

Data, Responsibly

Serge Abiteboul

Inria & ENS Paris, France

Julia Stoyanovich

Drexel University, USA



data RESPONSIBLY

People are data, data is people



The promise of Big Data

Power

5Vs: volume, velocity, variety, veracity, value

unprecedented data collection capabilities

enormous computational power

massively parallel processing

Opportunity

improve people's lives, e.g., recommendation

accelerate scientific discovery, e.g., medicine

boost innovation, e.g., autonomous cars

transform society, e.g., open government

optimize business, e.g., advertisement targeting



goal - progress

Online price discrimination

THE WALL STREET JOURNAL.

WHAT THEY KNOW

Websites Vary Prices, Deals Based on Users' Information

By JENNIFER VALENTINO-DEVRIES,
JEREMY SINGER-VINE and ASHKAN SOLTANI

December 24, 2012

It was the same Swingline stapler, on the same Staples.com website. But for Kim Wamble, the price was \$15.79, while the price on Trude Frizzell's screen, just a few miles away, was \$14.29.

A key difference: where Staples seemed to think they were located.

WHAT PRICE WOULD YOU SEE?



lower prices offered to buyers who live in more affluent neighborhoods

<https://www.wsj.com/articles/SB10001424127887323777204578189391813881534>

Online job ads

theguardian

Samuel Gibbs

Wednesday 8 July 2015 11.29 BST

Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs



One experiment showed that Google displayed adverts for a career coaching service for executive jobs 1,852 times to the male group and only 318 times to the female group. Photograph: Alamy

The AdFisher tool simulated job seekers that did not differ in browsing behavior, preferences or demographic characteristics, except in gender.

One experiment showed that Google displayed ads for a career coaching service for “\$200k+” executive jobs **1,852 times to the male group and only 318 times to the female group**. Another experiment, in July 2014, showed a similar trend but was not statistically significant.

<https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study>

Job-screening personality tests

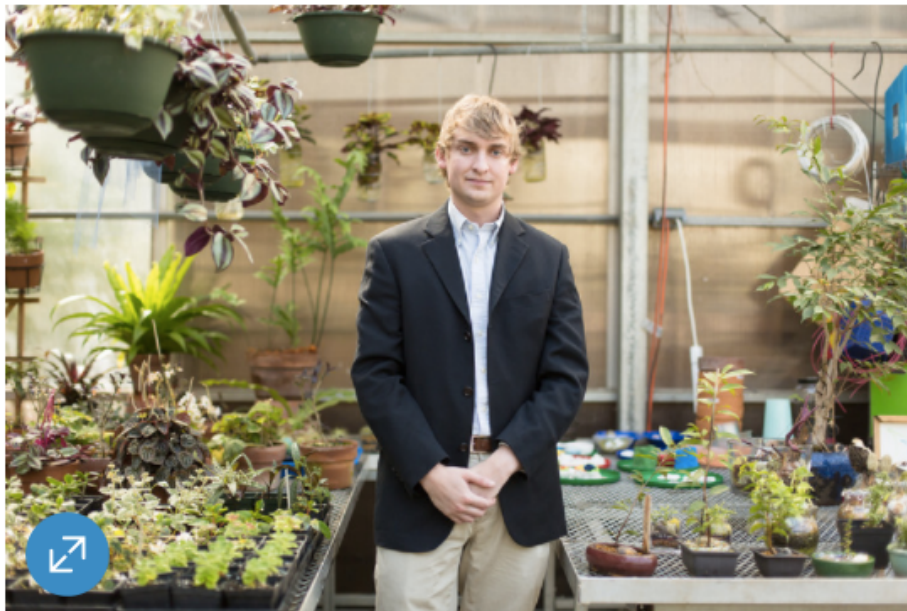
THE WALL STREET JOURNAL.

By **LAUREN WEBER** and **ELIZABETH DWOSKIN**

Sept. 29, 2014 10:30 p.m. ET

Are Workplace Personality Tests Fair?

Growing Use of Tests Sparks Scrutiny Amid Questions of Effectiveness and Workplace Discrimination



Kyle Behm accused Kroger and six other companies of discrimination against the mentally ill through their use of personality tests. TROY STAINS FOR THE WALL STREET JOURNAL

The Equal Employment Opportunity commission is **investigating whether personality tests discriminate against people with disabilities**.

As part of the investigation, officials are trying to determine if the tests **shut out people suffering from mental illnesses** such as depression or bipolar disorder, even if they have the right skills for the job.

<http://www.wsj.com/articles/are-workplace-personality-tests-fair-1412044257>

Racial bias in criminal sentencing

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

A commercial tool COMPAS automatically predicts some categories of future crime to assist in bail and sentencing decisions. It is used in courts in the US.



Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

The tool correctly predicts recidivism **61% of the time.**

Blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend.

The tool makes **the opposite mistake among whites**: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes.

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Racial bias in criminal sentencing

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Lack of diversity in data and methods

The New York Times

Artificial Intelligence's White Guy Problem

By KATE CRAWFORD JUNE 25, 2016



Like all technologies before it, artificial intelligence will reflect the values of its creators. So **inclusivity matters** — from who designs it to who sits on the company boards and which ethical perspectives are included.

Otherwise, **we risk constructing machine intelligence that mirrors a narrow and privileged vision of society**, with its old, familiar biases and stereotypes.

<http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>

Is Big Data impartial?

Claim: **Big Data is algorithmic, therefore it cannot be biased!** And yet...

- Algorithms **discriminate** just like humans do, but at a larger scale
- Processes are **opaque**, and defy public scrutiny
- It is our responsibility to understand the issues and offer **technological solutions** that address them
- Technology must be informed by **ethical** and **legal considerations**



<http://www.allenoverly.com/publications/en-gb/Pages/Protected-characteristics-and-the-perception-reality-gap.aspx>

Data, Responsibly

fairness



diversity



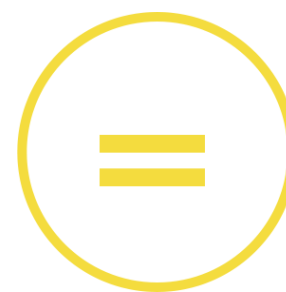
transparency



data protection



neutrality



Roadmap



- ✓ Introduction
- ➡ Responsible data analysis
 - Technical issues
 - Societal issues
 - Conclusion

Roadmap

- ✓ Introduction
- Responsible data analysis
 - ➔ Fairness
 - Diversity
 - Transparency
 - Neutrality
 - Data Protection
- Technical issues
- Societal issues
- Conclusion



Fairness is lack of bias

- Where does bias come from?
 - data collection
 - data analysis
 - result interpretation
- Analogy - scientific data analysis
 - collect a representative sample
 - do sound reproducible analysis
 - explain methodology, interpret results in context



when data is about people, bias can lead to discrimination

The evils of discrimination

Disparate treatment is the illegal practice of treating an entity differently based on a **protected characteristic** such as race, gender, age, religion, sexual orientation, or national origin.

Disparate impact is the result of systematic disparate treatment, where disproportionate **adverse impact** is observed on members of a **protected class**.



<http://www.allenoverly.com/publications/en-gb/Pages/Protected-characteristics-and-the-perception-reality-gap.aspx>



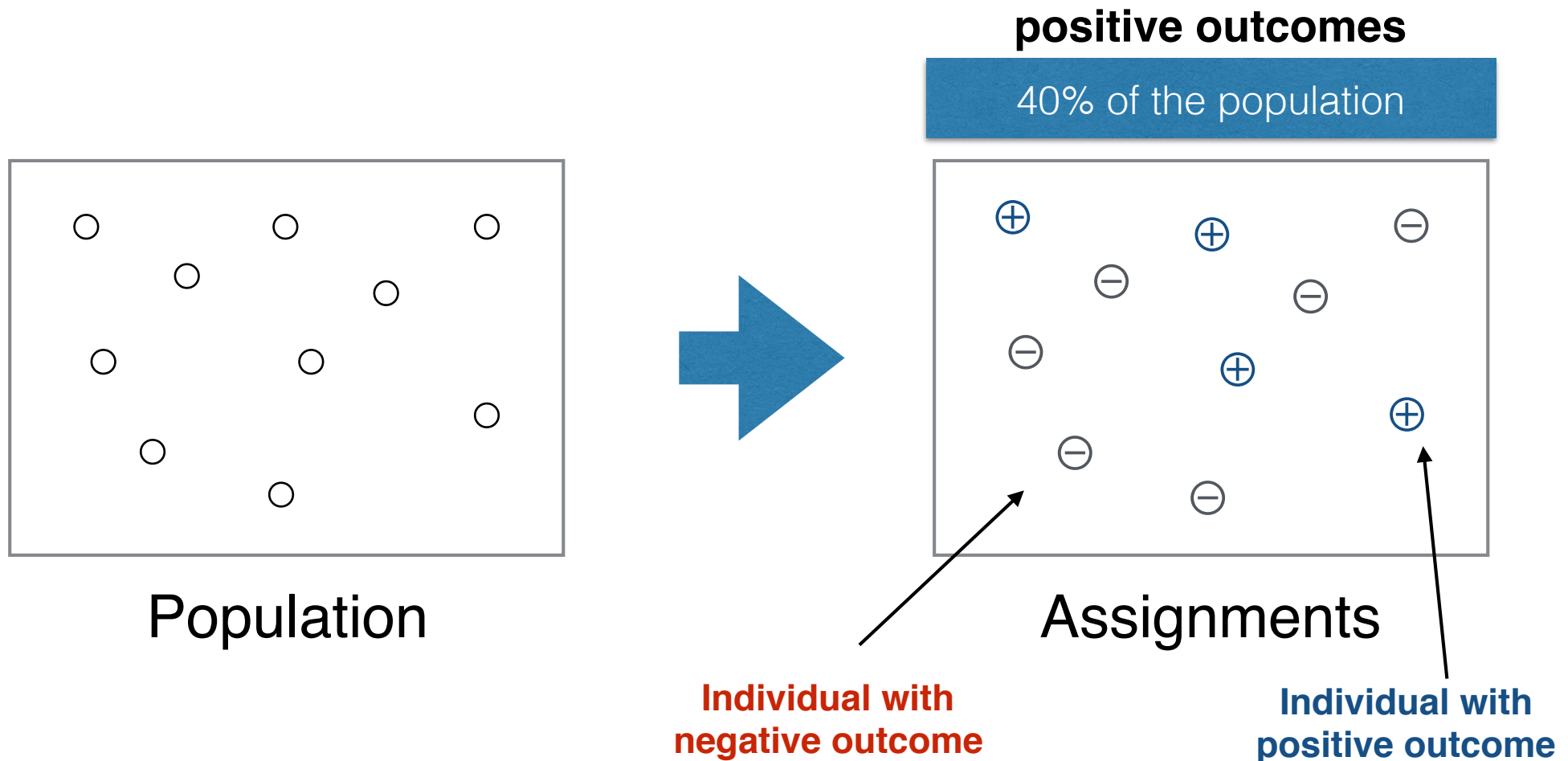
Outcomes

Consider a **vendor** assigning positive or negative **outcomes** to individuals.

Positive Outcomes	Negative Outcomes
offered employment	denied employment
accepted to school	rejected from school
offered a loan	denied a loan
offered a discount	not offered a discount

Assigning outcomes to populations

Fairness is concerned with how outcomes are assigned to a population



Sub-populations may be treated differently

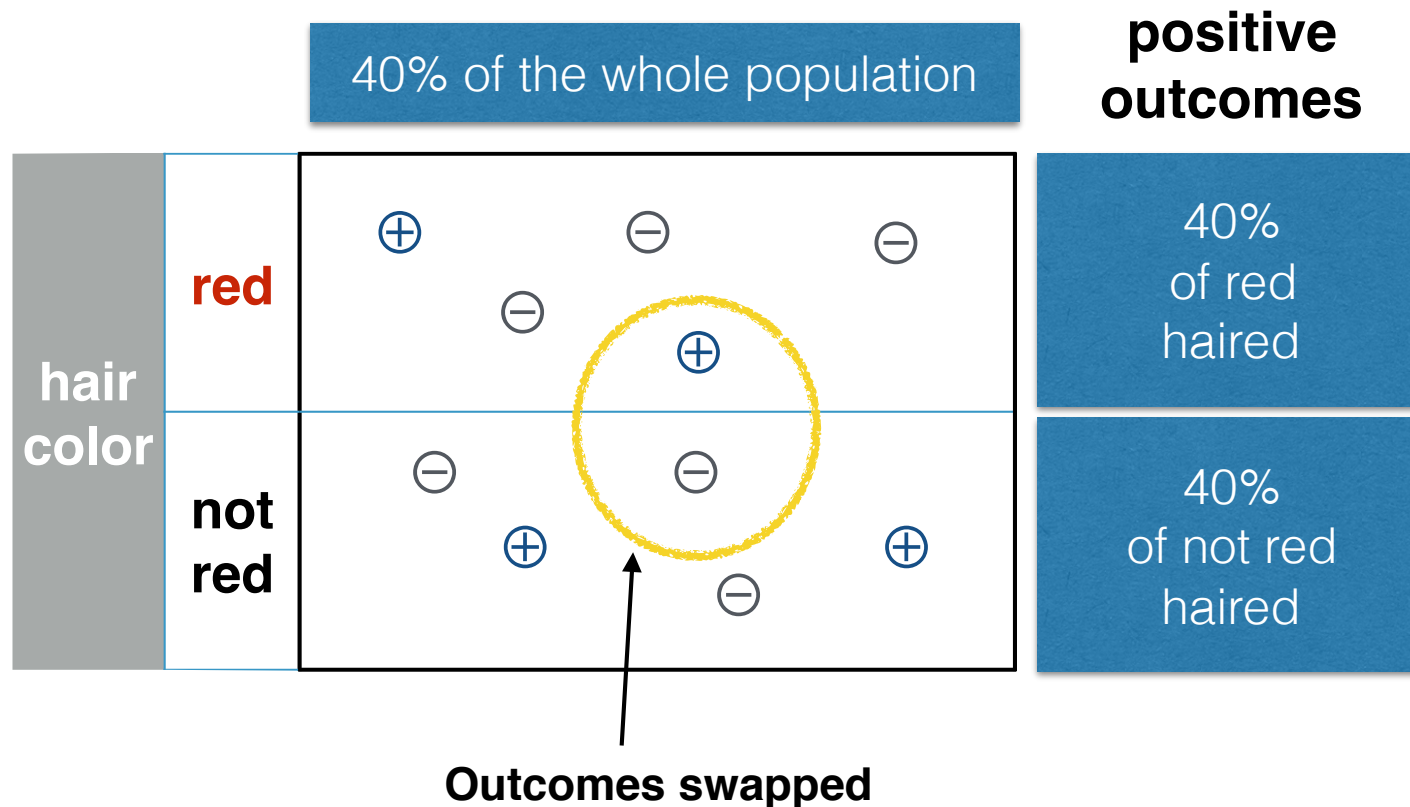
Sub-population: those with red hair
(under the same assignment of outcomes)



Enforcing statistical parity

Statistical parity (aka **group fairness**)

demographics of the individuals receiving any outcome are the same as demographics of the underlying population



Redundant encoding

Now consider the assignments under both **hair color** (protected) and **hair length** (innocuous)

		hair length		
		long	not long	
hair color	red	⊕	⊖ ⊖ ⊖ ⊖	positive outcomes 20% of red haired
	not red	⊕ ⊕ ⊕	⊖ ⊖	60% of not red haired

Deniability

The vendor has adversely impacted red-haired people, but claims that outcomes are assigned according to hair length.

Blinding does not imply fairness

Removing **hair color** from the vendor's assignment process does not prevent discrimination

		hair length		
		long	not long	
hair color	red	⊕	⊖ ⊖ ⊖ ⊖	positive outcomes 20% of red haired
	not red	⊕ ⊕ ⊕	⊖ ⊖	60% of not red haired

Assessing disparate impact

Discrimination is assessed by the effect on the protected sub-population, not by the input or by the process that lead to the effect.

