Information Storage and Retrieval

CSCE 670
Texas A&M University
Department of Computer Science & Engineering
Instructor: Prof. James Caverlee

The Dark Side
20 April 2017

A dangerous reasoning

To discriminate is to treat someone differently

(Unfair) discrimination is based on group membership, not individual merit

People's decisions include objective and subjective elements

Hence, they can be discriminate

Algorithmic inputs include only objective elements

Hence, they cannot discriminate?

Automated Inference on Criminality using Face Images

Xiaolin Wu
McMaster University
Shanghai Jiao Tong University
xwu510@gmail.com

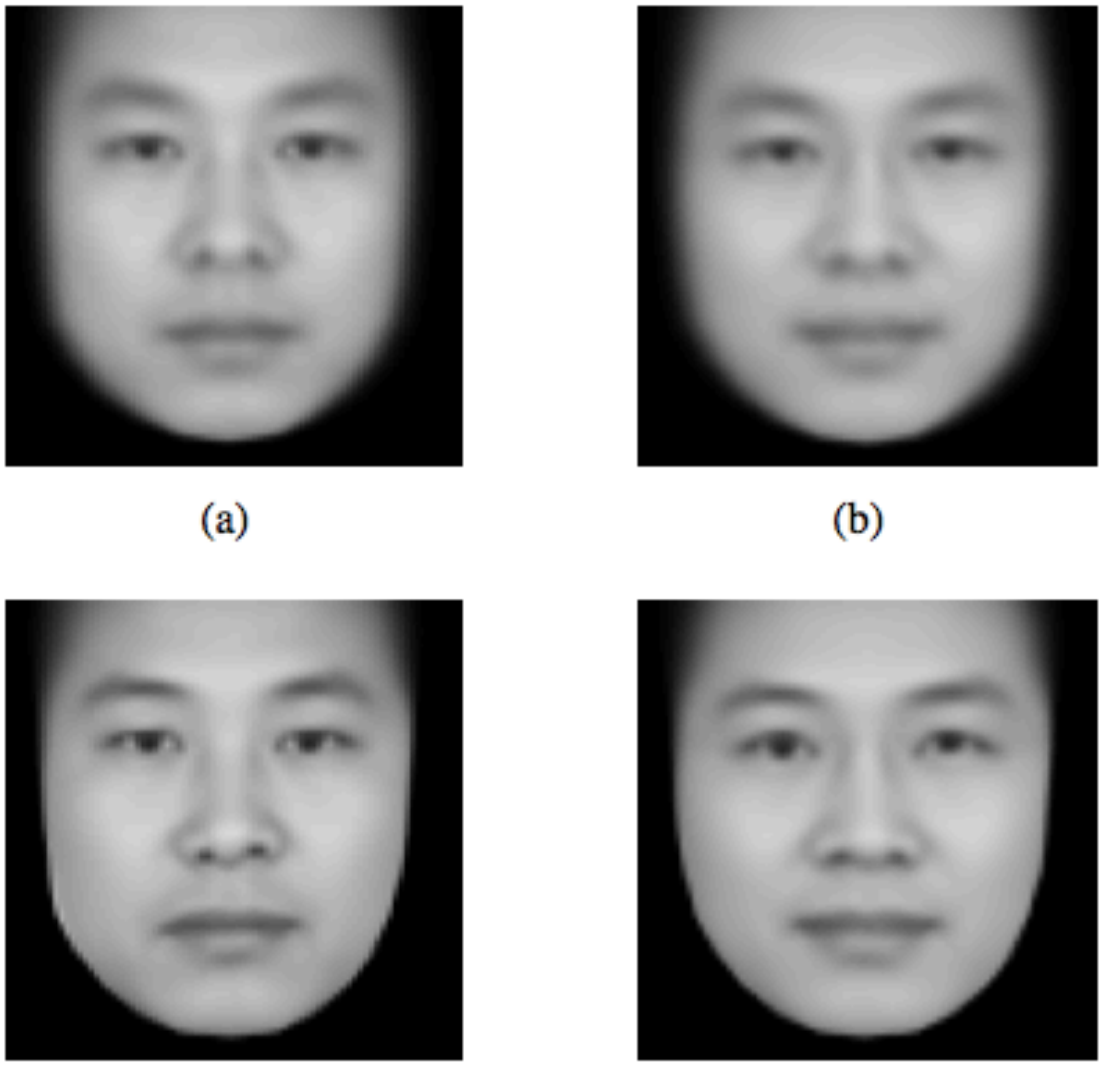
Xi Zhang
Shanghai Jiao Tong University
zhangxi_19930818@sjtu.edu.cn

We study, for the first time, automated inference on criminality based solely on still face images, which is free of any biases of subjective judgments of human observers. Via supervised machine learning, we build four classifiers (logistic regression, KNN, SVM, CNN) using facial images of 1856 real persons controlled for race, gender, age and facial expressions, nearly half of whom were convicted criminals, for discriminating between criminals and non-criminals. All four classifiers perform consistently well and empirically establish the validity of automated face-induced inference on criminality, despite the historical controversy surrounding this line of enquiry. Also, some discriminating structural features for predicting criminality have been found by machine learning. Above all, the most important discovery of this research is that criminal and non-criminal face images populate two quite distinctive manifolds. The variation among criminal faces is significantly greater than that of the non-criminal faces. The two manifolds consisting of criminal and non-criminal faces appear to be concentric, with the non-criminal manifold lying in the kernel with a smaller span, exhibiting a law of "normality" for faces of non-criminals. In other words, the faces of general law-biding public have a greater degree of resemblance compared with the faces of criminals, or criminals have a higher degree of dissimilarity in facial appearance than non-criminals.

