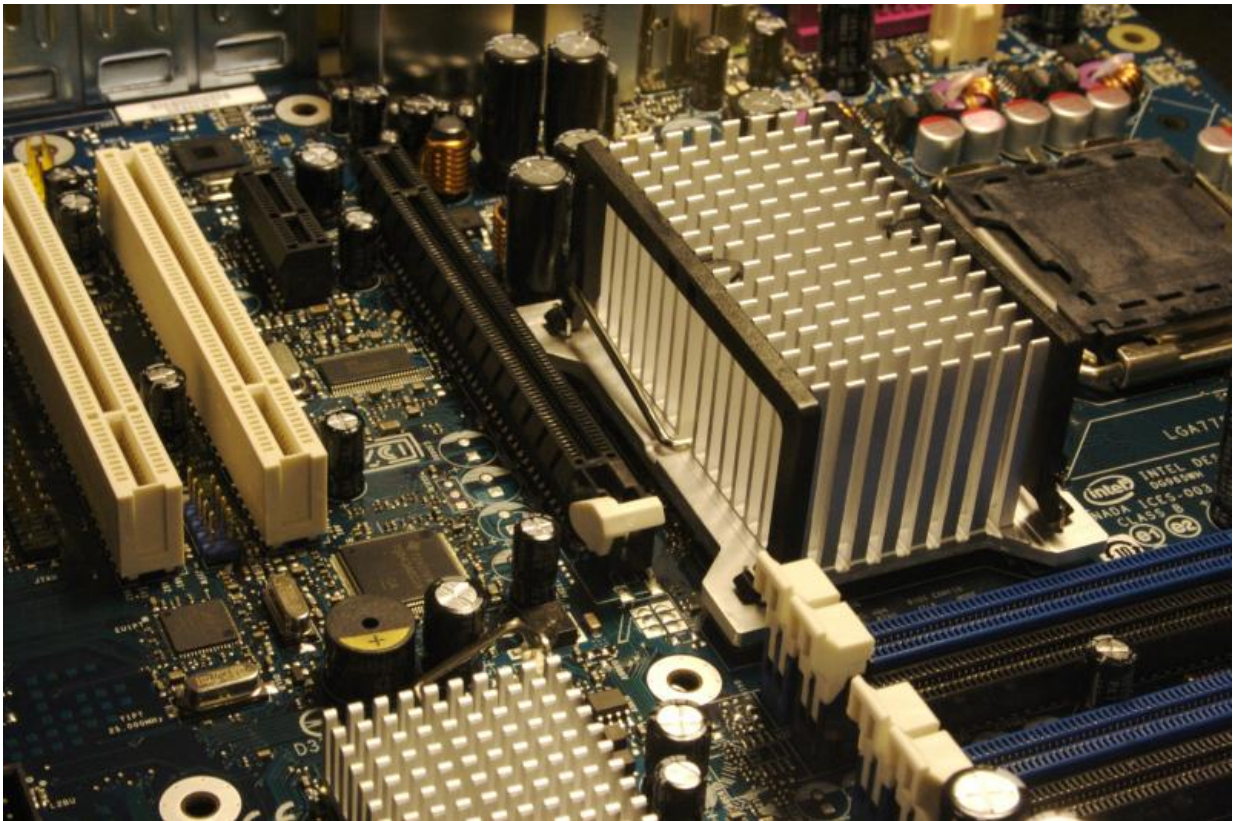# Big data algorithms can discriminate, and it's not clear what to do about it

August 13 2015, by Jeremy Kun

Is bias baked in? Credit: Justin Ruckman, CC BY

"This program had absolutely nothing to do with race…but multi-variable equations."

That's what Brett Goldstein, a former policeman for the Chicago Police Department (CPD) and current Urban Science Fellow at the University of Chicago's School for Public Policy, said about a predictive policing algorithm he deployed at the CPD in 2010. His algorithm tells police where to look for criminals based on where people have been arrested previously. It's a "heat map" of Chicago, and the CPD claims it helps them allocate resources more effectively.