Redundant encoding

Let's replace hair color with **race** (protected),
hair length with **zip code** (innocuous)

		zip code	
		10025	10027
race	black	⊕	⊖ ⊖ ⊖ ⊖
	white	⊕ ⊕ ⊕	⊖ ⊖

positive outcomes

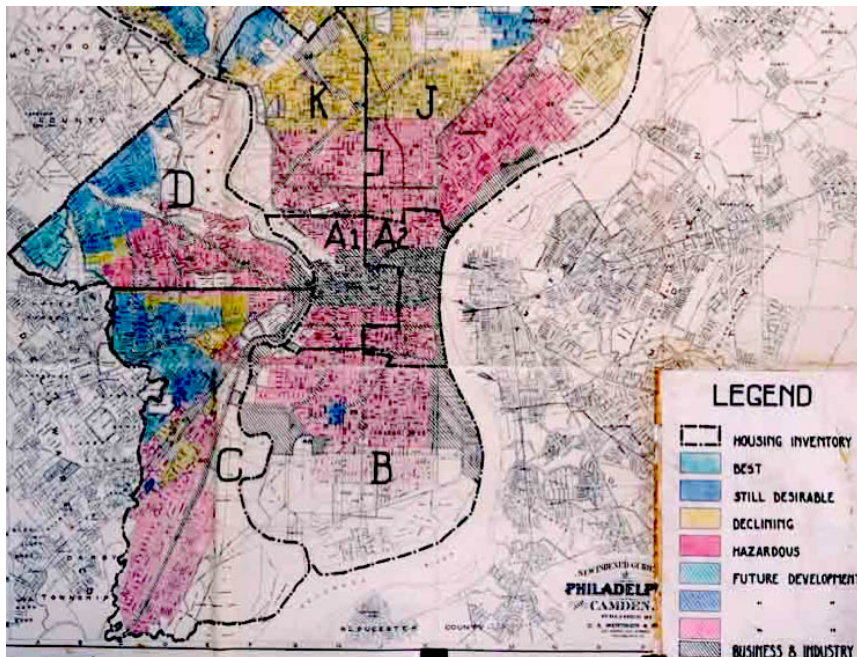
20%
of black

60%
of white

Redlining

Redlining is the practice of arbitrarily denying or limiting financial services to specific neighborhoods, generally because its residents are people of color or are poor.

Philadelphia, 1936



wikipedia

Households and businesses in the red zones could not get mortgages or business loans.

Discrimination may be unintended

Staples website estimated user's location, **offering discounts** to those near rival stores, leading to discrimination w.r.t. to average income.

		rival store proximity		
		close	far	
income	low	⊕	⊖ ⊖ ⊖ ⊖	positive outcomes 20% of low income
	high	⊕ ⊕ ⊕	⊖ ⊖	60% of high income

Discrimination
Whether intentional or not, discrimination is unethical and, in many countries, illegal.

Redundant encoding in criminal sentencing

In the ProPublica criminal sentencing investigation, **race was not one of the input features**

Input from 137 questions in categories: current charges, criminal history, non-compliance history, family criminality, peers, substance abuse, residence/stability, social environment, education, vocation, leisure/recreation, social isolation, criminal personality, anger, criminal attitudes.

Examples:

- “Was one of your parents ever sent to jail or prison?”
- “How many of your friends/acquaintances are taking drugs illegally?”
- Agree or disagree with: “A hungry person has a right to steal”

Imposing statistical parity

May be contrary to the goals of the vendor

positive outcome: offered a loan

		credit score	
		good	bad
race	black	\oplus	\ominus \ominus \oplus \ominus
	white	\oplus \ominus \oplus	\ominus \ominus

positive outcomes

40% of black

40% of white

Impossible to predict loan payback accurately.
Use past information, may itself be biased.

Justifying exclusion

Self-fulfilling prophecy

deliberately choosing the “wrong” (lesser qualified) members of the protected group to build bad track record

		credit score		
		good	bad	
race	black	⊕	⊖ ⊕ ⊖ ⊖	40% of black
	white	⊕ ⊖ ⊕	⊖ ⊖	40% of white

Effect on sub-populations

Simpson's paradox

disparate impact at the full population level disappears or reverses when looking at sub-populations!

		grad school admissions	
		admitted	denied
gender	F	1512	2809
	M	3715	4727

positive outcomes

35% of women

44% of men

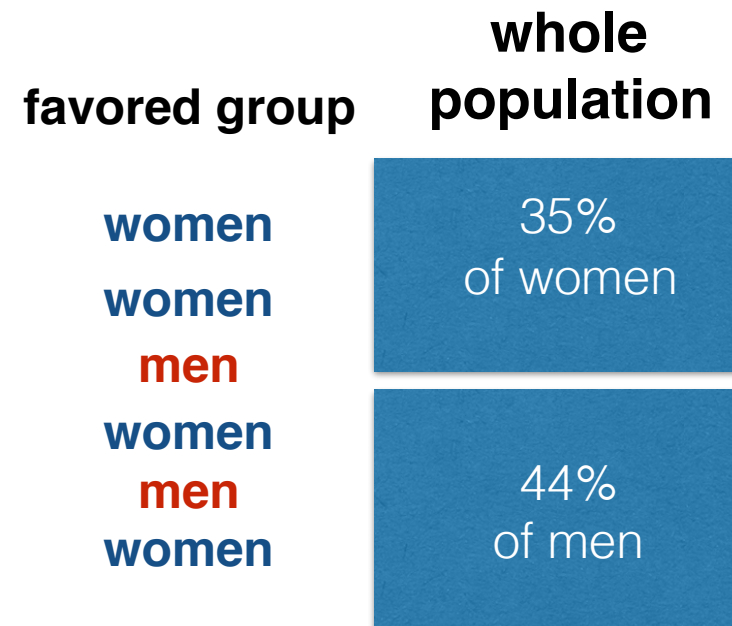
UC Berkeley 1973: it appears men were admitted at higher rate.

Effect on sub-populations

Simpson's paradox

disparate impact at the full population level disappears or reverses when looking at sub-populations!

Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	373	6%	341	7%

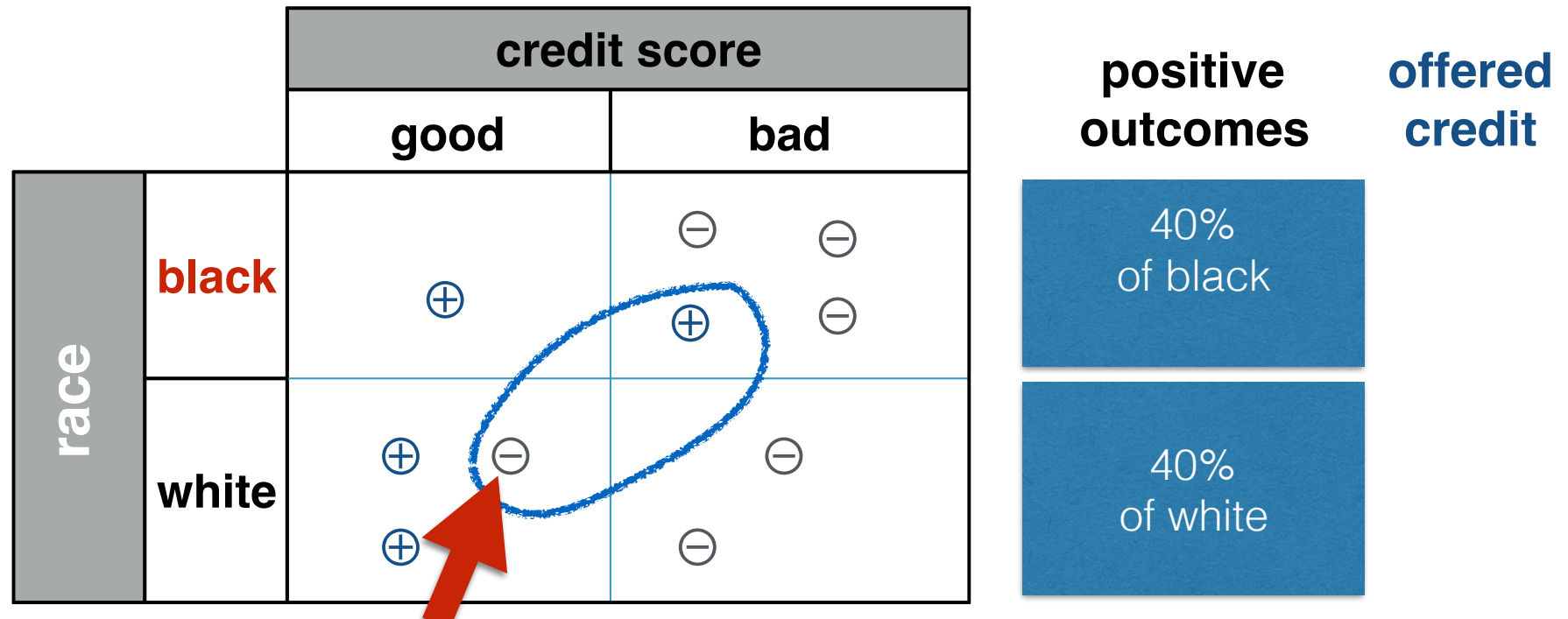


UC Berkeley 1973: women applied to more competitive departments, with low rates of admission among qualified applicants.

Is statistical parity sufficient?

Statistical parity (aka group fairness)

demographics of the individuals receiving any outcome are the same as demographics of the underlying population



Individual fairness

any two individuals who are similar w.r.t. a particular task should receive similar outcomes

Discrimination-aware data analysis

- **Detecting discrimination**

- mining for discriminatory patterns in (input) data
- verifying data-driven applications

- **Preventing discrimination**

- data pre-processing
- model post-processing
- model regularization

[Ruggieri *et al.*; 2010]

[Luong *et al.*; 2011]

[Pedresci *et al.*; 2012]

[Romei *et al.*; 2012]

[Hajian & Domingo-Ferrer; 2013]

[Mancuhan & Clifton; 2014]

[Kamiran & Calders; 2009]

[Kamishima *et al.*; 2011]

[Mancuhan & Clifton; 2014]

[Feldman *et al.*; 2015]

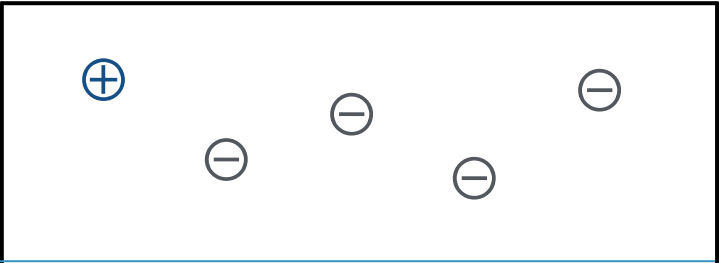
[Dwork *et al.*; 2012]

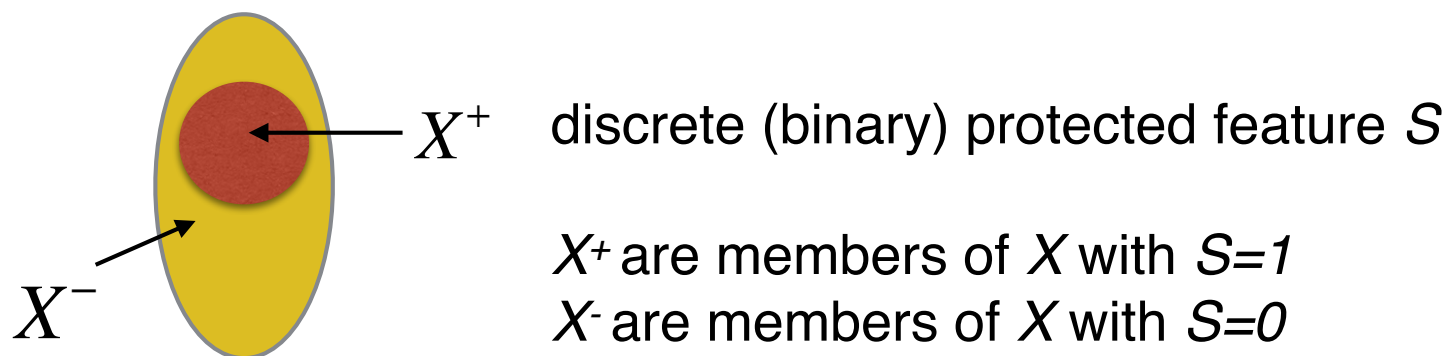
[Zemel *et al.*; 2013]

both rely on discrimination criteria

many more....

How do we quantify discrimination?

		40% of the whole population	positive outcomes	$Y = 1$
hair color	red		20% of red hair	$Y = 1 X^+$
	not red		60% of not red hair	$Y = 1 X^-$



Discrimination criteria

[I. Zliobaite, CoRR abs/1511.00148 (2015)]

- **Statistical tests** check how likely the difference between groups is due to chance - *is there discrimination?*
- **Absolute measures** express the absolute difference between groups, quantifying the *magnitude of discrimination*
- **Conditional measures** express how much of the difference between groups cannot be *explained by other attributes*, while also quantifying the *magnitude of discrimination*
- **Structural measures** *how wide-spread is discrimination?*
Measures the number of individuals impacted by direct discrimination.

Discrimination measures

[I. Zliobaite, CoRR abs/1511.00148 (2015)]

a proliferation of task-specific measures

Table III. Summary of absolute measures. Checkmark (✓) indicates that it is directly applicable in a given machine learning setting. Tilde (~) indicates that a straightforward extension exists (for instance, measuring pairwise).

Measure	Protected variable			Target variable		
	Binary	Categoric	Numeric	Binary	Ordinal	Numeric
Mean difference	✓	~		✓		✓
Normalized difference	✓	~		✓		
Area under curve	✓	~		✓	✓	✓
Impact ratio	✓	~		✓		
Elift ratio	✓	~		✓		
Odds ratio	✓	~		✓		
Mutual information	✓	✓	✓	✓	✓	✓
Balanced residuals	✓	~		~	✓	✓
Correlation	✓		✓	✓		✓

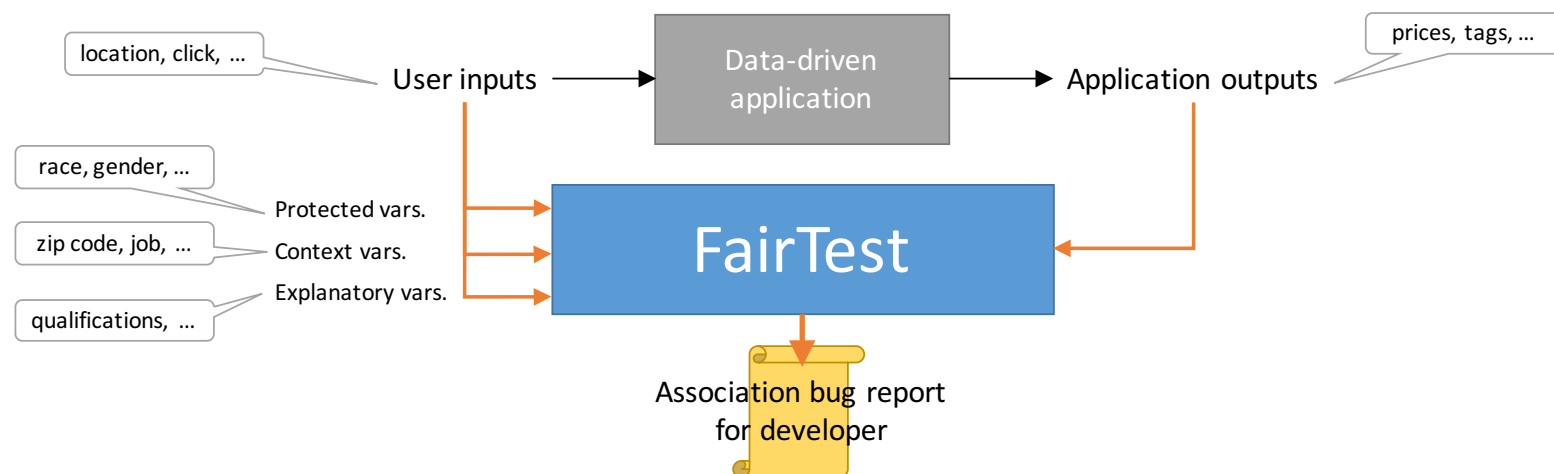
used for statistical parity:

$$\frac{\% \text{ of } + \text{ for protected class}}{\% \text{ of } + \text{ for population}}$$

FairTest: identifying discrimination

[F. Tramèr *et al.*, arXiv:1510.02377 (2015)]

- A test suite for data analysis applications
- Tests for **unintentional discrimination** according to several representative discrimination measures
- Automates search for **context-specific associations** between protected variables and application outputs
- Report findings, ranked by association **strength** and affected **population size**



FairTest: discrimination measures

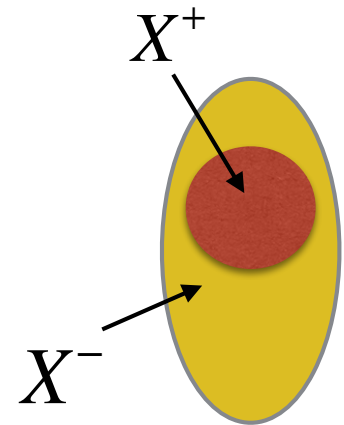
[F. Tramèr *et al.*, arXiv:1510.02377 (2015)]