(a) Three samples in criminal ID photo set S_c.

(b) Three samples in non-criminal ID photo set S_n

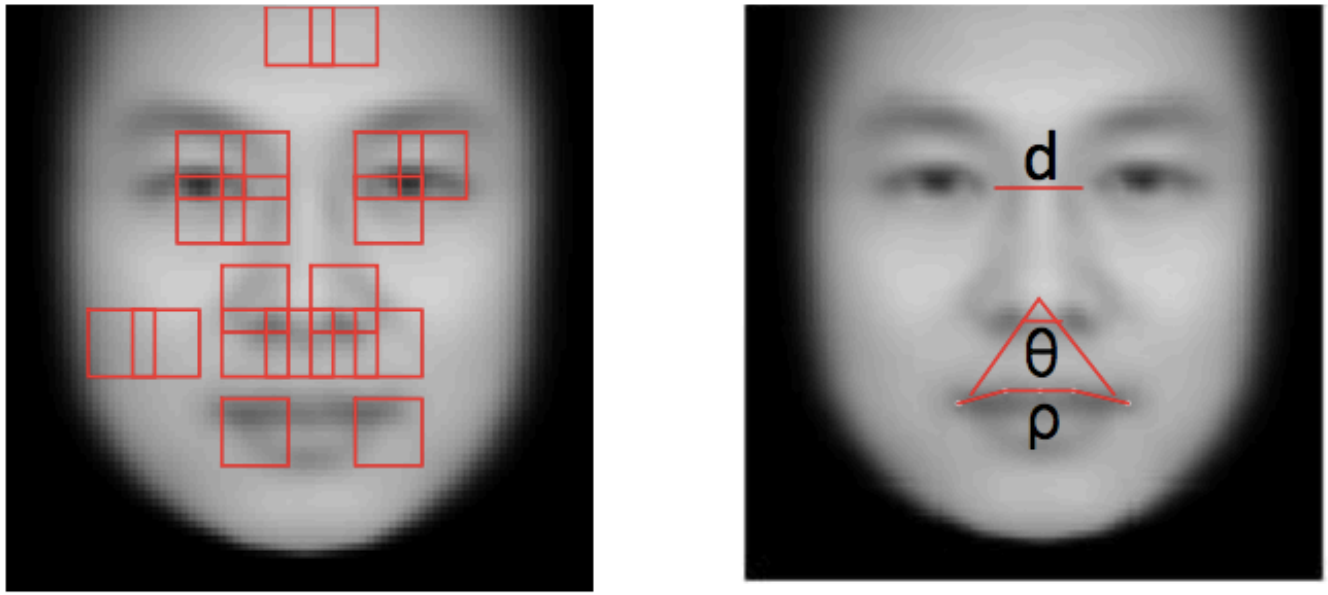
Figure 1. Sample ID photos in our data set.



(a) (b)

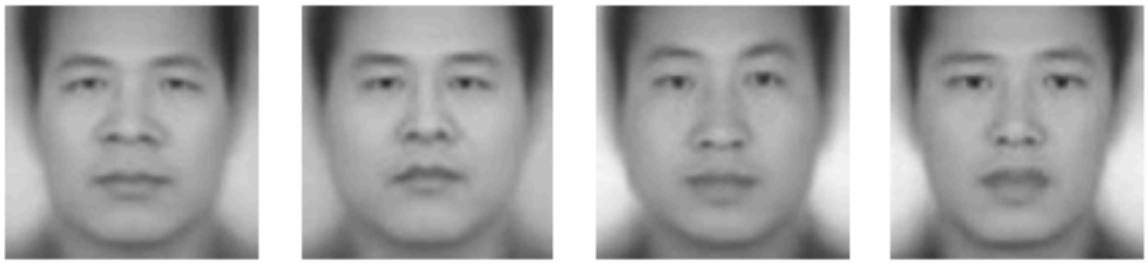
(c) (d)

Figure 10. (a) and (b) are "average" faces for criminals and non-criminals generated by averaging of eigenface representations ; (c) and (d) are "average" faces for criminals and non-criminals generated by averaging of landmark points and image warping.



(a) (b)

Figure 8. (a) FGM results; (b) Three discriminative features ρ, d and θ.



