

DATA FEMINISM

You have 2 free member-only stories left this month. [Sign up for Medium and get an extra one](#)

5 Questions on Data and Accountability with Julia Angwin



Catherine D'Ignazio

Follow

Feb 15, 2020 · 8 min read ★

By Catherine D'Ignazio with editing by Isabel Carter





Image courtesy of Julia Angwin

Julia Angwin is an award-winning investigative journalist and the best-selling author of *Dragnet Nation*. She has worked at the *Wall Street Journal*, where she oversaw the groundbreaking series “What They Know” about erosion of privacy in the age of Big Data. Angwin is a co-founder and editor-in-chief of The Markup, a nonprofit newsroom that investigates the impact of technology on society. Before that, she worked as a senior investigative reporter at ProPublica where she was a finalist for a Pulitzer for her series on algorithmic bias, including a story on “Machine Bias,” a story about the ways that risk assessment tools perpetuate discriminatory sentencing in the American criminal justice system.

Lauren Klein and I reference “Machine Bias” in the “Collect, Analyze, Imagine, Teach” chapter of *Data Feminism*, which talks about challenging power in data science. “Machine Bias” serves as an undeniable, real world example of the harmful outcomes that can result when demographic data is used without accountability to the people it describes. As we say in the book, “Machine Bias” — and the many examples like it — proves that “the data are never ‘raw.’ Data are always the product of unequal social relations — relations affected by centuries of history.”

I caught up with Angwin on March 1, 2019 to talk about her thoughts on the story. What follows is an excerpt from that conversation. It has been edited for clarity.

What has been the most impactful outcome of the “Machine Bias” story?

That’s a great question. You know, it’s a weird art. Like most investigative journalism, you kind of aim for getting the bad cop fired and having some impact right away. To take on a structural issue like risk assessment is different, because even though we only wrote about one particular tool, results were generalizable to the field. We were taking on the whole question of risk assessments. So change came kind of slowly.

The things I feel best about are: one, the computer science community took it so seriously. I feel really proud that we could contribute to a whole rethinking of fairness in algorithms, and I think that the impact of that will be so widespread that it won't be measurable really. It will be a whole new approach to fairness, which I think is incredible, and it's something I'm really proud of.

The other thing is that although no one in the risk assessment field really took the advice onboard, the debate around risk assessments has changed. Previously, bail reform advocates were pretty much in favor of risk assessments because it seemed like a really good alternative to bail, and after the story came out a lot of bail reform advocates were kind of pissed off at me, and said I wasn't really helping, and yet at the same time what happened was, slowly, the conversation changed.

I felt really emboldened when some 200 civil rights groups wrote a letter saying that based on their new understanding about bias and risk assessments, they no longer wholeheartedly endorse them as part of the bail reform movement. They said the risk assessments would have to meet all these fairness conditions before they would endorse their use. That is great because obviously bail reform is a really important thing, but replacing it with systematized bias just didn't seem like the right outcome. So it feels like that debate is moving.

Generally with investigative journalism, you kind of want your impact to come within six months, so you can get a prize. One thing about the COMPAS story is that it didn't win many prizes. It's probably the most impactful story of my career, and it's made me think a lot about how the incentives are not setup well for investigative journalism. There's a lot better incentives to get one bad cop fired than to take on a whole structural issue. I can't solve that incentive system, but I do think that's one of the things I'm bringing to this new nonprofit newsroom: the idea of making sure we focus on these harder issues, even if we know they aren't going to pay off in the awards cycle.

It's not really discussed in the story, but there's a line in your methodology about gender: "Surprisingly, given their lower levels of criminality overall, female defendants were 19.4 percent more likely to get a higher score than men, controlling for the same factors." Can you say more about gender and risk assessment algorithms?

Yeah, there was a significant bias in the risk assessments. Not only were those likelihood ratios 20 percent higher, but the recidivism rate for a woman who scored a “High” is not as high as a man’s at the high end. Because women just generally recidivate less, the whole scale is such that somebody with a high risk score that’s a woman is generally about the level of a medium risk score for a man. And that’s true of almost all of the risk score tools.

So there has been a lot of work on that through Patricia Van Voorhis at the University of Cincinnati who has studied it for like 10–15 years, and I came across her work while I was researching COMPAS. Because it was known but only in academic circles, I decided it was a less newsy finding, so we didn’t focus on that in the story.