Binary ratio / difference compares probabilities of a single output for two groups

$$\Pr(Y = 1 | X^+) - \Pr(Y = 1 | X^-)$$

Easy to extend to non-binary outputs,
not easy to overcome binary
protected class membership

$$\frac{\Pr(Y = 1 | X^+)}{\Pr(Y = 1 | X^-)} - 1$$



Mutual information measures statistical dependence between outcomes and protected group membership

Works for non-binary outputs, class membership,
can be normalized; bad for continuous values,
does not incorporate order among values

$$\sum \Pr(y, s) \ln \frac{\Pr(y, s)}{\Pr(y) \Pr(s)}$$

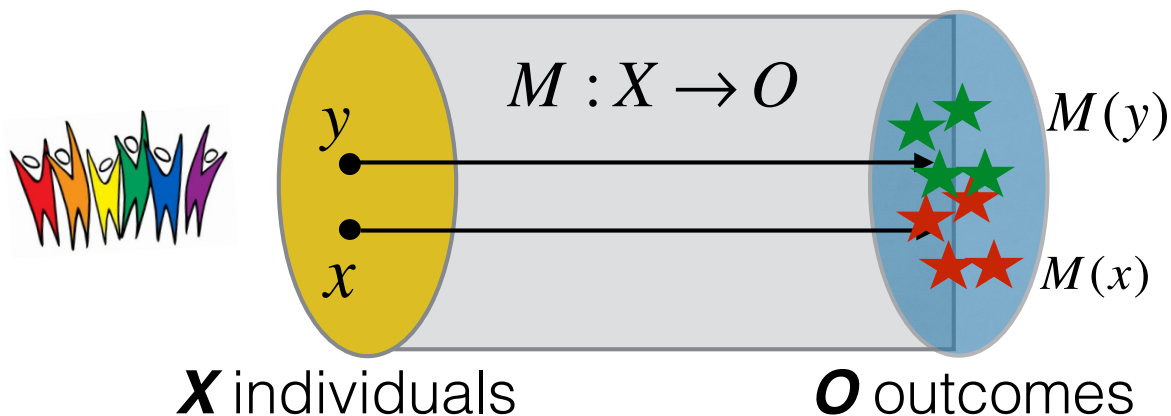
Pearson's correlation measures strength of linear relationship between outcomes and protected group membership

Works well for ordinal and continuous values, may detect non-linear correlations, is easy to interpret; finding a 0 correlation does not imply that S and Y are independent

Fairness through awareness

[C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. S. Zemel; *ITCS 2012*]

Fairness: Individuals who are **similar** for the purpose of classification task should be **treated similarly**.



A task-specific similarity metric is given $d(x, y)$

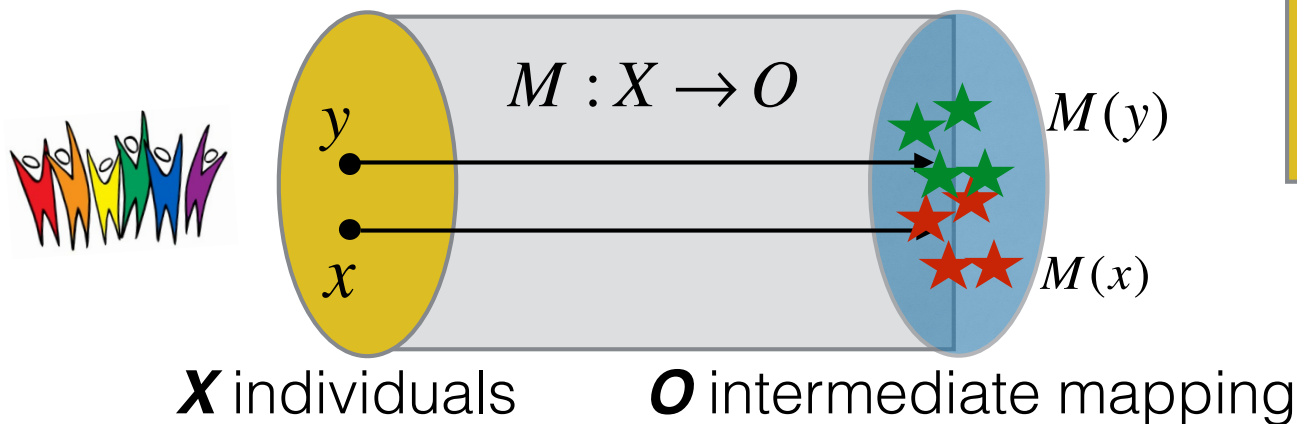


$M : X \rightarrow O$ is a **randomized mapping**: an individual is mapped to a distribution over outcomes

Fairness through a Lipschitz mapping

[C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. S. Zemel; *ITCS 2012*]

Individuals who are **similar** for the purpose of classification task should be **treated similarly**.



A task-specific similarity metric is given $d(x, y)$

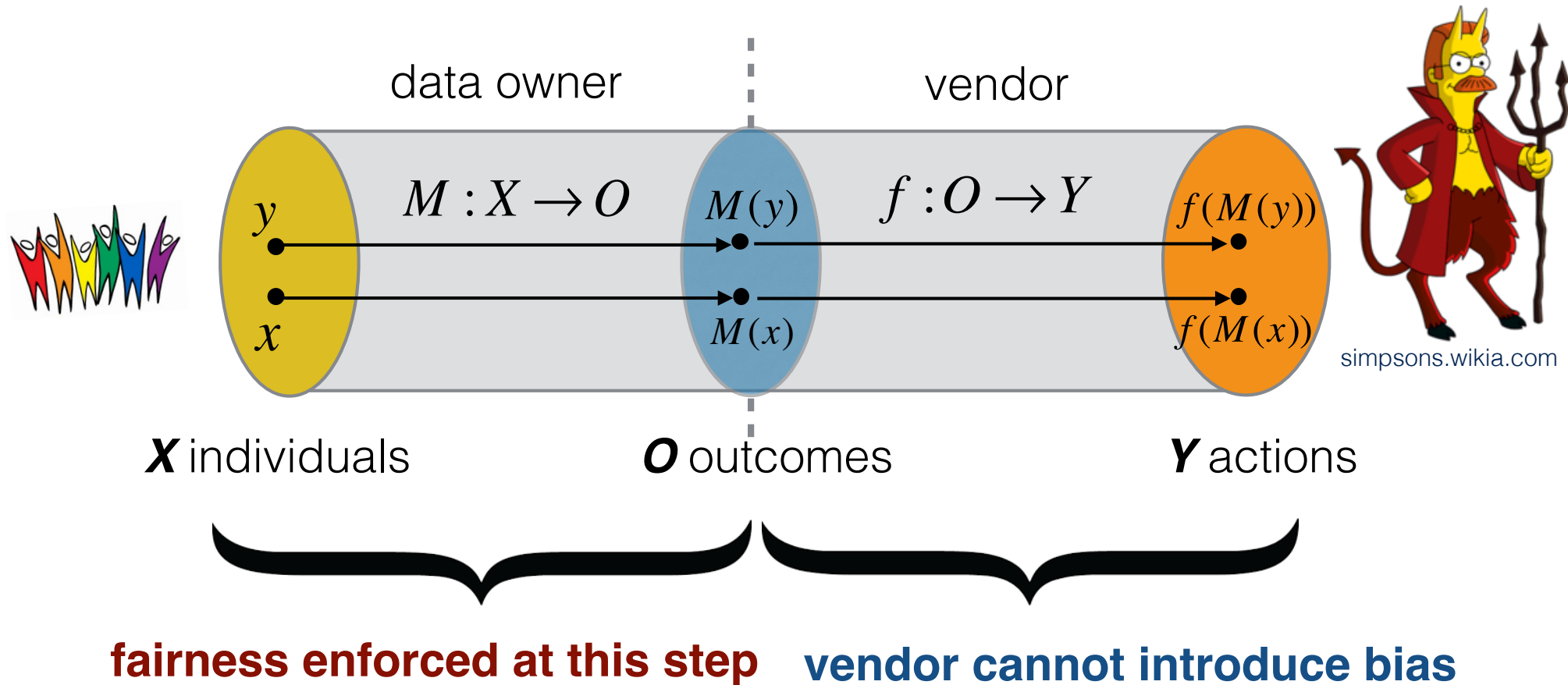


M is a Lipschitz mapping if $\forall x, y \in X \quad \|M(x), M(y)\| \leq d(x, y)$

close individuals map to close distributions

Fairness through awareness

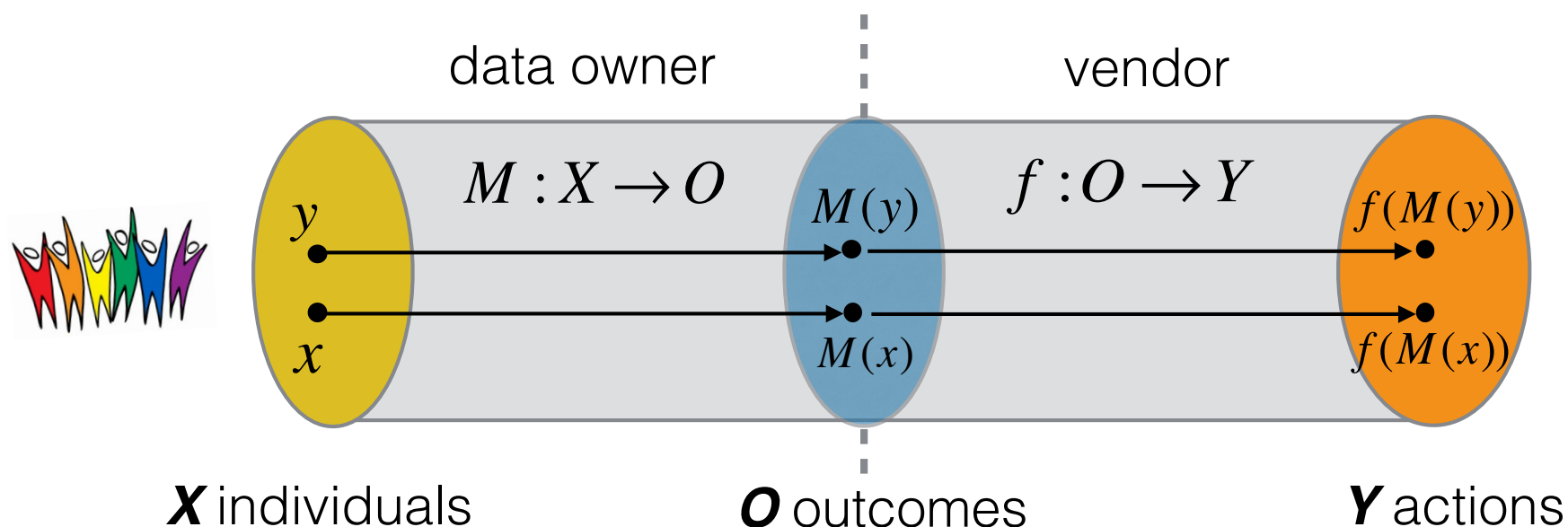
[C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. S. Zemel; *ITCS 2012*]



What about the vendor?

[C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. S. Zemel; *ITCS 2012*]

Vendors can efficiently maximize expected utility,
subject to the Lipschitz condition



Computed with a linear program of size $\text{poly}(|X|, |Y|)$

the same mapping can be used by multiple vendors

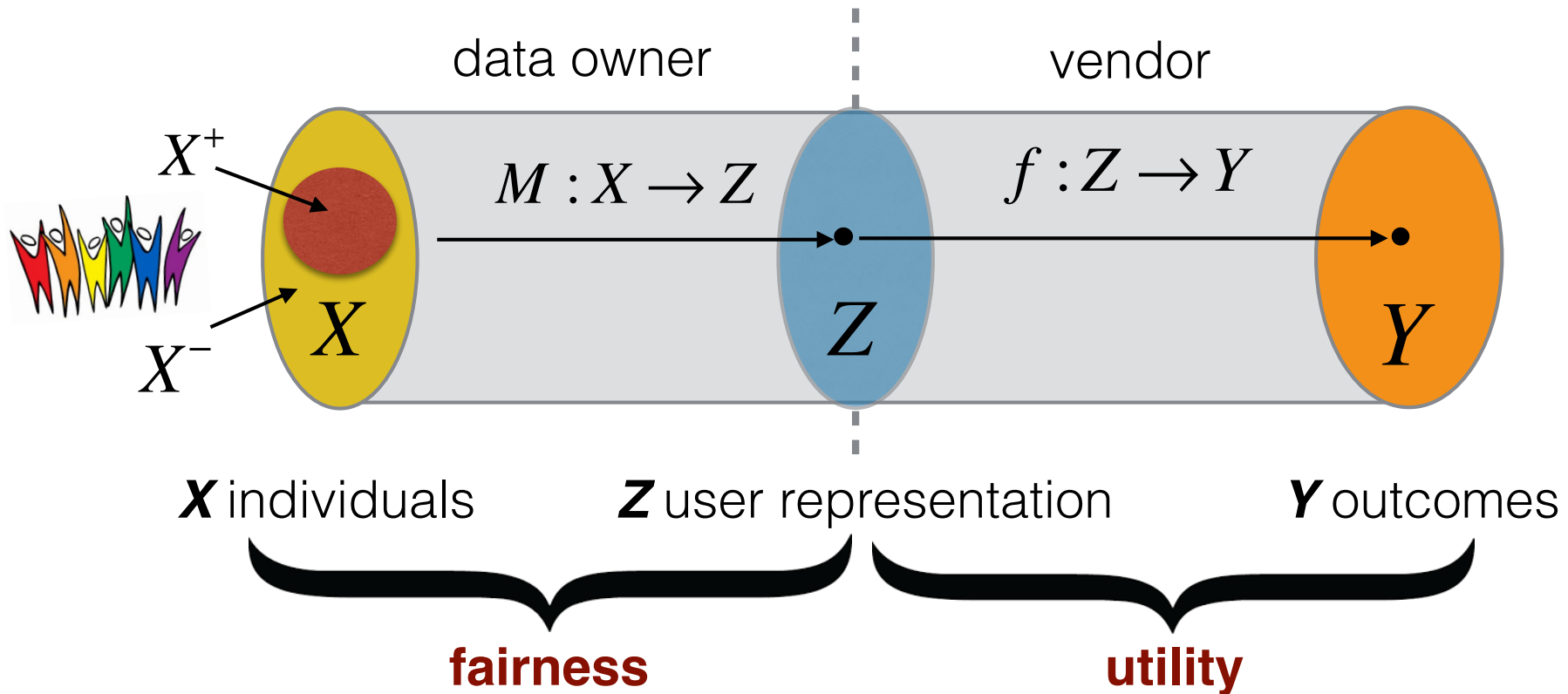
Fairness through awareness: summary

[C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. S. Zemel; *ITCS 2012*]

- An early work in this space, proposes a principled data pre-processing approach
- Stated as an **individual fairness** condition but also sometimes leads to **group fairness**
- Relies on an externally-supplied task-specific similarity metric - magic!
- Is not formulated as a learning problem, does not generalize to unseen data

Learning fair representations

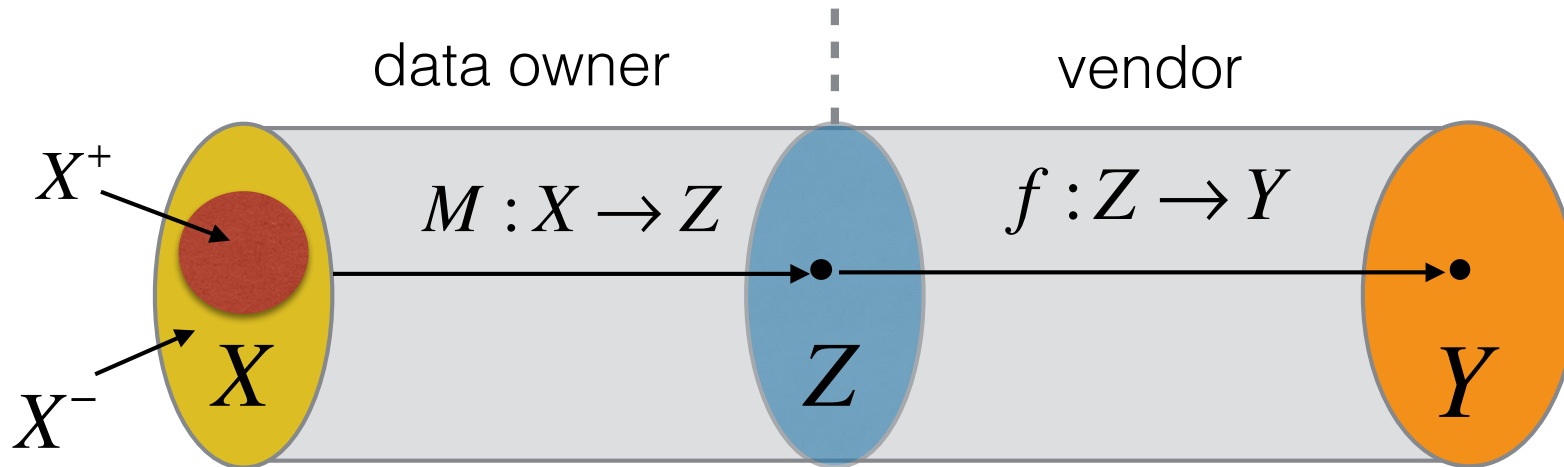
[R. S. Zemel, Y. Wu, K. Swersky, T. Pitassi, C. Dwork; *ICML 2013*]