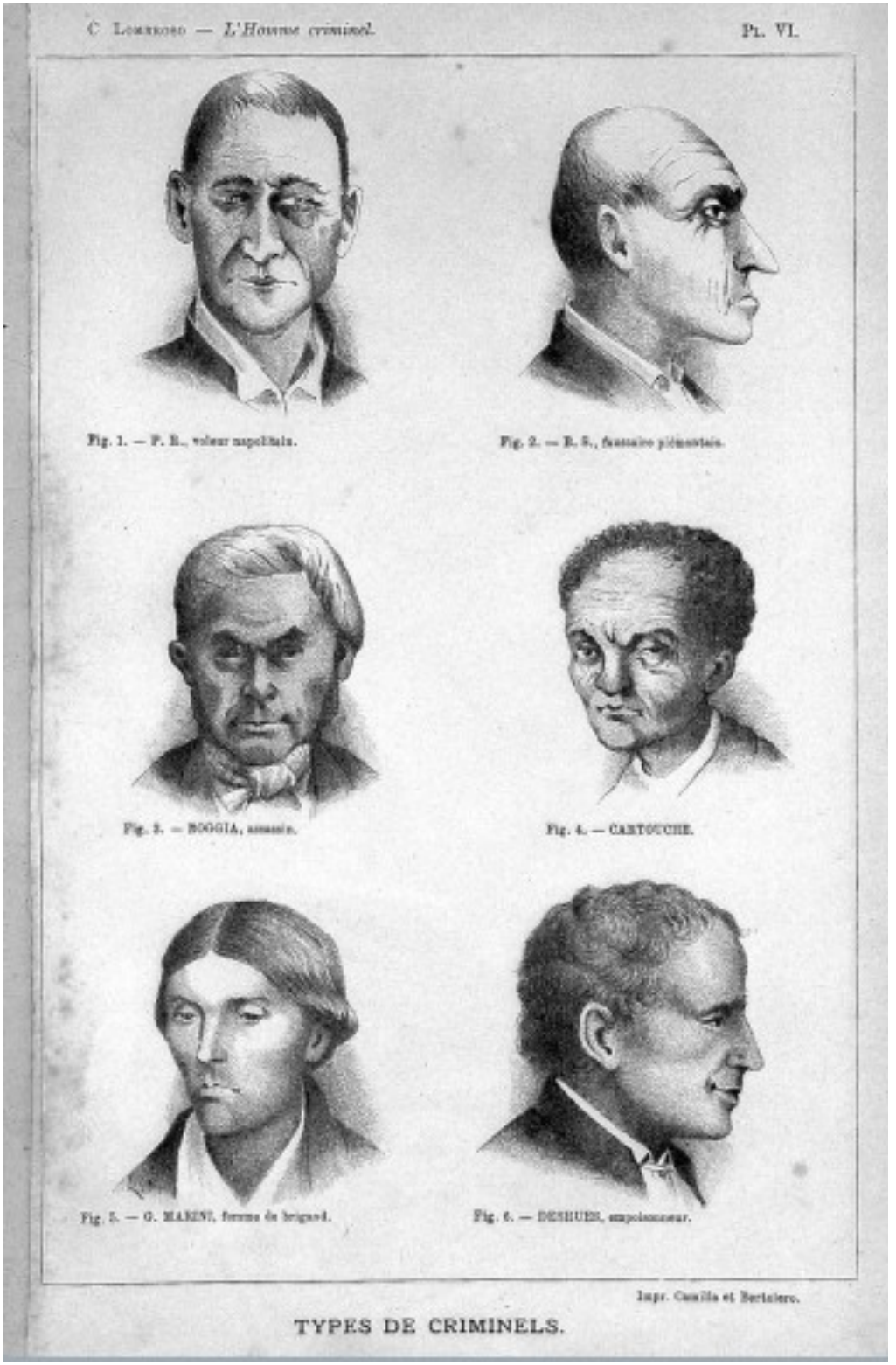
(a) -0.98 (b) -0.68 (c) -0.28 (d) -0.38

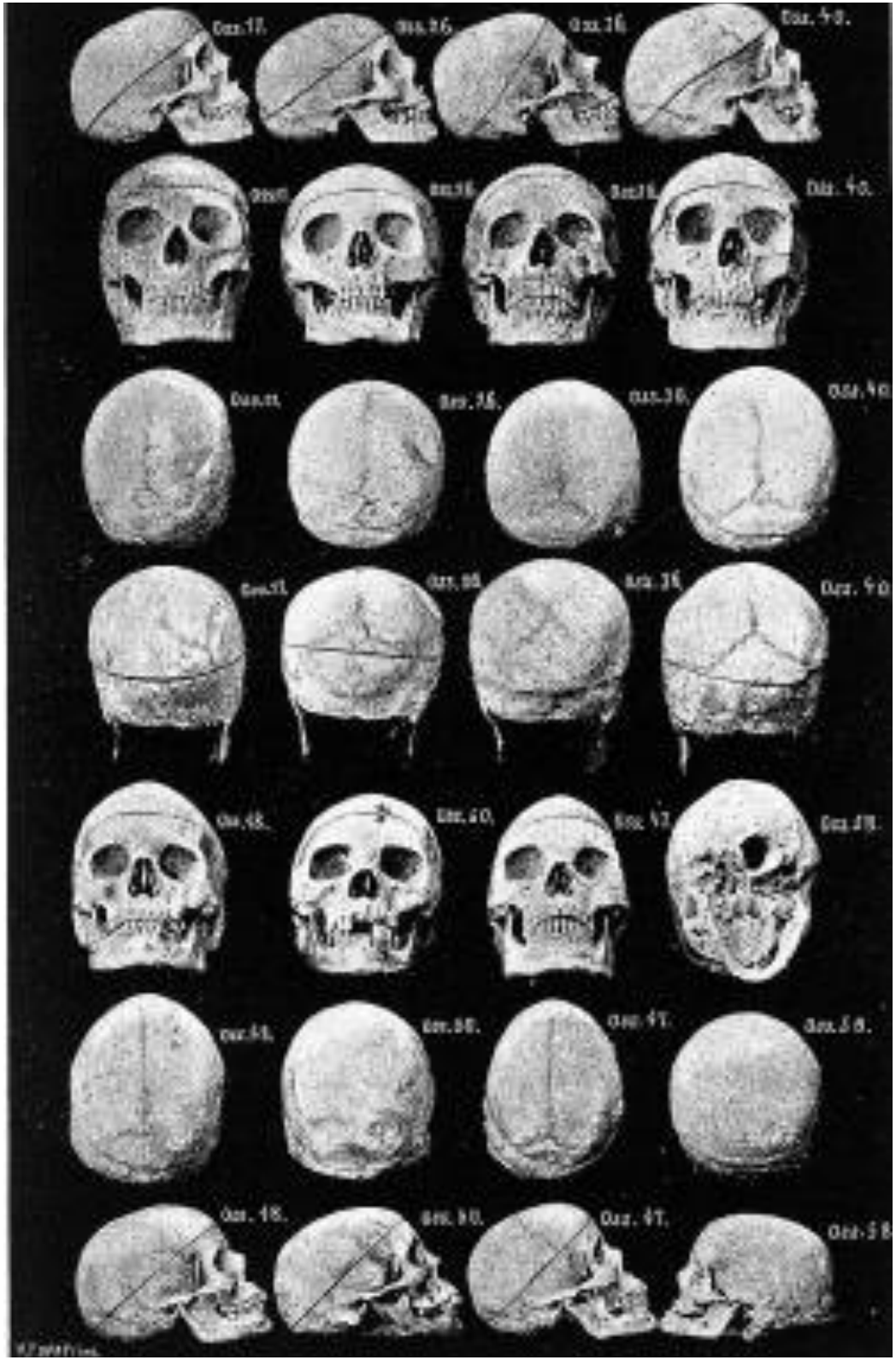
(e) 0.76 (f) 0.98 (g) 0.66

Figure 13. (a), (b), (c) and (d) are the four subtypes of criminal faces corresponding to four cluster centroids on the manifold of S_c; (e), (f) and (g) are the three subtypes of non-criminal faces corresponding to three cluster centroids on the manifold of S_n. The number associated with each face is the average score of human judges (-1 for criminals; 1 for non-criminals).