I think if I had to do it over again though, I’d probably go back and do a second story on the fact that there’s this known problem with all women’s risk scores. There could have been more to do on that front, but that’s also a journalism incentive thing. So I regret not putting enough focus on the women’s issue because it’s actually really shocking that judges are looking at these and thinking that high risk means the same thing for a man and a woman when it doesn’t.

How did your own identity as a woman in data journalism play into not want to talk about gender in relation to findings, if at all?

I think in general in my career as a woman focusing on tech, I have an expertise in programming and a math degree. I have emphasized all of those skills in order to create credibility for myself. I have leaned heavily into that, and I’ve also leaned away from stories about traditional women’s issues. Partly because there aren’t that many to write about in the tech industry, but also because it’s a pigeonhole right? It happens to Black writers all the time — the race beat — and then they have to only do that forever. Some people are okay with that, and that’s actually what they want to do, but for me, I didn’t want to be pigeonholed.

One of the things I woke up to during the #MeToo movement was how many decisions I had made like that over the years. I thought back to my years as a corporate reporter at the *Wall Street Journal*, and how many executives I wrote about who I knew were having affairs with their secretary or whatever. It didn’t occur to me

that that was news, being abusive to women, harassment. I didn't think about it until the #MeToo movement when I was like, "oh shoot, that's news."

But it's hard to be the first one out the gate with that. I'm really, really proud that the woman that was the first one out the gate with that was my [mentee], Emily Steel, at the *New York Times*. She came to work with me at the *Wall Street Journal* right out of college as an intern, and I trained her on my investigative team there. She wrote the first story in the *New York Times* about the Bill O'Reilly sexual harassment cases which then led to the Harvey Weinstein stuff.

While she was doing that work and while those stories were coming out, I started thinking about my own career, and how there were a lot of stories I had stepped aside from. I would put the COMPAS risk score in that category really, because I could have easily done a story on bias against women. It was well known, and well documented, and nobody had ever done a really big story on it. I didn't, mostly because it didn't occur to me that that was news.

What role does journalism have to play in the emerging conversations about ethics, data and artificial intelligence?

I feel like journalism has the same role here as it has in everything which is accountability, holding power to account. In the era of data and AI the challenge is that accountability is hard to prove and hard to trace. Anyone who's ever been denied a loan knows that it's sometimes difficult to know what the reason was or why you didn't get the job or the house you wanted to rent. Sometimes you know it's your credit score, right? But sometimes an employer might be looking at your Facebook page. We just don't know what data is being used against us.

The challenge for journalism is to try to make those linkages as concrete as possible, so we can show the world what the harms are. Journalism is a way to illuminate harm, so it can be solved, so it can be mitigated. You need a lot of technical expertise in order to do the analysis that would hold those things accountable. That's why I'm investing in this newsroom that's like half programmers.

Tell me about The Markup — what's the most exciting thing you all are working on?

The thing that is so awesome right now is hiring. There are so many people in so many newsrooms who actually have technical skills but haven't been given the freedom to generate stories. One of the problems I'm trying to solve here is the fact that people with technical skills are usually put at a data desk in a newsroom, and they are treated like a service. They have no agency over stories. Reporters come to them with a hypothesis which may or may not be true, and they demand some supporting evidence. What I found was that giving those people the opportunity to drive the story process, to collect the data originally, to think of the hypothesis is something a lot of people are really there for, so it's really exciting because I'm staffing all these people who — when they're unleashed — will have so many great stories to tell.

Plus, we're taking a very explicit scientific method approach to journalism. The idea is that we really do believe you should have a hypothesis and test your hypothesis and that's what journalism should do. That's an explicit repudiation of objectivity which was the previous lodestar for journalism and which, as I think we've all come to learn, has led to a lot of false equivalence, and not really informing the public, and sometimes really actively misinforming the public as if there were really two sides to every issue, right? Usually when you really think about what the two sides are, it's like the preponderance of evidence. 85 percent of it supports one side and 15 percent supports the other side.

I think it's incumbent upon us to be more honest about that. That's our goal in our reporting. What we should do is write about our best efforts to obtain the truth and what the limitations are of those findings. That's our goal, and still that's a challenge, because doing that in a way that's readable, and exciting, and feels like a narrative is going to be a challenge, but we're all really excited about it.

Learn more about [Julia Angwin](#) and check out [The Markup](#), Angwin's technology accountability news organization.

[Data Journalism](#)[Data Science](#)[Algorithms](#)[Feminism](#)[Criminal Justice Reform](#)[About](#) [Help](#) [Legal](#)

Get the Medium app