Idea: remove reliance on a “fair” similarity measure, instead **learn** representations of individuals, distances

Fairness and utility

[R. S. Zemel, Y. Wu, K. Swersky, T. Pitassi, C. Dwork; *ICML 2013*]



Learn a **randomized mapping** $M(X)$ to a set of K prototypes Z

$M(X)$ should lose information about membership in S $P(Z | S = 0) = P(Z | S = 1)$

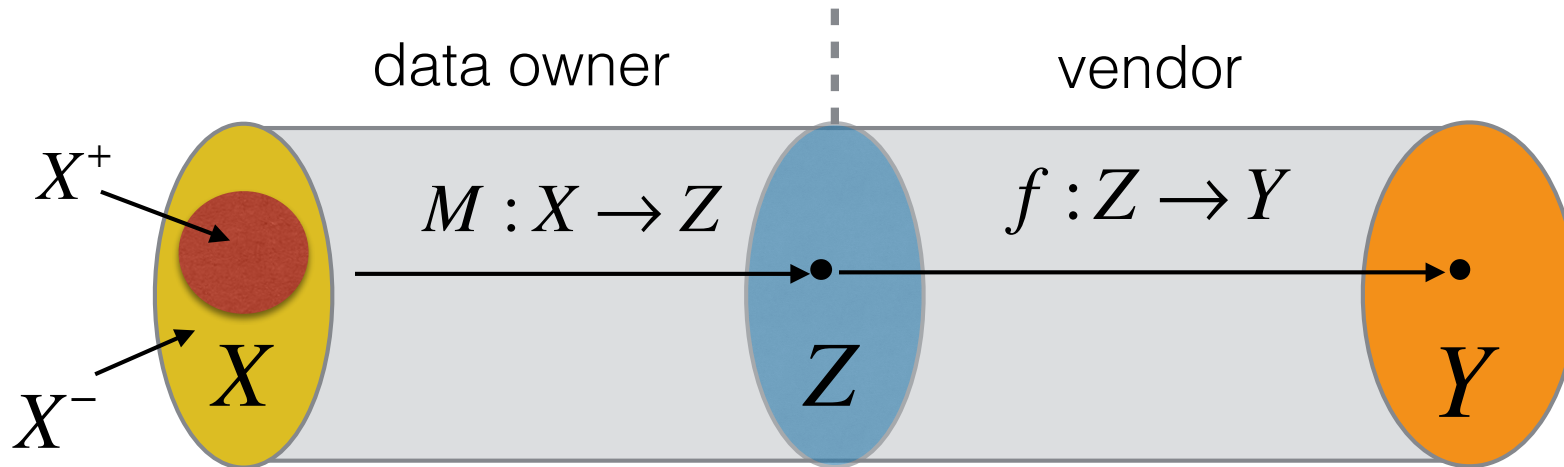
$M(X)$ should preserve other information so that vendor can maximize utility

$$L = A_z \cdot L_z + A_x \cdot L_x + A_y \cdot L_y$$

group fairness \nearrow **individual fairness** \nwarrow **utility**

The objective function

[R. S. Zemel, Y. Wu, K. Swersky, T. Pitassi, C. Dwork; *ICML 2013*]



$$L = A_z \cdot L_z + A_x \cdot L_x + A_y \cdot L_y$$

group fairness (points to A_z) **individual fairness** (points to A_x) **utility** (points to A_y)

$$P_k^+ = P(Z = k \mid x \in X^+)$$

$$P_k^- = P(Z = k \mid x \in X^-)$$

$$L_z = \sum_k |P_k^+ - P_k^-| \quad L_x = \sum_n (x_n - \hat{x}_n)^2$$

$$L_y = \sum_n -y_n \log \hat{y}_n - (1 - y_n) \log(1 - \hat{y}_n)$$

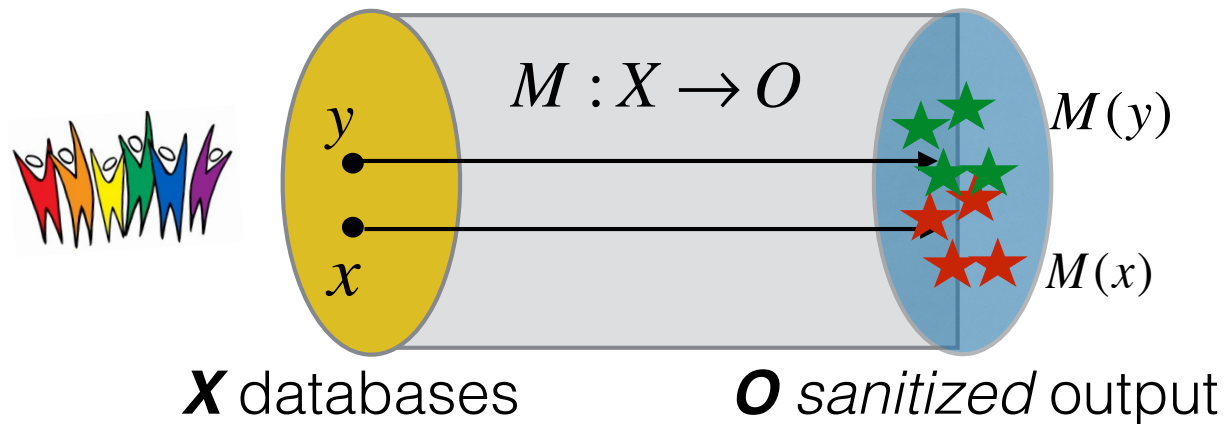
Learning fair representations: summary

[R. S. Zemel, Y. Wu, K. Swersky, T. Pitassi, C. Dwork; *ICML 2013*]

- A principled learning framework in the data pre-processing / classifier regularization category
- **Evaluation** of accuracy, discrimination (group fairness) and consistency (individual fairness), promising results on real datasets
- Not clear how to set K , so as to trade off accuracy / fairness
- The mapping is **task-specific**

Connection to privacy

Fairness through awareness generalizes differential privacy



close databases map to close output distributions

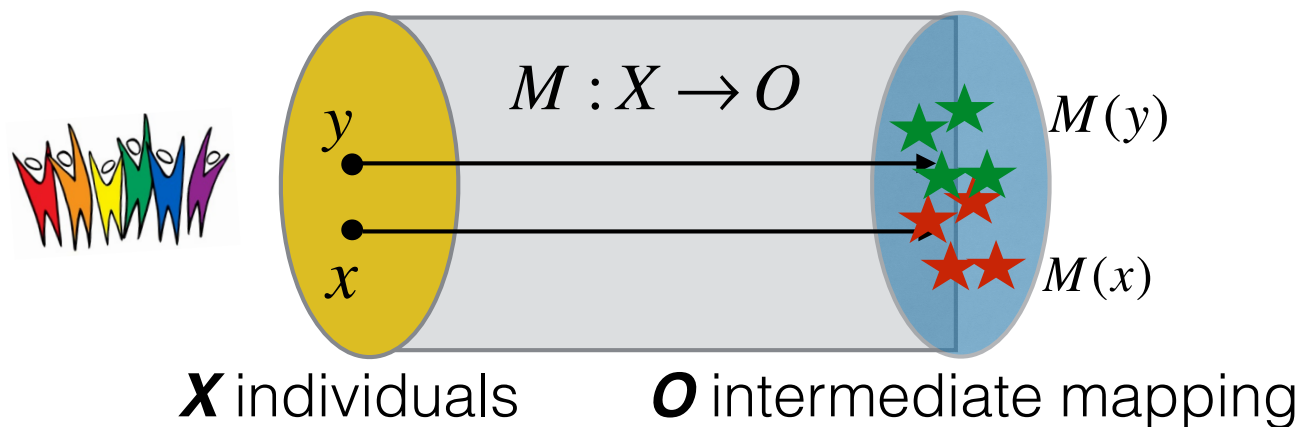


Databases that differ in one record.

Connection to privacy

Does the fairness mapping provide privacy?

Similar individuals (according to $d(x,y)$) are hard to distinguish in the intermediate mapping. This provides a form of protection similar to anonymity based privacy.



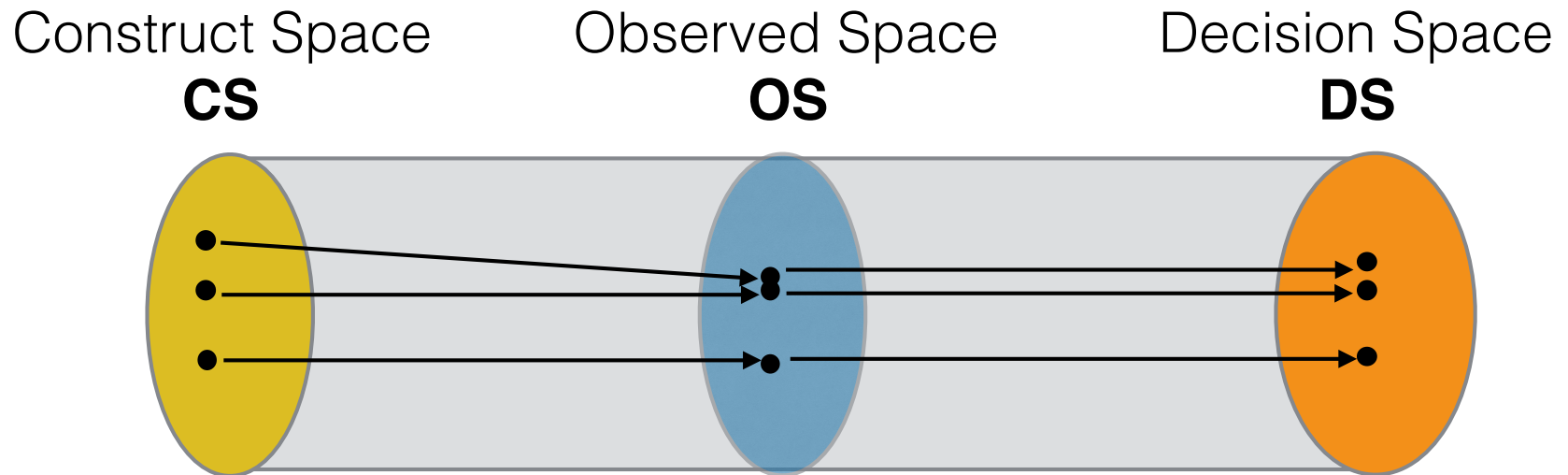
It depends on the metric d and on whether individual similarity is based on sensitive properties.

On the (im)possibility of fairness

[S. Friedler, C. Scheidegger and S. Venkatasubramanian, arXiv:1609.07236v1 (2016)]

Goal: tease out the difference between *beliefs* and *mechanisms* that logically follow from those beliefs.

Main insight: To study algorithmic fairness is to study the interactions between different spaces that make up the decision pipeline for a task



Examples of features and outcomes

[S. Friedler, C. Scheidegger and S. Venkatasubramanian, arXiv:1609.07236v1 (2016)]

Construct Space	Observed Space	Decision Space
intelligence	SAT score	performance in college
grit	high-school GPA	
propensity to commit crime	family history	recidivism
risk-averseness	age	

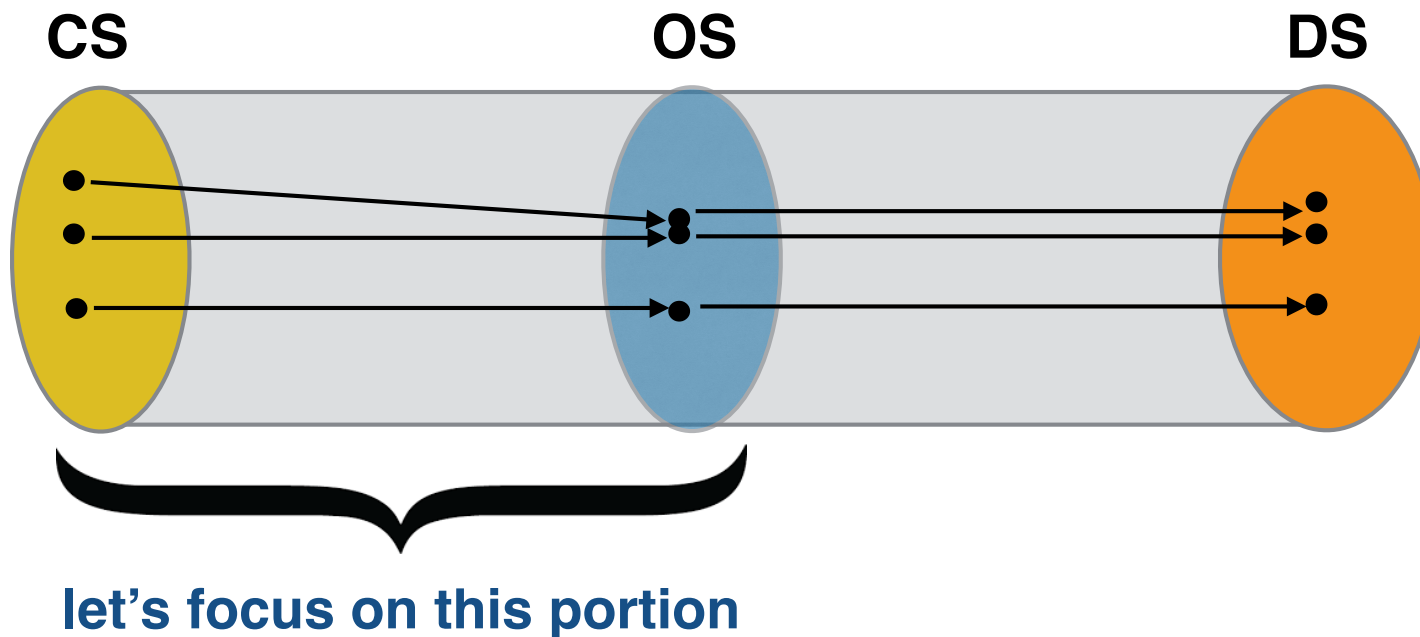
**define fairness through properties of mappings
between CS, OS and DS**

Fairness through mappings

[S. Friedler, C. Scheidegger and S. Venkatasubramanian, arXiv:1609.07236v1 (2016)]

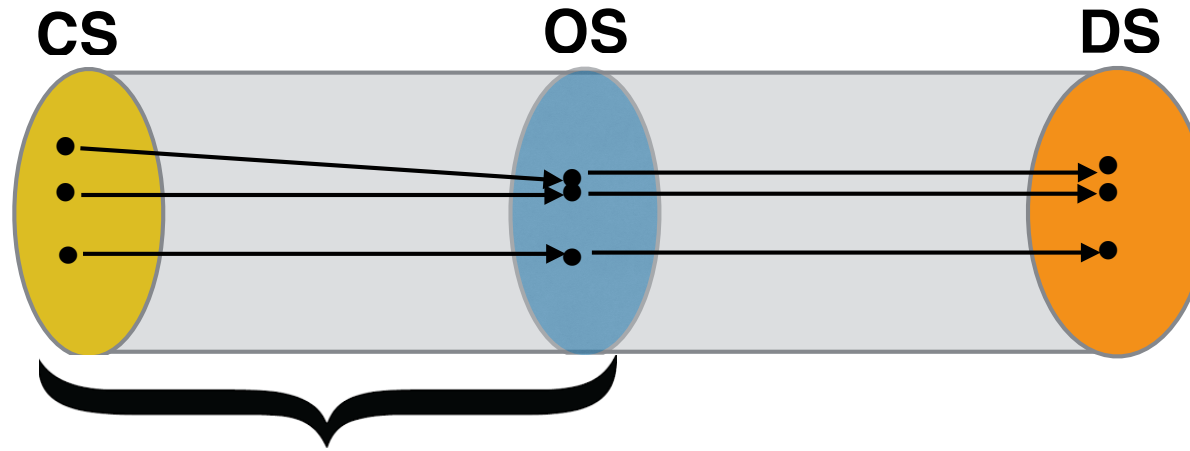
Fairness: a mapping from **CS** to **DS** is (ϵ, ϵ') -fair if two objects that are no further than ϵ in **CS** map to objects that are no further than ϵ' in **DS**.