- From the paper: “Unlike a human examiner/judge, a computer vision algorithm or classifier has absolutely no subjective baggages, having no emotions, no biases whatsoever due to past experience, race, religion, political doctrine, gender, age, etc., no mental fatigue, no preconditioning of a bad sleep or meal. The automated inference on criminality eliminates the variable of meta-accuracy (the competence of the human judge/ examiner) all together.”
- Thoughts?

Cesare Lombroso

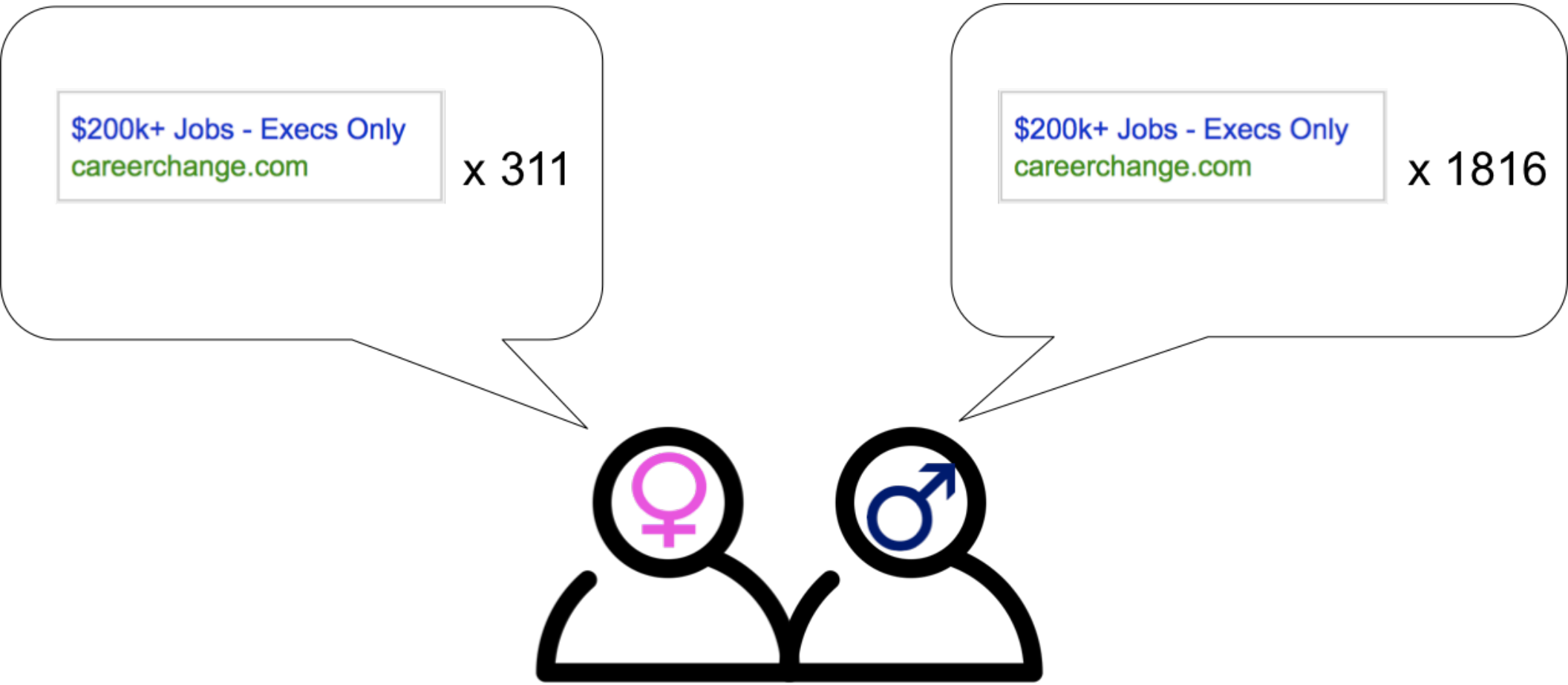




phrenology = the detailed study of the shape and size of the cranium as a supposed indication of character and mental abilities.