Chicago police also recently collaborated with Miles Wernick, a professor of electrical engineering at Illinois Institute of Technology, to algorithmically generate a "heat list" of 400 individuals it claims have the highest chance of committing a violent crime. In response to criticism, Wernick said the algorithm does not use "any racial, neighborhood, or other such information" and that the approach is "unbiased" and "quantitative." By deferring decisions to poorly understood algorithms, industry professionals effectively shed accountability for any negative effects of their code.

But do these algorithms discriminate, treating low-income and black neighborhoods and their inhabitants unfairly? It's the kind of question many researchers are starting to ask as more and more industries use algorithms to make decisions. It's true that an algorithm itself is quantitative – it boils down to a sequence of arithmetic steps for solving a problem. The danger is that these algorithms, which are trained on data produced by people, may reflect the biases in that data, perpetuating structural racism and negative biases about minority groups.

There are a lot of challenges to figuring out whether an algorithm embodies bias. First and foremost, many practitioners and "computer experts" still don't publicly admit that algorithms can easily discriminate. More and more evidence supports that not only is this possible, but it's happening already. The law is unclear on the legality of biased algorithms, and even algorithms researchers don't precisely understand

what it means for an algorithm to discriminate.

Being quantitative doesn't protect against bias

Both Goldstein and Wernick claim their algorithms are fair by appealing to two things. First, the algorithms aren't explicitly fed protected characteristics such as race or neighborhood as an attribute. Second, they say the algorithms aren't biased because they're "quantitative." Their argument is an appeal to abstraction. Math isn't human, and so the use of math can't be immoral.

Sadly, Goldstein and Wernick are repeating a common misconception about data mining, and mathematics in general, when it's applied to social problems. The entire purpose of data mining is to discover hidden correlations. So if race is disproportionately (but not explicitly) represented in the data fed to a [data-mining](link) algorithm, the algorithm can infer race and use race indirectly to make an ultimate decision.

Here's a simple example of the way algorithms can result in a biased outcome based on what it learns from the people who use it. Look at how how Google search suggests finishing a query that starts with the phrase "transgenders are":

Autocomplete features are generally a tally. Count up all the searches you've seen and display the most common completions of a given partial query. While most algorithms might be neutral on the face, they're designed to find trends in the data they're fed. Carelessly trusting an algorithm allows dominant trends to cause harmful discrimination or at least have distasteful results.
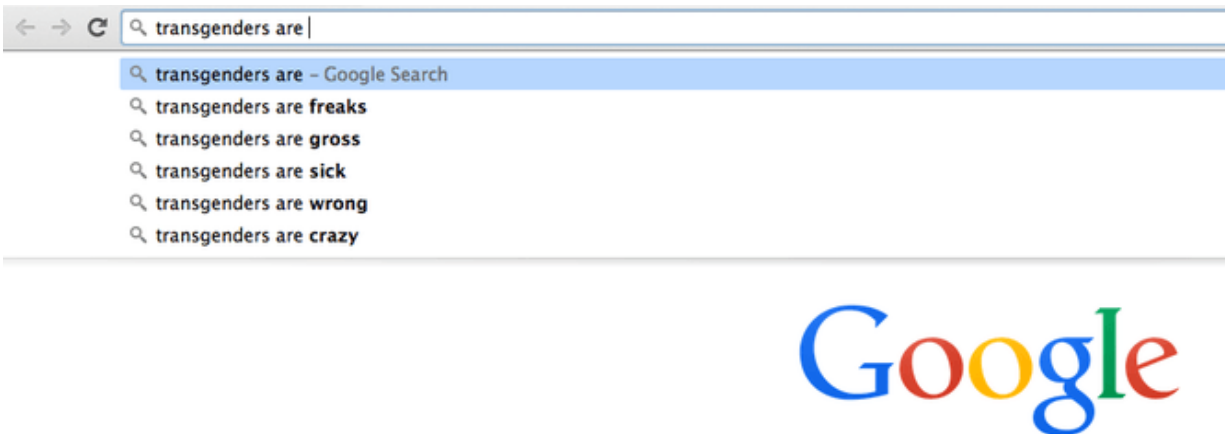
Taken from Google.com on 2015-08-10.

