that the 
system of Compas, for example, is being unfair. And at the same time that minorities in general feel they 
are being treated unfair by the judicial system. They don't really trust the judges or the lawyers they are 
getting.

Is there ever a consideration of, well, if there is this subjective component and there's a risk of, for 
example, breaching the fair trial right, why isn't it a right of the accused to define whether or not these 
systems are put in place since there is this subjective component of them having the right to be -- to 
believe that they are being treated fairly, not only objectively but subjectively?

JULIA ANGWIN: 
I'm not sure if I totally understand but I'm going to understand what I think or want the question to be, 
which is right now the way our criminal justice system works is all of the due process protections, which 
are the ones you think of of what is designed to embed fairness into the system, are only really required 
at trial. And nobody goes to trial. Right?

Pretrial is really the only decision. And then you plea. And so there are very few trials.

And so the due process requirement has been totally ignored during the pretrial phase. So for 
instance, people have argued that people should be able to contest their score during the pretrial and say, 
look it says I'm a 7. I'm a 4. Now the problem I don't know what that debate looks like, I'm a 4 not a 7. 
How is a judge going to adjudicate that? But at the very least you can have the conversation or you can 
have some other way to embed that discussion of risk into that.

But right now the defendant really has very little rights to fight that battle about their quote riskiness. 
And so some of the issue is just how to build more due process into what is effectively the judgment 
phase now.
JOI ITO:

Yeah. But I think there are some cases where I think like in Wisconsin where they tried to use due process to go after Compas score being used in sentencing. And to your point, they say, well, it's a secret. We can't tell you.

And it doesn't make sense. Because you couldn't say that if --

JULIA ANGWIN:

Well, the challenge -- is that the Compas challenge of due process I think is up at the Supreme Court now. But is that judges are really differential to other judges. So essentially every ruling so far on due process for risk assessment scores has been like, you know what, judges can consider whatever they want in sentencing. They can just not like you and sentence you.

And pretrial judges can consider whatever they want. Sometimes there's a bond schedule where you have to follow. But most judges have extreme latitude. And when it's appealed, the judge above them the judge is like, judges are awesome. They should really get what they want.

Over there.

JOI ITO:

I think Madaris had one and then we'll go --

>> Yeah. All right. We at the Media Lab have been working on new cryptographic techniques that will let you fuse data sets in a priority preserving way. For example you have datasets with criminal history and datasets that relate mental health nothing about each dataset gets (inaudible) but you still get detail about it. There's a natural question about weaponing on this. Because normally what you could compute on, you could also FOIA at least in certain things. And here this ability for you to peak under what computation was done under the rail would no longer simply exist.

Can you share some of your thoughts on what would be -- it would be like to do investigative journaling in a future like this?

JULIA ANGWIN:

Using homomorphic encryptions?

>> Using techniques like homomorphic encryptions.

JULIA ANGWIN:

It's really unusual. I'm really in love with math as you may have noticed. So I love stories that are like my friend at BuzzFeed just did, they have done a couple of these, where like statistically it's impossible that judges in figure skating are fair. The data shows that they favor their own country in ways -- there's no way for it to be explained but other than bias. Right?

But journalism is reluctant to do probabilistic findings. It's a difficult travel for me. Like I have to produce those people, Otis and Ryan. Like I have to have anecdotes. That's like what is the -- that's the currency of journalism is the narrative. So I love the idea of pioneering these areas where you're like, I can't see it but I know enough to know. But I think mainstream journalism is not quite there yet.

>> I asked Midaris if I could actually steal it since I was here. So my question is actuarial science is really expanding in the age of big data. And actuarial science is fundamentally based on this idea of risk. So my question is, is risk like fundamentally a reductionist neoliberalist concept? And if it is, is there an alternative concept that you would like to see data science orient itself around for modelling purposes? (Chuckles)
JULIA ANGWIN:

Whew. That's a hard one. I do think risk is often narrowly and politically defined. And people are unwilling to acknowledge that. So I think that is true. I still think it's useful in the sense that what I really love the most is the fact that I'm not expanding beyond the scope of risk. And I'm still showing these companies are not doing it. So I came to your playing field and I'm using your rules and you don't have it going on. That to me is still the best proof. I agree there's definitely like other ways to have those fights. But for me, like I like to wane on the playing field of the opponent.