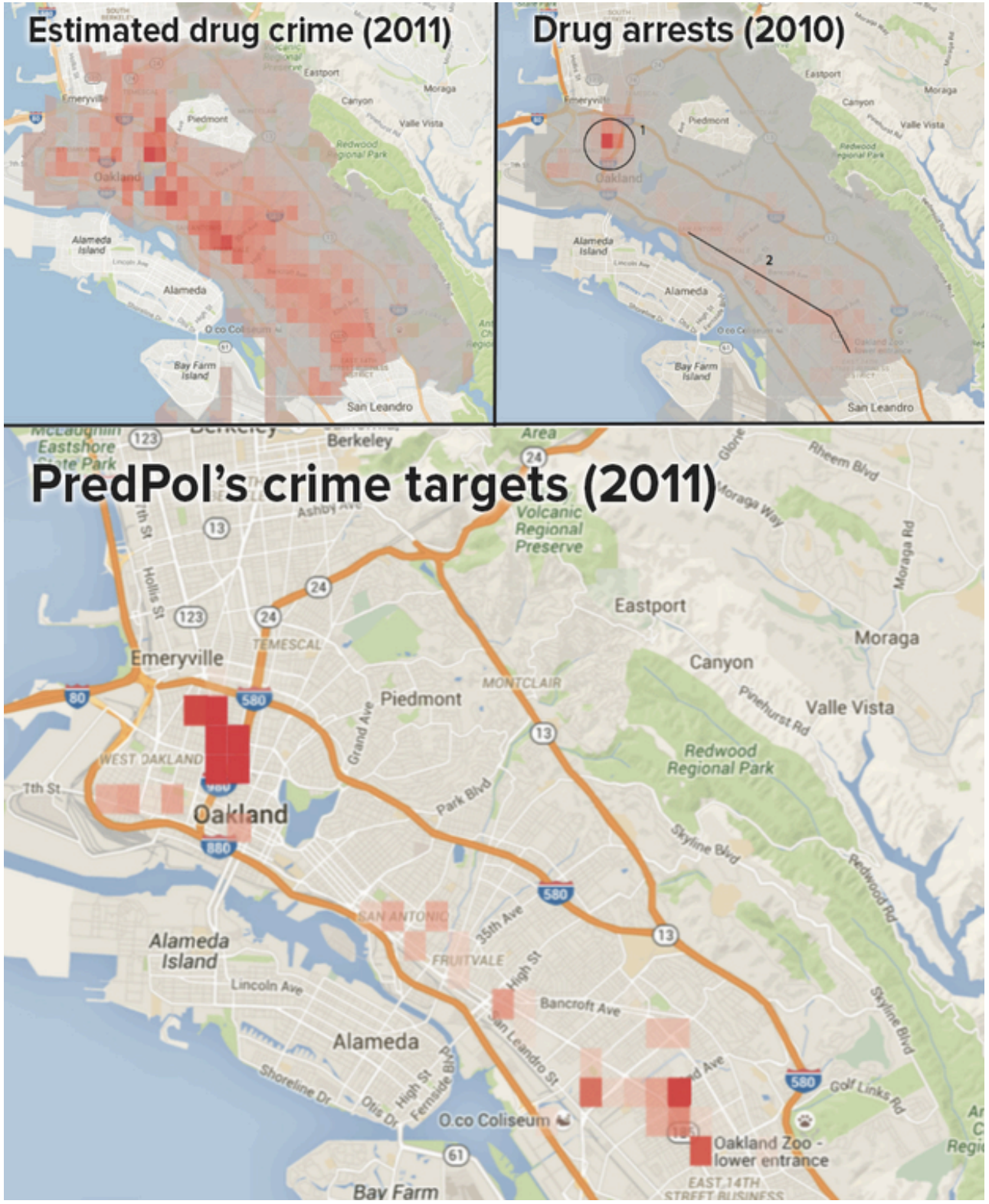
Discrimination and Opacity in Online Behavioral Advertising

- http://possibility.cylab.cmu.edu/adfisher/
- To study discrimination, we had AdFisher create 1000 fresh browser instances and assign them randomly to two groups. One group set their gender to male on Google's Ad Settings page, while the other set it to female. Then, all the browsers visited the top 100 websites for employment on Alexa. Thereafter, all the browsers collected the ads served by Google on the Times of India.



