# Comprehending Causation and Correlation

**Author :** Charles A. Sullivan

**Date :** August 4, 2017

James Grimmelmann & Daniel Westreich, *Incomprehensible Discrimination*, 7 **Calif. L. Rev. Online** 164 (2017), available at SSRN.

I'm a fan of off-beat approaches to legal scholarship, having attempted a couple of efforts myself. And I try to keep up with developments in the real world that threaten to impact our discipline, like the concern about "Big Data" that has begun to appear in the law journals. So it's no surprise that I was very taken by a particularly creative piece by Professors James Grimmelmann and Daniel Westreich, which combines an amusing conceit with dead-on analysis of an emerging and important question.

*Incomprehensible Discrimination* explores one aspect of a much longer article that appeared recently in the California Law Review by Solon Barocas and Andrew Selbst on data-driven algorithmic methods of making employment decisions.[1] The aspect Grimmelmann & Westreich explore is the ability of such analyses to find correlations between, say, job performance and any of a variety of data points with no apparent causal connection to better performance. Barocas and Selbst conclude that traditional disparate impact analysis is not likely to invalidate these kind of selection process. Given sufficient data and robust tests of significance, it's hard to conclude that reliance on such factors is irrational, even in the absence of any articulable explanation for what one has to do with the other. For Grimmelmann and Westreich, that's exactly the problem.

For example, it is at least conceivable that better job performance can be predicted from characteristics such as the applicant's favorite kind of music, whether she owns a car, or her zip code. But what confidence can we have in the resulting explanation?

Grimmelmann & Westreich's method of exploring this question is a "(fictional) opinion of the (fictional) Zootopia Supreme Court of the (fictional) State of Zootopia," which is designed to explore the disparate impact implications of this use of Big Data. They posit a hiring process that could easily result in racial (or, in the world of Zootopia, species) distinctions without any intent to discriminate. Thus, the Zootopia Police Department

> uses a mathematical model to predict which applicants will be successful police officers. Four facts about this model are undisputed. First, its scores are significantly correlated with a reasonable measure of job performance. Second, the model does not explicitly consider applicants' species. Third, it nonetheless systematically favors carnivorous applicants over herbivores. Fourth, no one has explained how and why the model works at predicting job performance or how and why it disadvantages herbivores.

The *second* undisputed fact is designed to rule out intentional discrimination. (Although it remains possible for a facially neutral model to be adopted for discriminatory purposes, that is very difficult to prove.) And, while the *third* raises a prima facie case of disparate impact, the *first* would seem to necessarily mean that the practice satisfied the business necessity/job relation standard under traditional analysis.

As for the *fourth*, that consideration does not currently have a place at the table under either disparate treatment or disparate impact analysis and, to be frank, wasn't especially relevant before Big Data raised exactly that question. But it is this possibility that Grimmelmann & Westreich argue should be determinative in any court challenge to such a practice. According to them, such algorithmic analysis may "both predict[] job performance and discriminate[] against herbivores" without anyone being able to provide a causal explanation for the result. (P. 170.) The Zootopia Supreme Court elaborates:

> The [plaintiff] League is correct that the factors that the model identified correlate with species, and the [employer] is correct that these factors also correlate with job performance. The problem is that there is no explanation in the record as to which of these two correlations, if either, is causal. It may be that the factors directly measure applicant characteristics that determine success in the challenging and dangerous field of police work, and that those characteristics happen to be unequally distributed in our diverse society. It may also be that these factors are instead measuring applicants' species and that they measure likely job performance *only because they are identifying species* in an applicant pool where the relevant characteristics are unequally distributed. (P. 170.)

To deal with this risk, the authors argue that the defendant's business necessity burden "requires it to show not just that its model's scores are not just *correlated* with job performance but *explain* it." (P. 170.) They suggest the necessity for "an explanation in terms of the chains of causation by which one state of affairs leads to another," and elaborate that a good explanation "is one that identifies the hidden and nondiscriminatory variables connecting the observed factors with the predicted target variable."

The reason, of course, is that, even a model that is scrubbed of any explicit race factors will still yield racially-slanted results if the factors actually used are correlated with race. Nor is this so strange. Although the authors don't mention it, the Supreme Court has repeatedly had to deal with the question of whether legislative redistricting decisions are race-motivated (unconstitutional absent a compelling state interest) or politically motivated (constitutional) given that African Americans vote very heavily Democratic so there is a very strong correlation between the two.

According to the authors, that kind of inquiry is necessitated whenever impenetrable algorithms yield disparate results for protected classes. They end with a cri de coeur:

> Our holding today is simple. *Incomprehensible discrimination will not stand*. Applicants who are judged and found wanting deserve a better explanation than, "The computer said so." Sometimes computers say so for the wrong reasons–and it is employers' duty to ensure that they do not. (P. 177.)

As is often the case, this Jot can't do justice to the Grimmelmann & Westreich piece, which is well worth a read, as is the Barocas & Selbst article to which it responds.

1. Solon Barocas & Andrew Selbst, *Big Data's Disparate Impact*, 104 **Calif. L. Rev.** 671 (2016). [?]

Cite as: Charles A. Sullivan, *Comprehending Causation and Correlation*, JOTWELL (August 4, 2017) (reviewing James Grimmelmann & Daniel Westreich, *Incomprehensible Discrimination*, 7 **Calif. L. Rev. Online** 164 (2017), available at SSRN), http://worklaw.jotwell.com/comprehending-causation-and-correlation/.