JOI ITO:

Can you throw it to this corner and then throw it to that corner.

JULIA ANGWIN:

Whoa, nice.

Hi. So we're talking about structure and kind of like how insurance companies looked at ZIP codes and determined how rates are based off of that.

But I was wondering how they respond to any changes in the urban environment. For example, if the socioeconomic factors of an environment change and if they are -- the risk or the rates change over time in response to those factors. And also the economics like you were saying how those correspond to those dynamics.

JULIA ANGWIN:

The insurance companies are interesting because they are -- they were like kind of the first algorithm users but they are sort of really old school because of that. So their systems are really legacies. So they kind of update their rates every couple of years.

So every couple years they will file something new, like this ZIP code is adjusted this way and it's supposed to be based on the risk that they have seen in terms of what they have had to pay out in those ZIP codes.

Now that data is a secret. So all I can see is the average that everyone has paid out. But I don't -- I don't know how well they are policing it because it feels like it's pretty disparate. Right? The reality versus what they are charging particularly like GEICO was insane. So I guess I don't know how often -- because they don't have much public scrutiny -- they do follow the regulators but what's interesting the way the regulators look at it they are not looking at this question. They are asking a very different question of the data, which is basically their main question for an insurance regulator, the only thing you care about is do they have enough money to fund all of the possible claims or are they going to go under? So that's your main question as a regulator. So they are not really looking at this question.

So I don't know how often they check. And I believe that basically when people don't check their metrics, they fail to update them.

In the back.

So considering the fact that it might be very hard for us to get criminal justice systems to stop using this data to make decisions based on what we think people might do in the future, do you think it's possible for us to start using this data to get criminal justice systems to repay defendants for wrongs that the justice system has done to them in the past?

JULIA ANGWIN:

To repay them? Can you expand a little bit on what you mean by that like pay them back their bail money or the time they lost from being in jail?
I know like Canada and California are doing things like that for non-violent drug crimes. So are there things like that. Or paying back bail money or somehow finding a way. It's a little hard to quantify. But finding some way to repay the injustice that the justice system has done to a defendant that was wrongly labeled as very high risk and therefore not given a chance.

JULIA ANGWIN:
I think generally I'm philosophically inclined towards reparations. I think if you can quantify a harm and do right by the person who has been harmed, it's a good idea. So I think it's very complex in the details.

JOI ITO:
I don't know if -- what harm measurement would work in this case. But in many cases like torts and stuff I think the harm is on what your potential income would be. And if you're a poor person, it would be much lower than a rich person. And that would be unfair, as well.
But I think it's interesting to figure out how you might do a retroactive fairness.

JULIA ANGWIN:
Yeah. Over that way.

JOI ITO:
Yeah, she represents Twitter here. The Twitter community. Not the company.

>> I am Twitter. I have a question from Twitter that I'm going to combine with a question of my own, if you don't mind. So CJ on Twitter asks, what's your best practice for being such a thorn in the side of justice systems or unjust systems that they have to listen to the data lobbed at them? And I want to follow up with asking you about the experience with the backfire effect that this idea when people presented with data about their facts and biases or facts like climate change or racism, say, that's clearly not true. You have just reinforced my position to the opposite. I guess what I'm asking if you read the comments in your articles but he's asking if you found a way to go get through there.

JULIA ANGWIN:
First of all I never read the comments. That's my first best practice in life. No comments.