$$f : CS \rightarrow DS \qquad d_{CS}(x, y) < \epsilon \Rightarrow d_{DS}(f(x), f(y)) < \epsilon'$$



Individual fairness

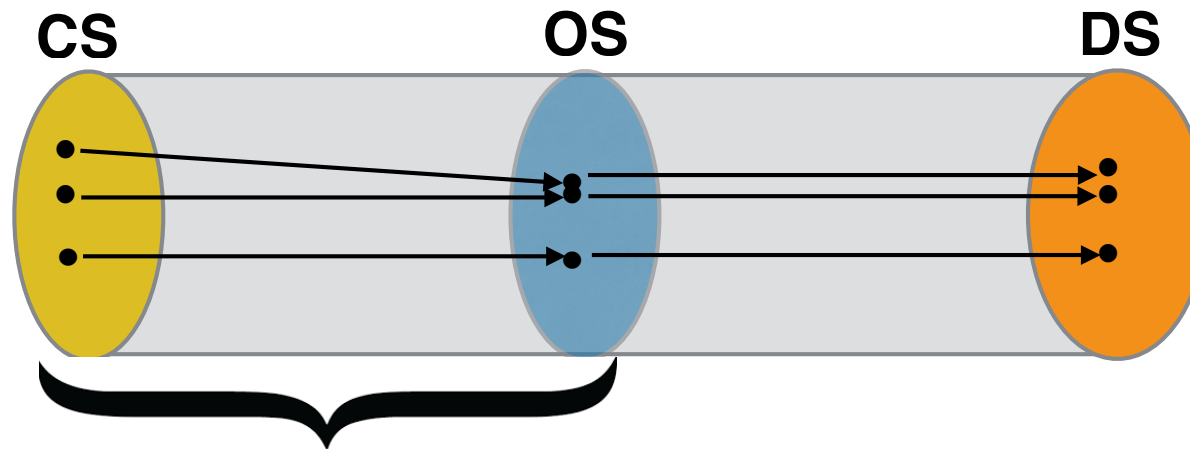
[S. Friedler, C. Scheidegger and S. Venkatasubramanian, arXiv:1609.07236v1 (2016)]



What you see is what you get (**WYSIWYG**): there exists a mapping from **CS** to **OS** that has low distortion. That is, we believe that OS faithfully represents CS. **This is the individual fairness world view.**

Group fairness

[S. Friedler, C. Scheidegger and S. Venkatasubramanian, arXiv:1609.07236v1 (2016)]



We are all equal (**WAE**): the mapping from CS to OS introduces **structural bias** - there is a distortion that aligns with the group structure of CS. **This is the group fairness world view.**

Structural bias examples: SAT verbal questions function differently in the African-American and in the Caucasian subgroups in the US. Other examples?

Two notions of fairness

individual fairness



equality

group fairness



equity

two intrinsically different world views

Roadmap

- ✓ Introduction
- Properties of responsible data analysis
 - ✓ Fairness
 - ➔ Diversity
 - Transparency
 - Neutrality
 - Data protection
- Common technical threads
- Societal issues
- Conclusion



Diversity & friends

- For a given user consuming information in search and recommendation, relevance is important, but so are:
 - **diversity** - avoid returning similar items
 - **novelty** - avoid returning known items
 - **serendipity** - surprise the user with unexpected items
- For a set of users
 - uncommon information needs must be met: **less popular**
“in the tail” queries constitute the overwhelming majority
 - lack of diversity can lead to **exclusion**



Jonas Lerman: “... the nonrandom, systematic omission of people who live on big data’s margins, whether due to poverty, geography, or lifestyle...”



Online dating

[J. Stoyanovich, S. Amer-Yahia, T. Milo; EDBT 2011]

Dating query: female, 40 or younger, at least some college, in order of decreasing income

Results are homogeneous at top ranks

Both the seeker (asking the query) and the matches (results) are dissatisfied

Crowdsourcing, crowdfunding, ranking of Web search results, ... - all subject to this problem

the rich get richer, the poor get poorer



MBA, 40 years old
makes \$150K



MBA, 40 years old
makes \$150K



MBA, 40 years old
makes \$150K



MBA, 40 years old
makes \$150K

... 999 matches



PhD, 36 years old
makes \$100K

... 9999 matches



BS, 27 years old
makes \$80K

Diversity models

Given a set of items I , select a diverse set of items S of size k , as quantified by diversity measure div .

$$S = \operatorname{argmax}_{S' \subseteq I, |S'|=k} div(S')$$

Diversity is an aspect of **quality of a collection** of items S . It is often traded off with **per-item quality** (utility).

Other variants: **aggregate diversity** (e.g., diversify recommendations to each user and across users) and **bundle diversity** (e.g., dinner and a movie).

Diversity measures

[M. Drosou, HV Jagadish, E. Pitoura, J. Stoyanovich; BigData 2017]

- **Distance-based**: the most common
 - **MaxSum**: maximize total (or average) pair-wise distance in S
 - **MaxMin**: maximize lowest pair-wise distance in S , a variant of the p-dispersion problem
- **Coverage-based**: “Noah’s Arc”, based on a set of pre-defined discrete categories (topics, demographics...)
- **Novelty-based**: relative to elements seen in the past (e.g., Maximal Marginal Relevance - MMR)

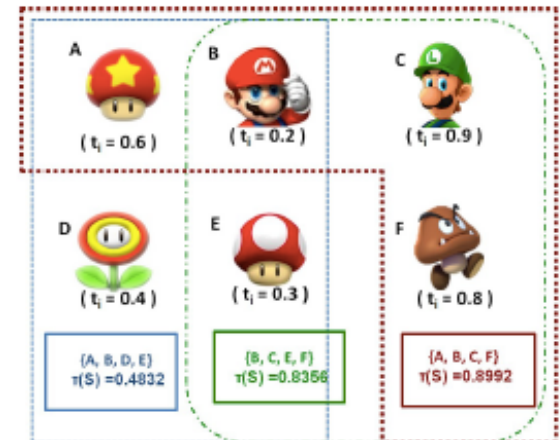
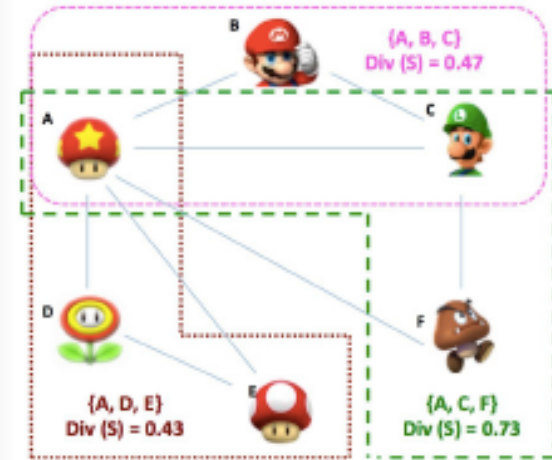
see our survey for diversity models, measures, algorithms

<https://www.cs.drexel.edu/~julia/documents/big.2016.0054.pdf>

Diversity can improve accuracy


[T. Wu, L. Chen, P. Hui, C.J. Zhang, W. Li; PVLDB 2015]

- Importance of diversity of opinion for **accuracy** is well-understood in the social sciences
 - Diversity is crucial in crowdsourcing, see Surowiecki “*The Wisdom of the Crowds*” 2005
 - The “Diversity trumps ability theorem”
- Crowd diversity: an aggregate of pair-wise diversity
- **S-Model**: similarity-driven / task-independent
- **T-Model**: task-driven, opinions are probabilistic




Diversity can improve user engagement


[J. Stoyanovich, S. Amer-Yahia, T. Milo; EDBT 2011]




MBA, 40 years old
makes \$150K



MBA, 40 years old
makes \$150K




MBA, 40 years old
makes \$150K




MBA, 40 years old
makes \$150K

... 999 matches



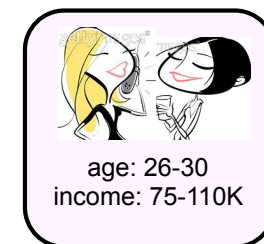
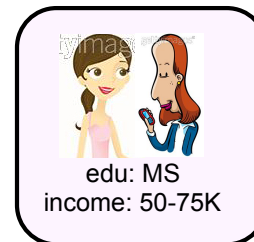
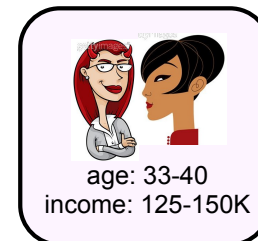
PhD, 36 years old
makes \$100K

... 9999 matches



BS, 27 years old
makes \$80K

Return clusters that expose **best from among comparable** items (profiles) w.r.t. user preferences



More diverse items seen, and liked, by users

Users are more engaged with the system

Why is diversity important?

- Unlike statistical parity and individual fairness, there is **no legal reason** to enforce diversity
- However, there are strong **utilitarian reasons**: diversity leads to better user satisfaction (IR, recommendation), higher quality of results (crowdsourcing, team formation), more efficient resource allocation (matchmaking)
- Further, diversity levels the playing field and **improves fairness in the long run**



there is, as of yet, no technical work on fairness / diversity in pipelines, and on understanding the feedback loops

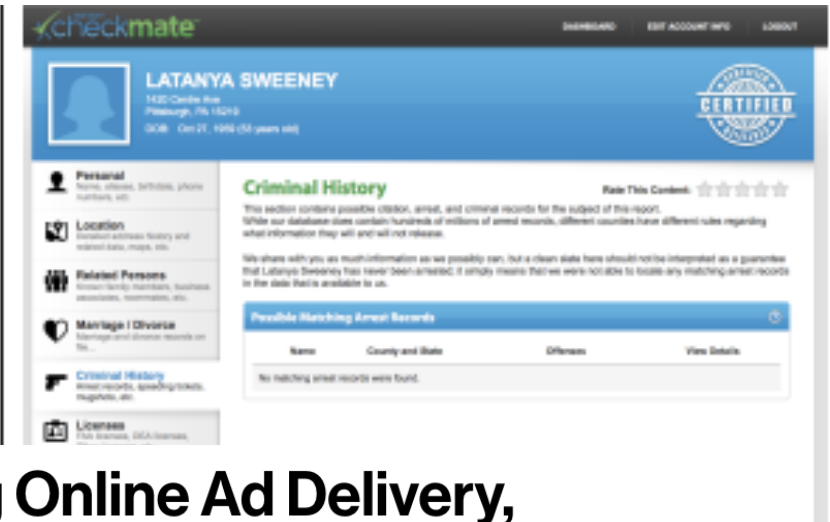
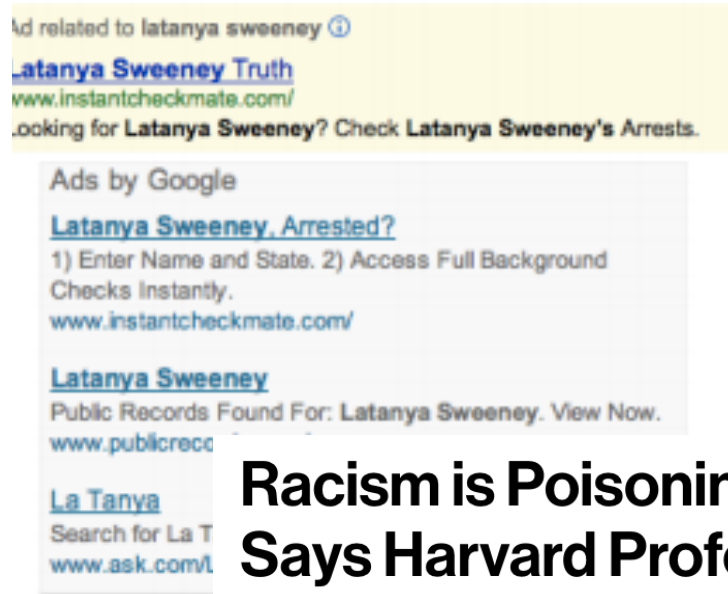
Roadmap

- ✓ Introduction
- Properties of responsible data analysis
 - ✓ Fairness
 - ✓ Diversity
 - ➔ Transparency
 - Neutrality
 - Data protection
- Common technical threads
- Societal issues
- Conclusion



Racially identifying names

[Latanya Sweeney; *CACM 2013*]



Racism is Poisoning Online Ad Delivery, Says Harvard Professor

Google searches involving black-sounding names are more likely to serve up ads suggestive of a criminal record than white-sounding names, says computer scientist

racially identifying names trigger ads suggestive of a criminal record

<https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/>

Transparency and accountability

- Users and regulators must be able to **understand** how raw data was selected, and what operations were performed during analysis
- Users want to **control** what is recorded about them and how that information is used
- Users must be able to **access** their own information and correct any errors (US Fair Credit Reporting Act)
- **Transparency** facilitates **accountability** - verifying that a service performs as it should, and that data is used according to contract



the problem is broad, we focus on a specific case

Online job ads

theguardian

Samuel Gibbs

Wednesday 8 July 2015 11.29 BST

Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs



One experiment showed that Google displayed adverts for a career coaching service for executive jobs 1,852 times to the male group and only 318 times to the female group. Photograph: Alamy

The AdFisher tool simulated job seekers that did not differ in browsing behavior, preferences or demographic characteristics, except in gender.

One experiment showed that Google displayed ads for a career coaching service for “\$200k+” executive jobs **1,852 times to the male group and only 318 times to the female group**. Another experiment, in July 2014, showed a similar trend but was not statistically significant.

<https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study>

Example: Ad targeting online

- **Users** browse the Web, consume content, consume ads (see / click / purchase)
- **Content providers** outsource advertising to third-party ad networks, e.g., Google's DoubleClick
- **Ad networks** track users across sites, to get a global view of users' behaviors
- **Google Ad Settings** aims to provide **transparency** / give **control to users** over the ads that they see

do users truly have transparency / choice or is this a placebo button?