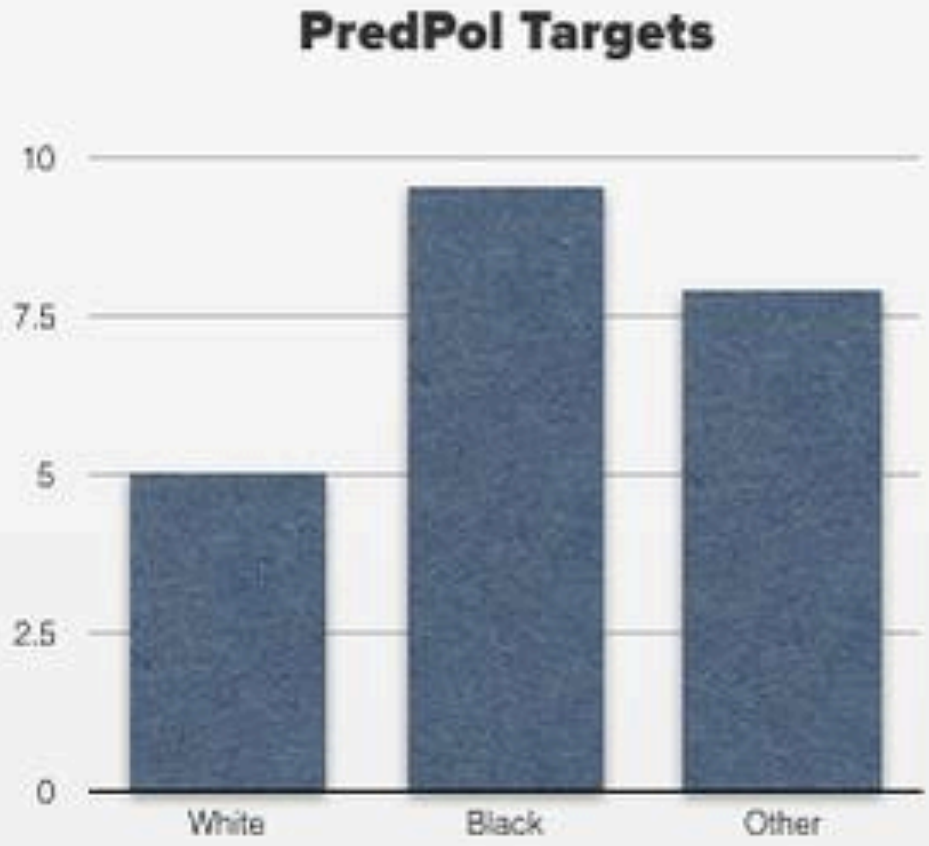
- The top two ads served to the male group was from a career coaching service called careerchange.com that promised high-paying executive level jobs. The top ad was served 1816 times to the male users, but only 311 times to the female users. Of the 500 simulated male users, 402 received the ad at least once, but only 60 female users received the same ad at least once.

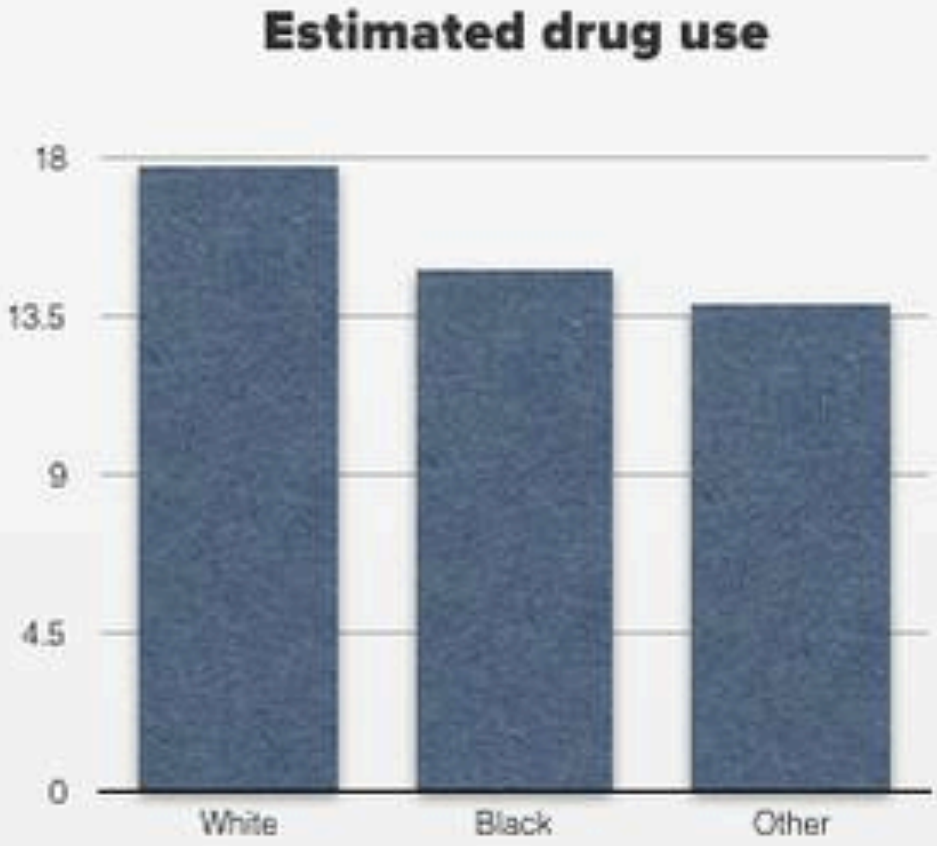

PREDPOL®
The Predictive Policing Company®
MORE THAN A HOTSPOT TOOL
PROVEN & FIELD TESTED
EASY TO DEPLOY & ACCESS
IN THE NEWS


Estimated drug crime (2011)
Drug arrests (2010)
1
2
PredPol's crime targets (2011)
Berkeley
Emeryville
Piedmont
Oakland
Alameda
Alameda Island
Bay Farm Island
San Leandro
Eastport
Canyon
Moraga
Valle Vista
Redwood Regional Park
Volcanic Regional Preserve
Lincoln Ave
Bancroft Ave
Golf Links Rd
Skyline Blvd
Oakland Zoo - lower entrance
O.co Coliseum