Beyond biased data, such as Google autocompletes, there are other pitfalls, too. Moritz Hardt, a researcher at Google, describes what he calls the [sample size disparity](). The idea is as follows. If you want to predict, say, whether an individual will click on an ad, most algorithms optimize to reduce error based on the previous activity of users.

But if a small fraction of users consists of a racial minority that tends to behave in a different way from the majority, the algorithm may decide it's better to be *wrong* for all the minority users and lump them in the "error" category in order to be more accurate on the majority. So an algorithm with 85% accuracy on US participants could err on the entire black sub-population and still seem very good.

Hardt continues to say it's hard to determine why data points are erroneously classified. Algorithms rarely come equipped with an explanation for why they behave the way they do, and the easy (and dangerous) course of action is not to ask questions.

Extent of the problem

While researchers clearly understand the theoretical dangers of algorithmic discrimination, it's difficult to cleanly measure the scope of the issue in practice. No company or public institution is willing to publicize its data and algorithms for fear of being labeled racist or sexist, or maybe worse, having a great algorithm stolen by a competitor.

Even when the Chicago Police Department was [hit with a Freedom of Information Act request](#), they did not release their algorithms or heat list, claiming a credible threat to police officers and the people on the list. This makes it difficult for researchers to identify problems and potentially provide solutions.

Legal hurdles

Existing discrimination law in the United States isn't helping. At best, it's unclear on how it applies to algorithms; at worst, it's a mess. Solon Barocas, a postdoc at Princeton, and Andrew Selbst, a law clerk for the Third Circuit US Court of Appeals, [argued together](#) that US hiring law fails to address claims about discriminatory algorithms in hiring.

The crux of the argument is called the "business necessity" defense, in which the employer argues that a practice that has a discriminatory effect is justified by being directly related to job performance. According to Barocas and Selbst, if a company algorithmically decides whom to hire, and that algorithm is blatantly racist but even mildly successful at predicting job performance, this would count as business necessity – and not as illegal discrimination. In other words, the law seems to *support* using biased algorithms.

What *is* fairness?

Maybe an even deeper problem is that nobody has agreed on what it means for an algorithm to be fair in the first place. Algorithms are mathematical objects, and mathematics is far more precise than law. We can't hope to design fair algorithms without the ability to precisely demonstrate fairness mathematically. A good mathematical definition of fairness will model biased decision-making in any setting and for any subgroup, not just hiring bias or gender bias.

And fairness seems to have two conflicting aspects when applied to a population versus an individual. For example, say there's a pool of applicants to fill 10 jobs, and an algorithm decides to hire candidates completely at random. From a population-wide perspective, this is as fair as possible: all races, genders and orientations are equally likely to be selected.

But from an individual level, it's as unfair as possible, because an extremely talented individual is unlikely to be chosen despite their qualifications. On the other hand, hiring based only on qualifications reinforces hiring gaps. Nobody knows if these two concepts are inherently at odds, or whether there is a way to define fairness that reasonably captures both. Cynthia Dwork, a Distinguished Scientist at Microsoft Research, and her colleagues have been studying the relationship between the two, but even Dwork admits they have just scratched the surface.

Get companies and researchers on the same page

There are immense gaps on all sides of the algorithmic fairness issue. When a panel of experts at this year's Workshop on Fairness, Accountability, and Transparency in Machine Learning was asked what the low-hanging fruit was, they struggled to find an answer. My opinion is that if we want the greatest progress for the least amount of work, then businesses should start sharing their data with researchers. Even with

[proposed "fair" algorithms](link) starting to appear in the literature, without well-understood benchmarks we can't hope to evaluate them fairly.

This story is published courtesy of [The Conversation](link) *(under Creative Commons-Attribution/No derivatives).*

Source: The Conversation

Citation: Big data algorithms can discriminate, and it's not clear what to do about it (2015, August 13) retrieved 2 May 2024 from https://phys.org/news/2015-08-big-algorithms-discriminate.html