Google Ads Settings

Your Google profile



Gender



Age

Ads based on your interests



Improve your ad experience when you are signed in to Google sites

With Ads based on your interests ON

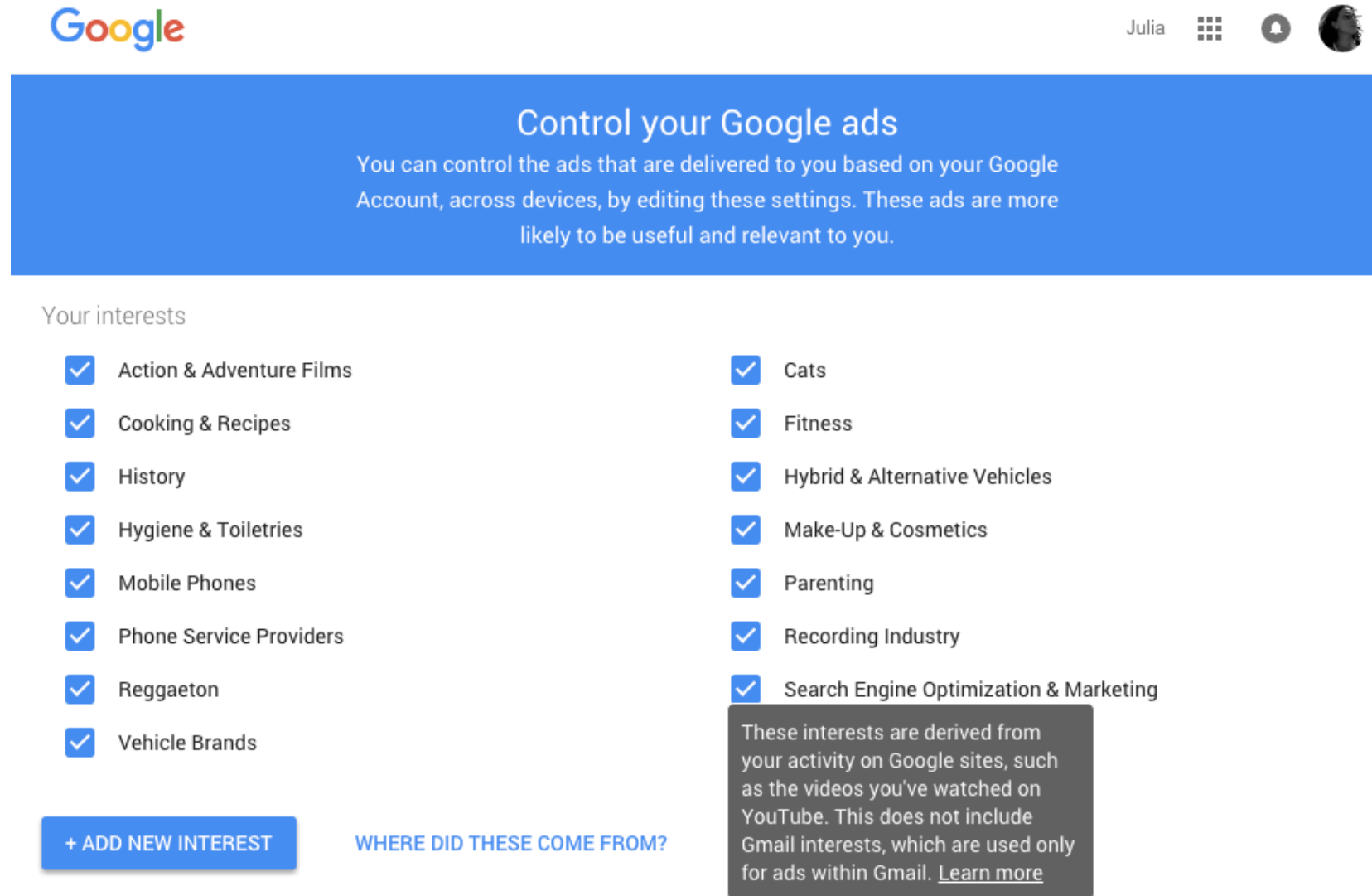
- The ads you see will be delivered based on your prior search queries, the videos you've watched on YouTube, as well as other information associated with your account, such as your age range or gender
- On some Google sites like YouTube, you will see ads related to your interests, which you can edit at any time by visiting this page
- You can block some ads that you don't want to see

With Ads based on your interests OFF

- You will still see ads and they may be based on your general location (such as city or state)
- Ads will not be based on data Google has associated with your Google Account, and so may be less relevant
- You will no longer be able to edit your interests
- All the advertising interests associated with your Google Account will be deleted

<http://www.google.com/settings/ads>

Google Ads Settings



The screenshot shows the Google Ads Settings page. At the top is the Google logo and the user's name 'Julia'. Below this is a blue header with the title 'Control your Google ads' and a sub-header explaining that users can control ads based on their Google Account settings. The main section is titled 'Your interests' and lists 15 interests, each with a blue checkmark icon. The interests are arranged in two columns. At the bottom left is a blue button labeled '+ ADD NEW INTEREST'. To its right is a link labeled 'WHERE DID THESE COME FROM?'. A grey tooltip box is positioned over the 'Search Engine Optimization & Marketing' interest, containing text about the source of these interests and a link to 'Learn more'.

Google

Julia

Control your Google ads

You can control the ads that are delivered to you based on your Google Account, across devices, by editing these settings. These ads are more likely to be useful and relevant to you.

Your interests

- ☒ Action & Adventure Films
- ☒ Cooking & Recipes
- ☒ History
- ☒ Hygiene & Toiletries
- ☒ Mobile Phones
- ☒ Phone Service Providers
- ☒ Reggaeton
- ☒ Vehicle Brands
- ☒ Cats
- ☒ Fitness
- ☒ Hybrid & Alternative Vehicles
- ☒ Make-Up & Cosmetics
- ☒ Parenting
- ☒ Recording Industry
- ☒ Search Engine Optimization & Marketing

+ ADD NEW INTEREST

[WHERE DID THESE COME FROM?](#)

These interests are derived from your activity on Google sites, such as the videos you've watched on YouTube. This does not include Gmail interests, which are used only for ads within Gmail. [Learn more](#)

<http://www.google.com/settings/ads>

AdFisher

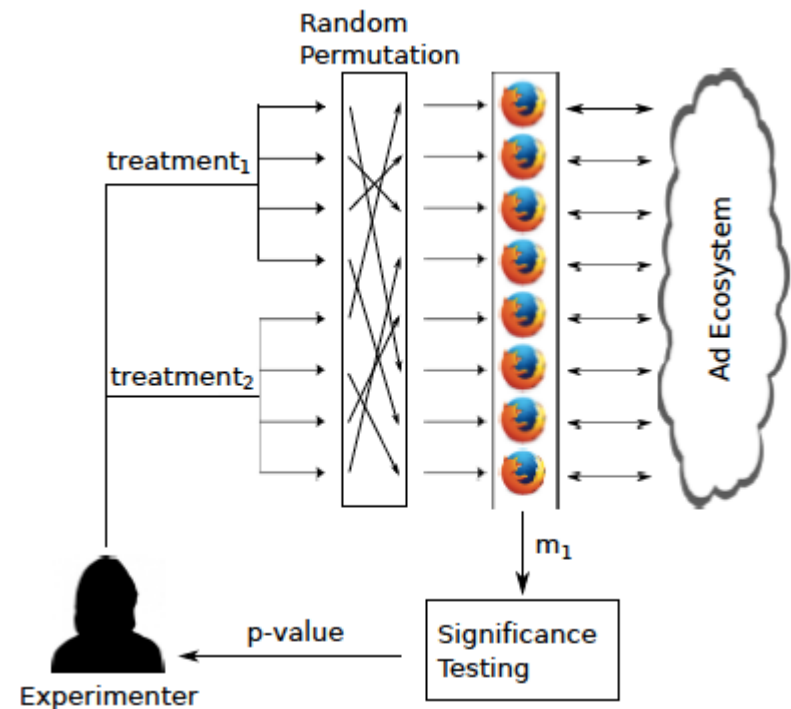
[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

From anecdotal evidence to statistical insight:

How do user behaviors, ads and ad settings interact?

Automated randomized controlled experiments for studying online tracking

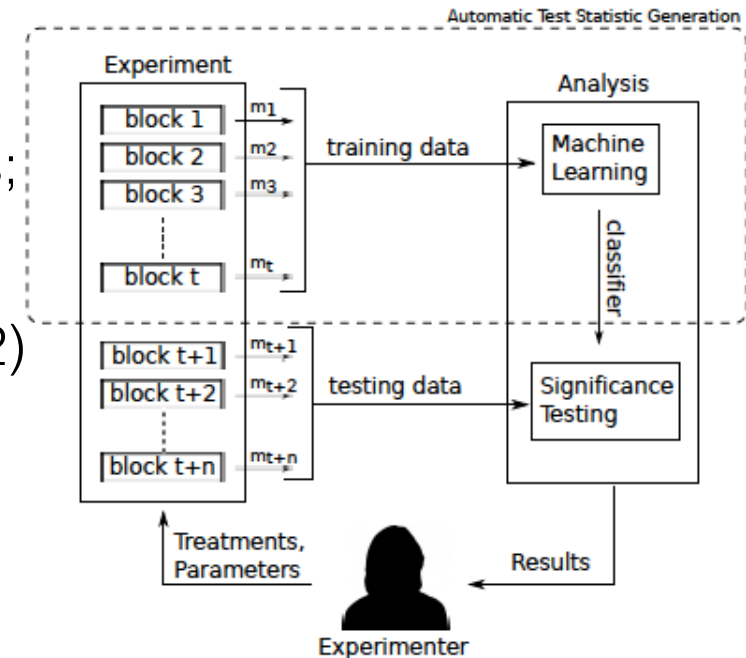
Individual data use transparency: ad network must share the information it uses about the user to select which ads to serve to him



AdFisher: methodology

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

- Browser-based experiments, simulated users
 - **input:** (1) visits to content providing websites; (2) interactions with Google Ad Settings
 - **output:** (1) ads shown to users by Google; (2) change in Google Ad Settings
- Fisher randomized hypothesis testing
 - **null hypothesis** inputs do not affect outputs
 - control and experimental treatments
 - AdFisher can help select a test statistic



AdFisher: gender and jobs

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

Non-discrimination: Users differing only in protected attributes are treated similarly

Causal test: Find that a protected attribute changes ads

Experiment 1: **gender and jobs**

Specify gender (male/female) in Ad Settings, simulate interest in jobs by visiting employment sites, collect ads from Times of India or the Guardian

Result: males were shown ads for higher-paying jobs significantly more often than females (1852 vs. 318)

violation

AdFisher: substance abuse

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

Transparency: User can view data about him used for ad selection

Causal test: Find attribute that changes ads but not settings

Experiment 2: **substance abuse**

Simulate interest in substance abuse in the experimental group but not in the control group, check for differences in Ad Settings, collect ads from Times of India

Result: no difference in Ad Settings between the groups, yet significant differences in what ads are served: rehab vs. stocks + driving jobs

violation

AdFisher: online dating

[A. Datta, M. Tschantz, A. Datta; *PETS 2015*]

Ad choice: Removing an interest decreases the number of ads related to that interest.

Causal test: Find that removing an interest causes a decrease in related ads

Experiment 3: **online dating**

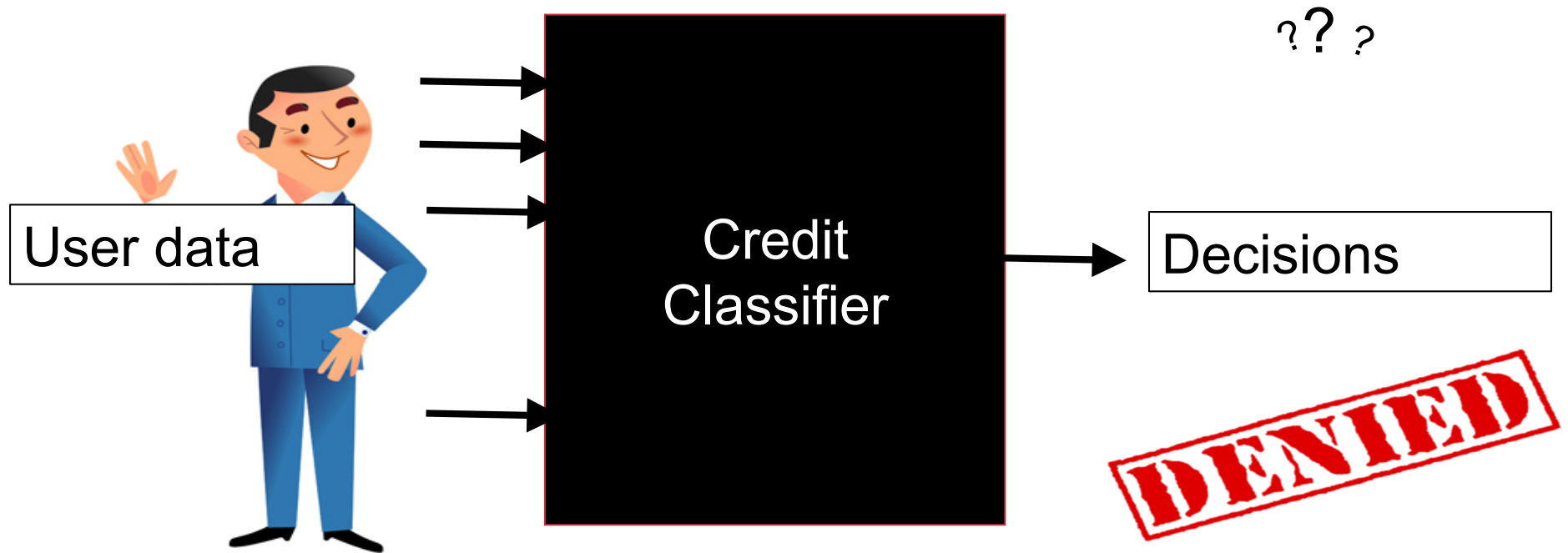
Simulate interest in online dating in both groups, remove “Dating & Personals” from the interests on Ad Settings for experimental group, collect ads

Result: members of experimental group do not get ads related to dating, while members of the control group do

compliance

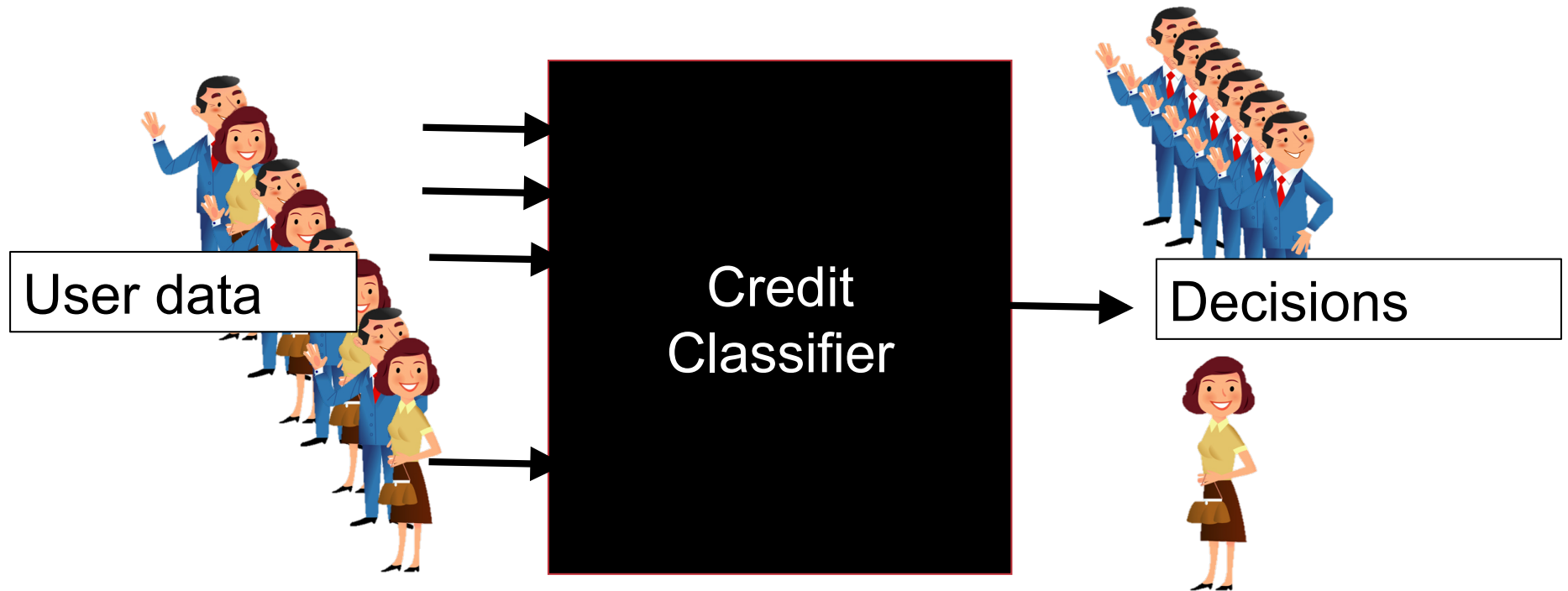
Algorithmic transparency

[A. Datta, S. Sen, Y. Zick; *SP 2016*]



Algorithmic transparency

[A. Datta, S. Sen, Y. Zick; *SP 2016*]



Quantifying influence of inputs on outcomes

[A. Datta, S. Sen, Y. Zick; *SP 2016*]

QII: quantitative input influence framework

Goal: determine how much influence an input, or a set of inputs, has on a **classification outcome** for an individual or a group

Transparency queries / quantities of interest

Individual: Which inputs have the most influence in my credit denial?