Predictive policing in Oakland vs. actual drug use

The chart on the left shows the demographic breakdown of people targeted for policing based on a simulation of PredPol in Oakland. The chart on the right shows actual estimated use of illicit drugs.





source: National Survey on Drug Use and Health , Human Rights Data Analysis Group

Mic

Self-perpetuating algorithmic biases

Credit scoring algorithm suggests Joe has high risk of defaulting

Hence, Joe needs to take a loan at a higher interest rate

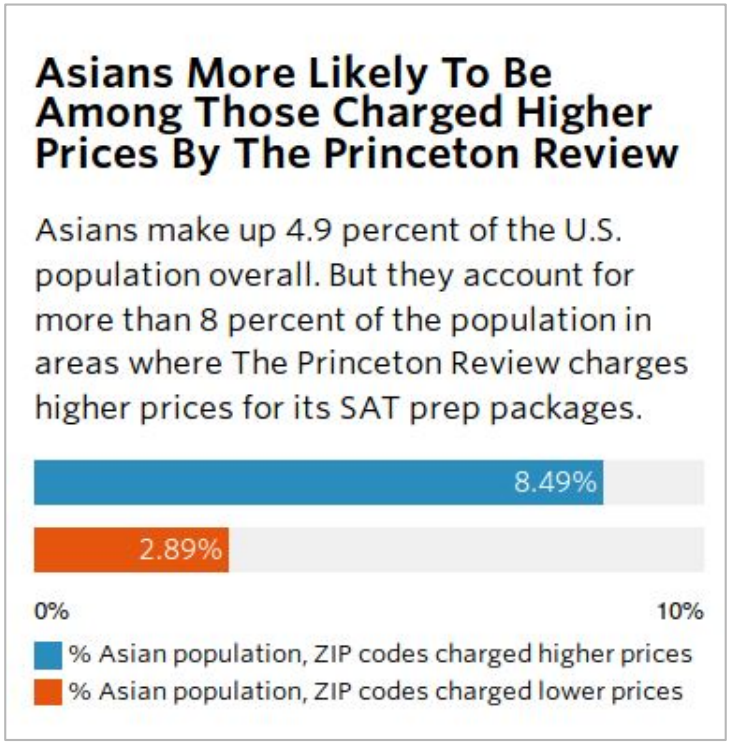
Hence, Joe has to make payments that are more onerous

Hence, Joe's risk of defaulting has increased

The same happens with stop-and-frisk of minorities further increasing incarceration rates

Geography and race: the "Tiger Mom Tax"

Pricing of SAT tutoring by The Princeton Review in the US doubles for Asians, due to geographical price discrimination



J. Angwin and J. Larson (2015). The tiger mom tax. ProPublica.

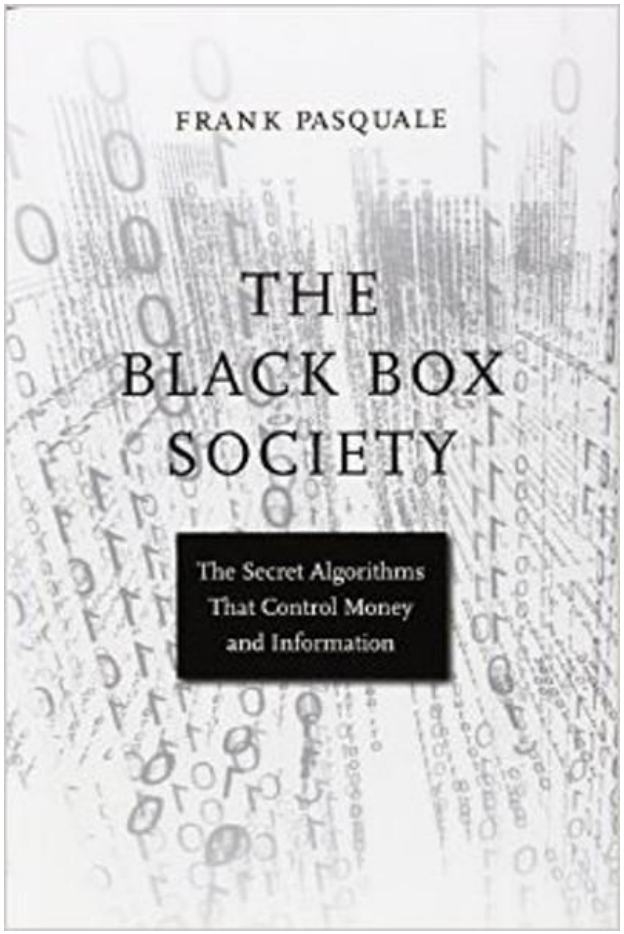
Judiciary use of COMPAS scores

COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is a 137-questions questionnaire and a predictive model for "risk of recidivism" and "risk of violent recidivism." The model is a proprietary secret of Northpointe, Inc.

Prediction accuracy of recidivism for blacks and whites is about the same (63% and 59%), but errs by being too lenient with whites and too harsh with blacks:

- Blacks that did not reoffend
  were classified as high risk twice as much as whites that did not reoffend
- Whites who did reoffend
  were classified as low risk twice as much as blacks who did reoffend

To make things worse ...