(Chuckles0

JULIA ANGWIN:
I have a whole Jihad about how I do journalism, which I'll give you a short version of. Which I believe that journalism has -- needs a guiding light. For the long time I was raised under the idea that objectivity was our guiding light and that became false equivalence. Everyone has agreed that it's no longer good but doesn't have a new load star. I'm arguing, mostly shouting into the wind, that we should use a scientific method as our load star. The scientific method is super nice because it's actually a little loosey-goosey when you really look at it. It's like, do you have a hypothesis? Do you collect evidence for it? And then, you know, do you have reproducible results? That's your goal. And those are my goals. And that's how I run my investigations is we come up with a hypothesis and then we figure out what are the tools and data we need to test this. And mostly we do lots of testing.
So I always tell a story about a research on Amazon. I had heard that there was price discrimination on Amazon. If you used a mobile browser versus desktop, you would get different prices.
So we set up these big experiments in the cloud and Amazon accounts and we were running it for months and the data was not there. There was no difference.

So then we were like we saw something weird between Prime and non-Prime. So we were like, okay, let's test that. So for months we were testing Prime versus non-Prime were there differences in prices? No. Sad.

Then I had given up on it. And three months later I went to a bar with a guy who is an expert on antitrust and he was telling me about how terrible Amazon was to the book sellers. And I was like, yawn. Then I was like I've been doing this test but I just don't have anything. He was like well the thing you need to test is does Amazon advantage itself when it's a seller versus third party sellers. That's the test.

So we went back and ran that. Because we already had all of the accounts set up. It was in the Amazon cloud running away. And boom immediate results. Right? That is like I have 7 of those going at any time. And most of them are total miserable failures.

JOI ITO:
Is this legal?

JULIA ANGWIN:
Oh, do we have to talk about that?
(Chuckles).

JOI ITO:
Okay, sorry.
Prank calls. No. Sorry.

JULIA ANGWIN:
We were just looking at prices on Amazon. That's legal.

JOI ITO:
Totally legal.

JULIA ANGWIN:
So I believe in the idea that you don't know what your story is until you've done the tests. Right? And most journalists get a tip and then they report out three anecdotes. And they are done. Then they go to the data desk and say, build me a visualization. And those guys say the data doesn't support your anecdotes and then they have a fight and data guys get sad and they quit. That's what happens. So I'm trying to build a new way of journalists and programmers working together. And my team is two programmers and a journalist and a researcher. And we are like four people. We work collectively from the beginning on these projects.

JOI ITO:
The is it legal part is only half a joke in that at MIT we actually can't do a lot of the studies we want to do because -- I'll just advertise this because I think some people will know about it. But there's a law called the Computer Fraud and Abuse Act that was created after the WarGames movie because everybody got afraid that people would hack into computers. And it says that if you use a computer in a way that is against the intent of the person who runs the computer and it's online, that is a felony that will throw you in jail. And Terms of Service has been deemed as a description of how the person wants you
to use the computers. So if you go on to Facebook and try all of these experiments, it could turn into a felony. And we have seen cases of that obviously.

So it's interesting, also, how -- I actually am going to name names. A lot of these companies who are really into trying to like do the right thing, when it comes down to these laws, there's also the anti-circumvention law, which is it's a felony to break copyright protection on anything, except for a very small number of exceptions.

So if you have an algorithm running on your computer, but it's protected, you can't audit it. And these are really stifling things for researchers. But you can imagine Hollywood doesn't want to loosen up copyright protections. And software companies and online companies don't want to loosen up your ability to research how their systems work. And the fact that a lot of people who talk about Internet freedom and all of this stuff don't talk about the fact that these laws are impeding research. I think it's sort of a shame.

JULIA ANGWIN:
Yeah. They are not impeding my research. But --

JOI ITO:
Your research is all legal. But . . .
(Chuckles).

JOI ITO:
But I do think it's something that we need to push against because unless you push against it, it won't change.
Good on that happy note, I would like to --
(Chuckles).

JOI ITO:
Thank you so much, Julia. This was really amazing. Thank you.

JULIA ANGWIN:
Thank you.
(Applause)