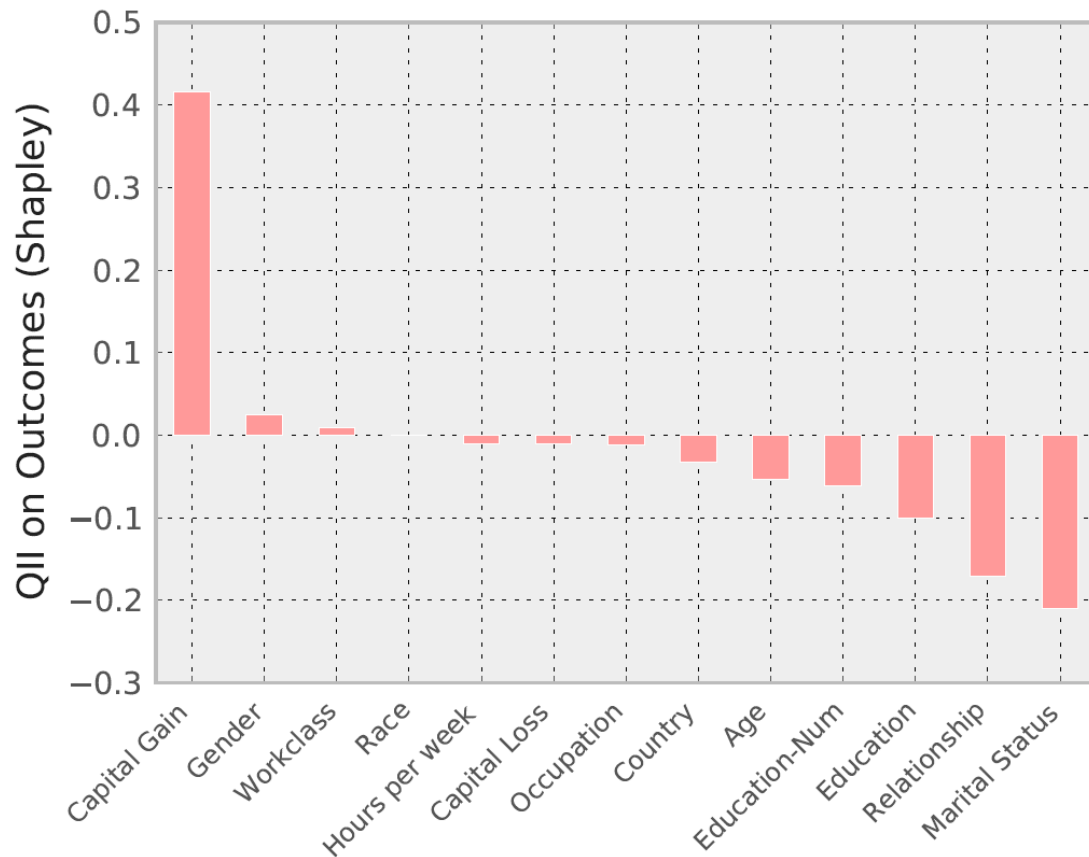
Group: Which inputs have most influence on credit decisions for women?

Disparity: Which inputs influence men getting more positive outcomes than women?

Uses **causal inference** to deal with correlated inputs. Useful as a building block to detect proxy discrimination (redundant encoding).

Explanation: Mr X

[A. Datta, S. Sen, Y. Zick; *SP 2016*]



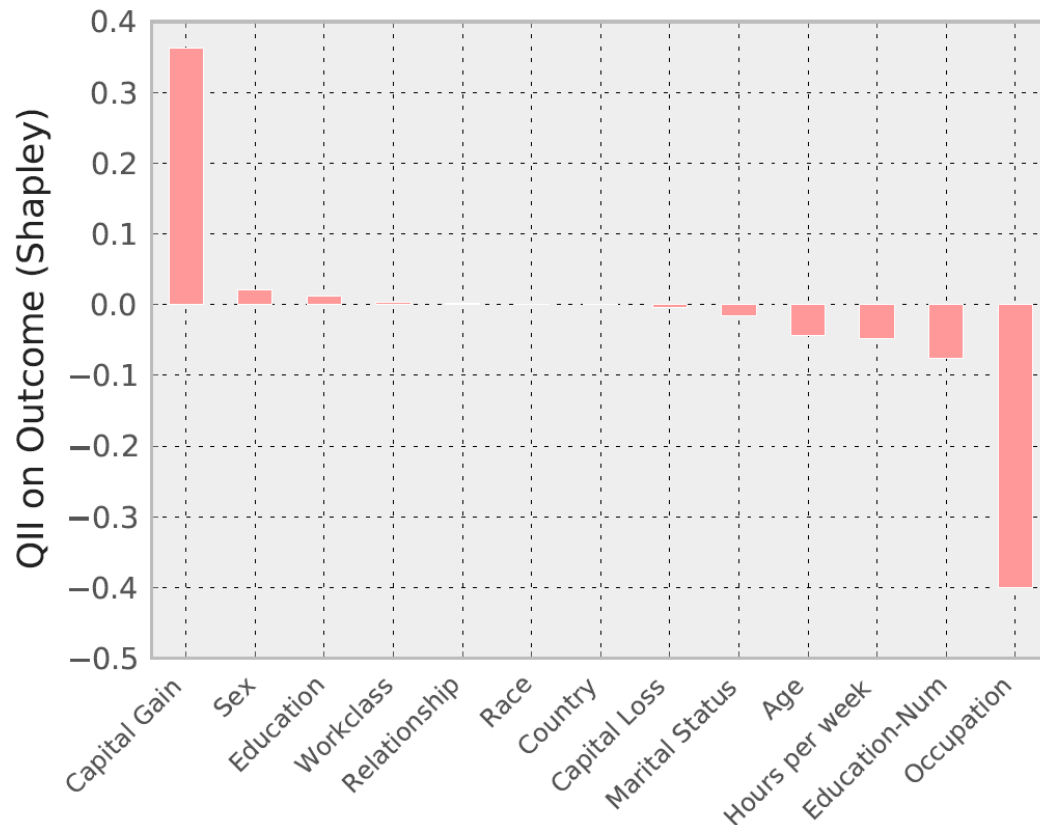
DENIED

Age	23
Workclass	Private
Education	11 th
Marital Status	Never married
Occupation	Craft repair
Relationship to household income	Child
Race	Asian-Pac Island
Gender	Male
Capital gain	\$14344
Capital loss	\$0
Work hours per week	40
Country	Vietnam

income

Explanation: Mr Y

[A. Datta, S. Sen, Y. Zick; *SP 2016*]



DENIED

Age	27
Workclass	Private
Education	Preschool
Marital Status	Married
Occupation	Farming-Fishing
Relationship to household income	Other Relative
Race	White
Gender	Male
Capital gain	\$41310
Capital loss	\$0
Work hours per week	24
Country	Mexico

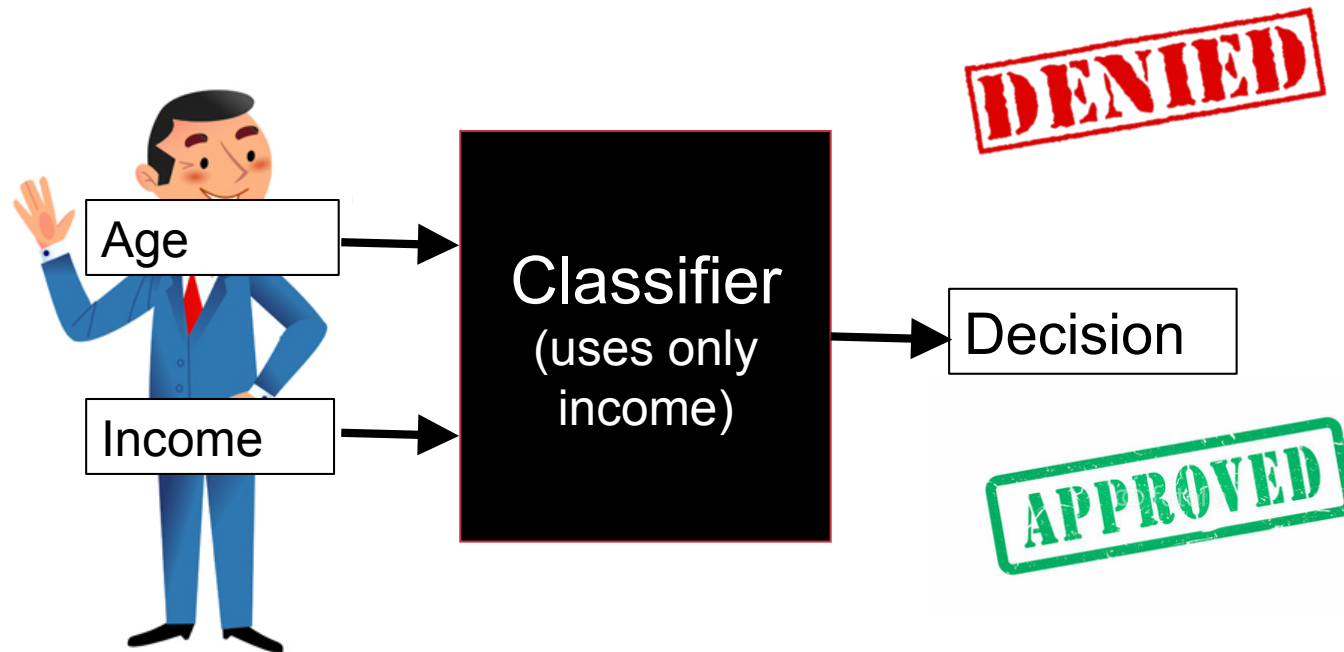
income

explanations for superficially similar individuals can be different

Unary QII

[A. Datta, S. Sen, Y. Zick; *SP 2016*]

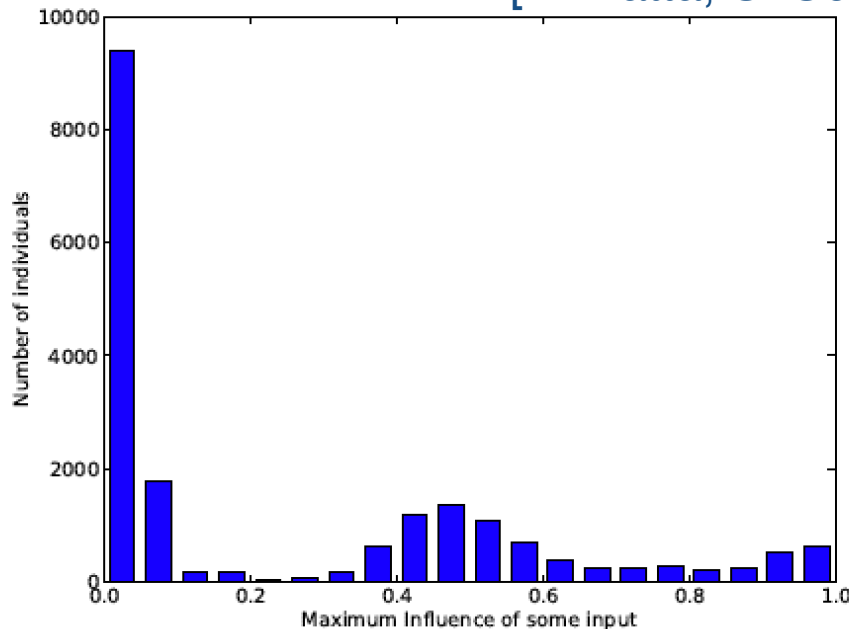
For a quantity of influence Q and an input feature i , the QII of i on Q is the difference in Q when i is changed via an **intervention**.



replace features with random values from the population, examine the distribution over outcomes

Set and Marginal QII

[A. Datta, S. Sen, Y. Zick; *SP 2016*]



A histogram of the highest specific causal influence for some feature across individuals in the UCI adult dataset. **Alone, most inputs have very low influence.**

Set QII measures the **joint influence** of a set of features S on the quantity of interest Q .

Marginal QII measures the **added influence** of feature i with respect to a set of features S on the quantity of interest Q . Use cooperative games (Shapley value) to aggregate marginal influence

Transparency in ranking

<https://freedom-to-tinker.com/2016/08/05/revealing-algorithmic-rankers/>

Input: database of items (individuals, colleges, cars, ...)

Score-based ranker: computes the score of each item using a **known** formula, e.g., monotone aggregation, then sorts items on score

Output: permutation of the items (complete or top-k)

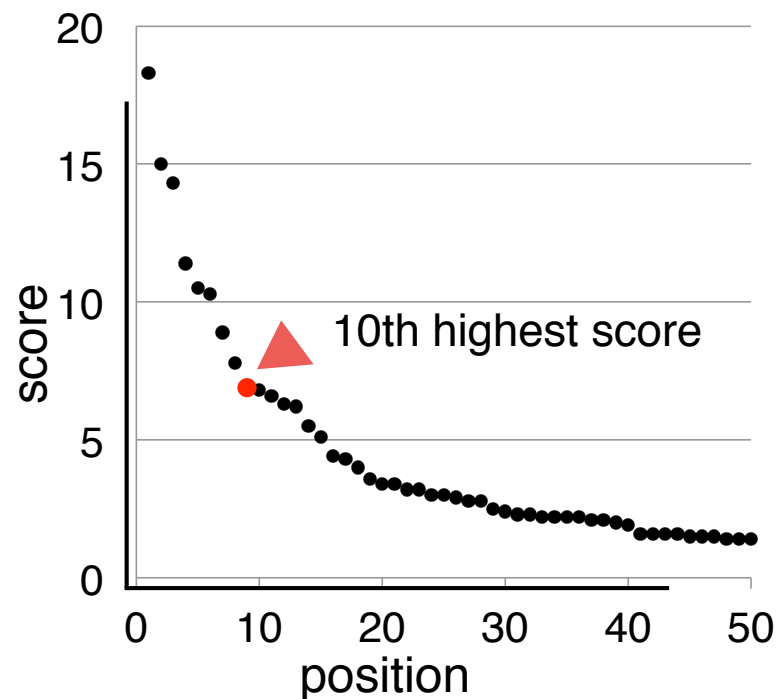
Do we have transparency?

We have syntactic transparency, but lack interpretability!

Opacity in algorithmic rankers

Reason 1: The scoring formula alone does not indicate the relative rank of an item.

Scores are absolute, rankings are relative. Is 5 a good score? What about 10? 15?



Opacity in algorithmic rankers

Reason 2: A ranking may be unstable if there are tied or nearly-tied items.

Rank	Institution	Average Count	Faculty
1	► Carnegie Mellon University	18.4	123
2	► Massachusetts Institute of Technology	15.6	64
3	► Stanford University	14.8	56
4	► University of California - Berkeley	11.5	50
5	► University of Illinois at Urbana-Champaign	10.6	56
6	► University of Washington	10.3	50
7	► Georgia Institute of Technology	8.9	81
8	► University of California - San Diego	8	51
9	► Cornell University	7	45
10	► University of Michigan	6.8	63
11	► University of Texas - Austin	6.6	43
12	► University of Massachusetts - Amherst	6.4	47



Opacity in algorithmic rankers

Reason 3: A ranking methodology may be unstable: small changes in weights can trigger significant re-shuffling.

THE NEW YORKER

DEPT. OF EDUCATION FEBRUARY 14 & 21, 2011 ISSUE

THE ORDER OF THINGS

What college rankings really tell us.



By Malcolm Gladwell

1. Chevrolet Corvette 205
2. Lotus Evora 195
3. Porsche Cayman 195

1. Lotus Evora 205
2. Porsche Cayman 198
3. Chevrolet Corvette 192

1. Porsche Cayman 193
2. Chevrolet Corvette 186
3. Lotus Evora 182

Opacity in algorithmic rankers

Reason 4: The weight of an attribute in the scoring formula does not determine its impact on the outcome.

Rank	Name	Avg Count	Faculty	Pubs	GRE
1	CMU	18.3	122	2	791
2	MIT	15	64	3	772
3	Stanford	14.3	55	5	800
4	UC Berkeley	11.4	50	3	789
5	UIUC	10.5	55	3	772
6	UW	10.3	50	2	796
		...			
39	U Chicago	2	28	2	779
40	UC Irvine	1.9	28	2	787
41	BU	1.6	15	2	783
41	U Colorado Boulder	1.6	32	1	761
41	UNC Chapel Hill	1.6	22	2	794
41	Dartmouth	1.6	18	2	794

Given a score function:

$0.2 * faculty +$

$0.3 * avg\ cnt +$

$0.5 * gre$

Rankings are not benign!

THE NEW YORKER

DEPT. OF EDUCATION FEBRUARY 14 & 21, 2011 ISSUE

THE ORDER OF THINGS

What college rankings really tell us.



By Malcolm Gladwell

Rankings are not benign. They enshrine very particular ideologies, and, at a time when American higher education is facing a crisis of accessibility and affordability, we have adopted a de-facto standard of college quality that is uninterested in both of those factors. And why? Because a group of magazine analysts in an office building in Washington, D.C., decided twenty years ago to value selectivity over efficacy, to use proxies that scarcely relate to what they're meant to be proxies for, and to pretend that they can compare a large, diverse, low-cost land-grant university in rural Pennsylvania with a small, expensive, private Jewish university on two campuses in Manhattan.



Harms of opacity

1. **Due process / fairness.** The subjects of the ranking cannot have confidence that their ranking is meaningful or correct, or that they have been treated like similarly situated subjects - *procedural regularity*

2. **Hidden normative commitments.** What factors does the vendor encode in the scoring ranking process (syntactically)? What are the *actual* effects of the scoring / ranking process? Is it stable? How was it validated?