Algorithms are "black boxes" protected by

- Industrial secrecy
- Legal protections
- Intentional obfuscation

Discrimination becomes invisible

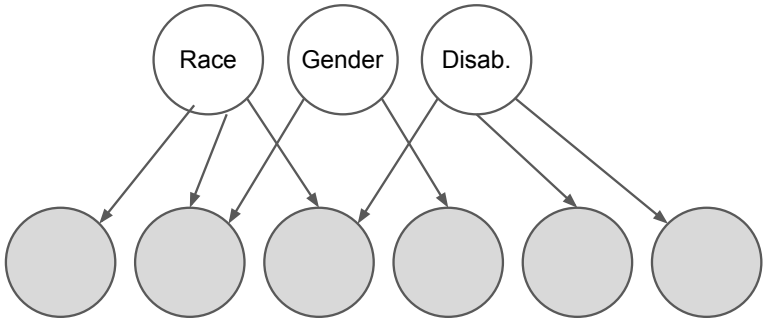
Mitigation becomes impossible

F. Pasquale (2015): The Black Box Society. Harvard University Press.

Some sources of algorithmic bias

Data as a social mirror

Protected attributes redundantly encoded in observables

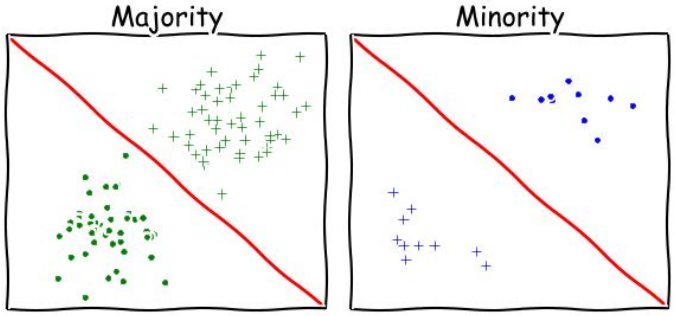


Correctness and completeness

Garbage in, garbage out (GIGO)

Sample size disparity: learn on majority

Errors concentrated in the minority class



M. Hardt (2014): "How big data is unfair". Medium.

Data mining assumptions might not hold

Data mining assumptions are not always observed in reality

- Variables might not be independently identically distributed
- Samples might be biased
- Labels might be incorrect

Errors might be concentrated in a particular class

Sometimes, we might be seeking more simplicity than what is possible

T. Calders and I. Žliobaitė (2013). Why unbiased computational processes can lead to discriminative decision procedures. Chapter 3 of: Discrimination and Privacy in the Information Society. Springer.

Two areas of concern: data and algorithms

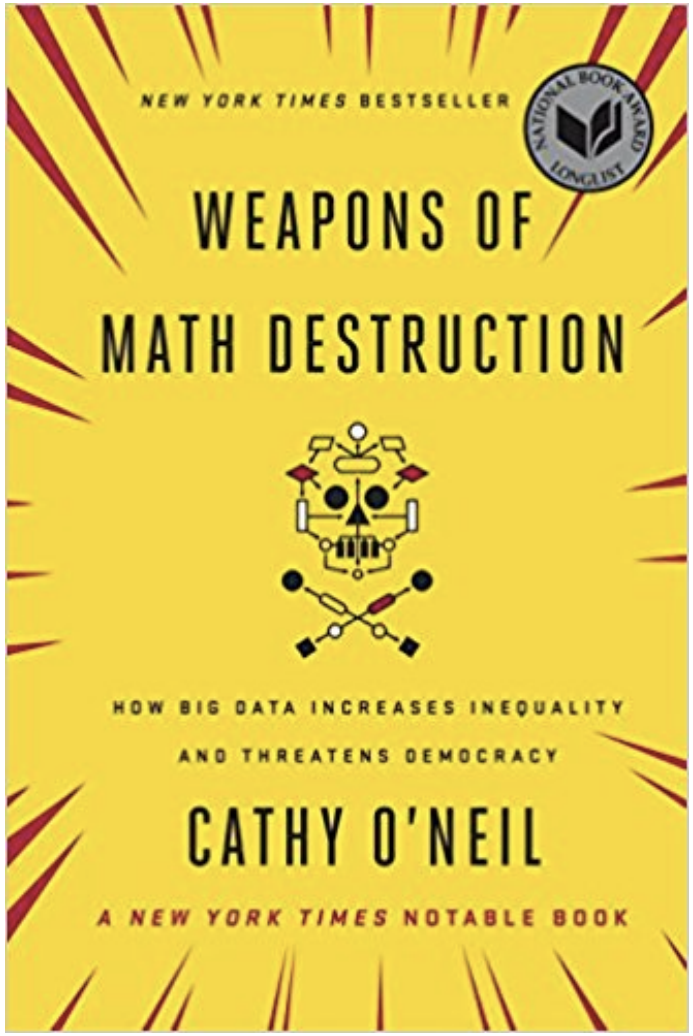
Data inputs:

- Poorly selected (e.g., observe only car trips, not bicycle trips)
- Incomplete, incorrect, or outdated
- Selected with bias (e.g., smartphone users)
- Perpetuating and promoting historical biases (e.g., hiring people that "fit the culture")

Algorithmic processing:

- Poorly designed matching systems
- Personalization and recommendation services that narrow instead of expand user options
- Decision making systems that assume correlation implies causation
- Algorithms that do not compensate for datasets that disproportionately represent populations
- Output models that are hard to understand or explain hinder detection and mitigation of bias

Executive Office of the US President (May 2016): "Big Data: A Report on Algorithmic Systems,Opportunity,and Civil Rights"


NEW YORK TIMES BESTSELLER
NATIONAL BOOK AWARD LONGLIST
WEAPONS OF MATH DESTRUCTION
HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY
CATHY O'NEIL
A NEW YORK TIMES NOTABLE BOOK