Harms of opacity

3. Interpretability. Especially where ranking algorithms are performing a public function, **political legitimacy** requires that the public be able to interpret algorithmic outcomes in a meaningful way. Avoid *algocracy*: the rule by incontestable algorithms.

4. Meta-methodological assessment. Is *a* ranking / *this* ranking appropriate here? Can we use a process if it cannot be explained? Probably yes, for recommending movies; probably not for college admissions.

Ranking Facts

Recipe

sort by decreasing income
income correlates with **age**

Ingredients

top-10

40 38 35
median
age
parity fails!

150 125 90
median
income

white 70%
black 10%
asian 20%
race
parity fails!

overall

40 32 18
median
age

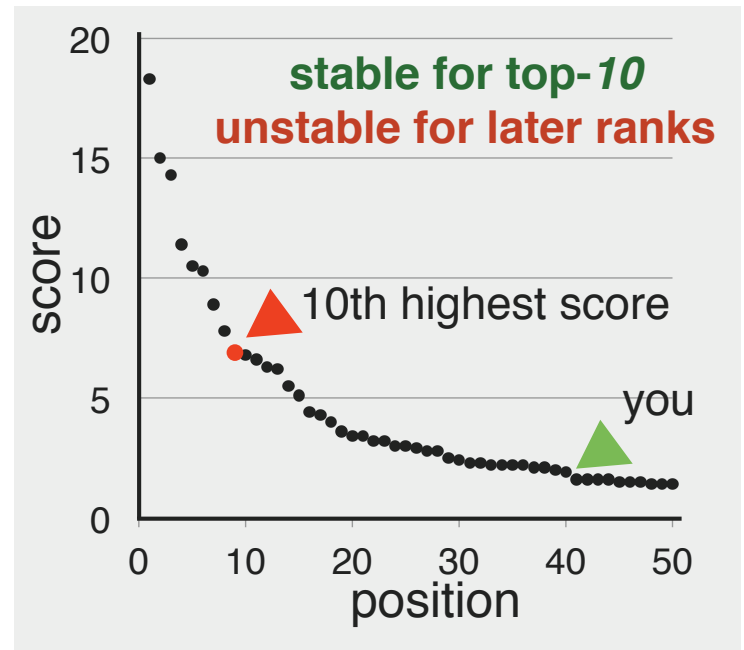
150 50 25
median
income

white 50%
black 40%
asian 10%
race

Your Outcome

rank 45
increase income by 20K
to move to top-10

Stability



Lots of other interesting work

- **Privacy**: awareness of privacy leaks, usability of tools
- **Tracking**: awareness of tracking, reverse-engineering
- Pricing transparency, e.g., Uber surge pricing [L. Chen, A. Mislove, C. Wilson; *IMC 2015*]
- Data Transparency Lab: technology + policy, see datatransparencylab.org for pointers

Is this down to privacy?

A shift from privacy and consent to responsible use!
[E. Kenneally; *SIGCAS 2015*]

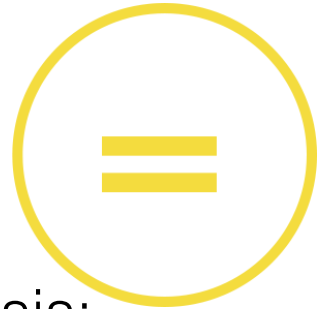
DATA
TRANSPARENCY
LAB

coffee break



data *RESPONSIBLY*

Technology is not the whole answer



Technology is needed to enable responsible data analysis:
specify and verify

- But will companies simply feel compelled to act responsibly?
- Who sets the standards for what is ethical and legal?

Users and regulators!

Power comes with responsibility

power

A handful of big players command most of the world's computational resources and most of the data, including all of your personal data - an **oligopoly**

danger



can destroy business competition

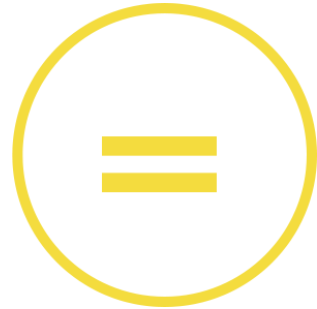
control what information you receive

can guide your decisions

can infringe on your privacy and freedom

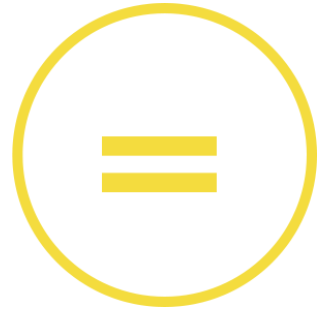
the rich get richer, the poor get poorer

User organization



- Users are data, users are consumers of data, users have **tremendous power!**
- Example: Instagram 2012, gave FB (new owner) broad access to user data and photos for commercial use. Forced to change back under pressure from users.
- Limitations: user education, lack of proper tools

Public policy



- Should the government regulate the Big Data industry?
 - regulate
 - define good practices
 - evaluate responsibility
- Issues:
 - which government?
 - lack of competence, agility

US legal mechanisms

[Big Data: A tool for inclusion or exclusion? FTC Report; 2016]

<https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>

- Fair Credit Reporting Act - applies to consumer reporting agencies, must ensure correctness, access and ability to correct information
- Equal opportunity laws - prohibit discrimination based on race, color, religion, ... - plaintiff must show disparate treatment / disparate impact
- FTC Act - prohibits unfair or deceptive acts or practices to companies engaged in data analytics

lots of gray areas, much work remains, enforcement is problematic since few auditing tools exist

EU legal mechanisms

- **Data protection**

- Different countries are developing specific laws, e.g., portability against user lock-in (France)

- **Transparency**

- Open data policy: legislation on re-use of public sector information
- Open access to research publications and data

- **Neutrality**

- Net neutrality: a new law, but with some limitations
- Platform neutrality: the first case against Google search

Japan legal mechanisms

個人情報保護委員会

Personal Information Protection Commission

法人番号：4000012010025

[> 本文へ](#) [> サイトマップ](#)

文字サイズ変更 標準 大きめ

[ホーム](#)

[委員会の概要](#)

[個人情報保護法について](#)

[マイナンバーについて](#)

[委員会の活動](#)



Personal Information Protection Commission

The Personal Information Protection Commission (PPC) was established on January 1, 2016, changed from the Specific Personal Information Protection Commission.



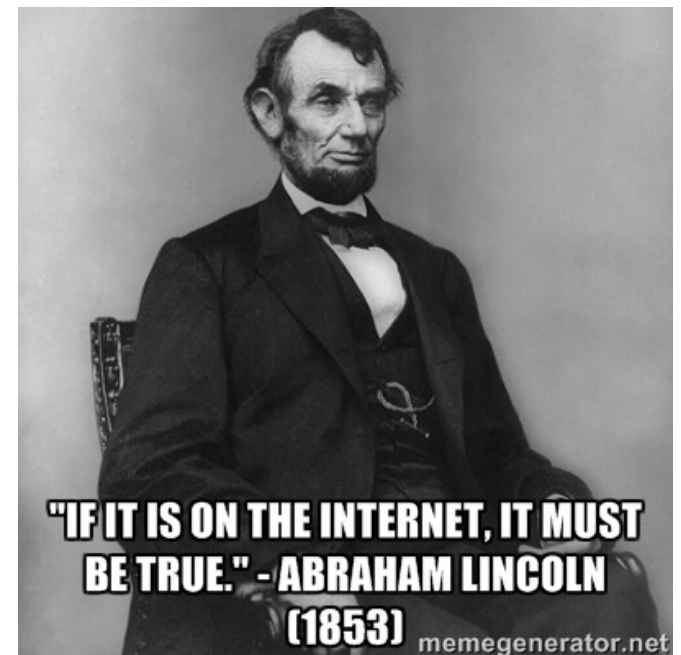
Article 17 (2) A business operator handling personal information must not acquire sensitive personal information without **in advance obtaining the person's consent** to do so, except in the following cases;

Article 2 (3) The term “**sensitive personal information**” ... require special consideration in handling so as to **avoid any unfair discrimination**, prejudice or other disadvantage to an individual based on person's race, creed, social status, medical history, criminal records or the fact that a person has incurred damages through an offense, etc.

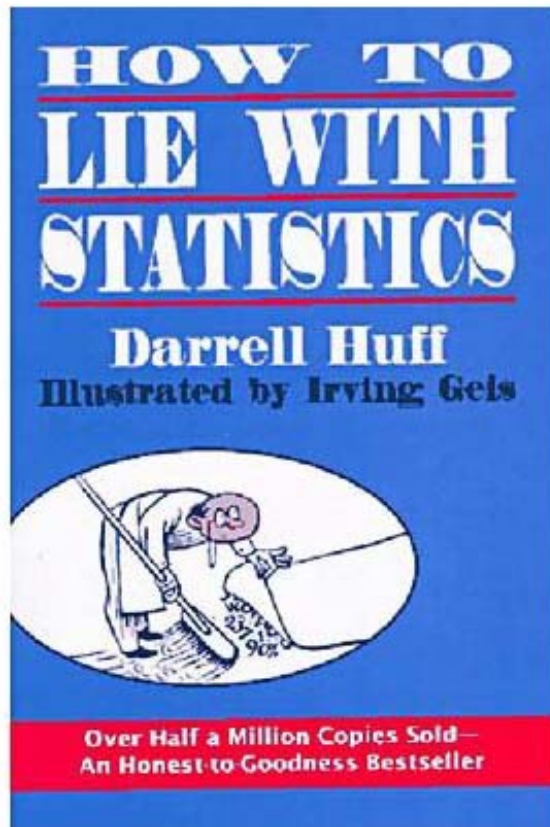
Education

- Concepts
 - **understanding** data acquisition methods and data analysis processes
 - **verifying** the data and the process: provenance, credit attribution, trust
 - **interpreting** results
- Tools: computer science, probability and statistics, what people need to know about **data science**!

learn and teach to question!



Data literacy



statistics



BIG DATA



Statistics scares people, Big Data REALLY scares people!



Report from Dagstuhl Seminar 16291

Data, Responsibly

Edited by

Serge Abiteboul¹, Gerome Miklau², Julia Stoyanovich³, and
Gerhard Weikum⁴

- 1 ENS – Cachan, FR, serge.abiteboul@inria.fr
- 2 University of Massachusetts – Amherst, US, miklau@cs.umass.edu
- 3 Drexel University – Philadelphia, US, stoyanovich@drexel.edu
- 4 MPI für Informatik – Saarbrücken, DE, weikum@mpi-inf.mpg.de

The goals of the seminar were to assess the state of data analysis in terms of fairness, transparency and diversity, identify new research challenges, and derive an agenda for computer science research and education efforts in responsible data analysis and use.

An important goal of the seminar was to **identify opportunities for high-impact contributions to this important emergent area specifically from the data management community.**

http://drops.dagstuhl.de/opus/volltexte/2016/6764/pdf/dagrep_v006_i007_p042_s16291.pdf

Research Directions for Principles of Data Management (Dagstuhl Perspectives Workshop 16151)

Edited by

Serge Abiteboul, Marcelo Arenas, Pablo Barceló, Meghyn Bienvenu, Diego Calvanese, Claire David, Richard Hull, Eyke Hüllermeier, Benny Kimelfeld, Leonid Libkin, Wim Martens, Tova Milo, Filip Murlak, Frank Neven, Magdalena Ortiz, Thomas Schwentick, Julia Stoyanovich, Jianwen Su, Dan Suciu, Victor Vianu, and Ke Yi

1 Introduction

In April 2016, a community of researchers working in the area of Principles of Data Management (PDM) joined in a workshop at the Dagstuhl Castle in Germany. The workshop was organized jointly by the Executive Committee of the ACM Symposium on Principles of Database Systems (PODS) and the Council of the International Conference on Database Theory (ICDT). The mission of this workshop was to identify and explore some of the most important research directions that have high relevance to society and to Computer Science today, and where the PDM community has the potential to make significant contributions. This report describes the family of research directions that the workshop focused on from three perspectives: potential practical relevance, results already obtained, and research questions that appear surmountable in the short and medium term. This report organizes the identified research challenges for PDM around seven core themes, namely *Managing Data at Scale*, *Multi-model Data*, *Uncertain Information*, *Knowledge-enriched Data*, *Data Management and Machine Learning*, *Process and Data*, and *Ethics and Data Management*. Since new challenges in PDM arise all the time, we note that this list of themes is not intended to be exclusive.

<https://arxiv.org/pdf/1701.09007.pdf>



Serge Abiteboul and
Julia Stoyanovich

NOVEMBER 20, 2015

DATA, RESPONSIBLY

≡ Big Data

(This blog post is an extended version of an October 12, 2015 Le Monde op-ed article (in French))

Our society is increasingly relying on algorithms in all aspects of its operation. We trust algorithms not only *to help carry out routine tasks*, such as accounting and automatic manufacturing, but also *to make decisions on our behalf*. The sorts of decisions with which we now casually entrust algorithms range from unsettling (killer drones), to tedious (automatic trading), or deeply personal (online dating). Computer technology has tremendous power, and with that power comes immense responsibility. Nowhere is the need to control the power and to judiciously use technology more apparent than in massive data analysis, known as big data.

Big data technology holds incredible promise of improving people's lives, accelerating scientific discovery and innovation, and bringing about positive societal change. The goal of big data analysis is to efficiently sift through oceans of data, identifying valuable knowledge. The more data is available, the more knowledge can be derived. This gives a strong incentive for data acquisition, as well as for data sharing. Data sharing may be fully unrestricted, as is the case with the Open Data movement, or more controlled, as is the case with medical data (for privacy) and scientific or commercial data

wp.sigmod.org/?p=1900



Thank you!



data *RESPONSIBLY*

ProPublica: details and an exercise

- Details of the ProPublica criminal sentencing investigation
- A hands-on exercise

Automated risk assessment

Goal: is to predict the likelihood of some category of future crime.

Used to: assign bail amounts, report given to judges for sentencing

Input: attributes of an individual, drawn from 137 questions answered by defendants or derived from records.

Output: risk scores [1,10]: general recidivism, violent recidivism, risk of failure to appear

The analysis

COMPAS is sold by a company called Northpointe, is among the most widely-used in the U.S.

Through a public records request, ProPublica obtained two years worth of COMPAS scores from the Broward County Sheriff's Office in Florida. They received data for all 18,610 people who were scored in 2013 and 2014.

Each pretrial defendant received at least three COMPAS scores: "Risk of Recidivism," "Risk of Violence" and "Risk of Failure to Appear."

ProPublica built a profile of each person's criminal history, both before and after they were scored.

Does COMPAS work?

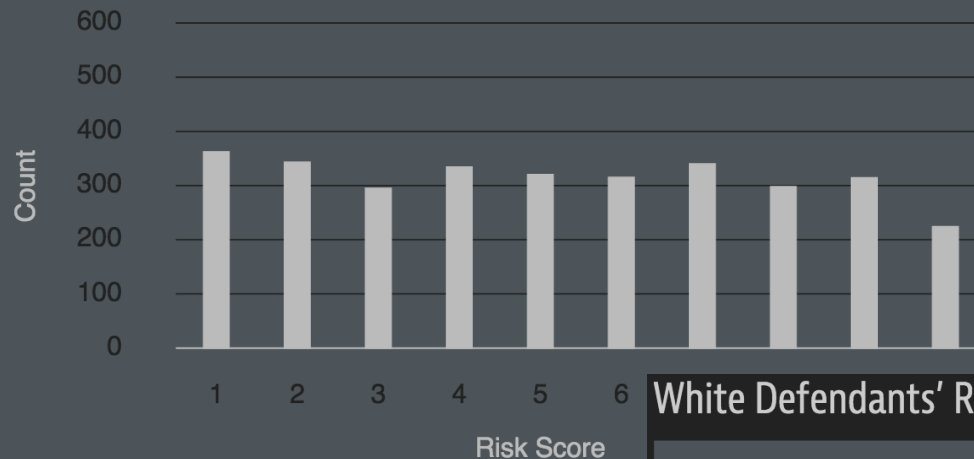
Per ProPublica *“somewhat more accurate than a coin flip”* (?)

Only 20 percent of people predicted to commit violent crimes actually went on to do so. When all crimes are considered (including misdemeanors): Of those deemed likely to re-offend, 61 percent were arrested for any subsequent crimes within two years.

“Northpointe does not agree that the results of our analysis, or the claims being made based upon that analysis, are correct or that they accurately reflect the outcome from the application of the model.”

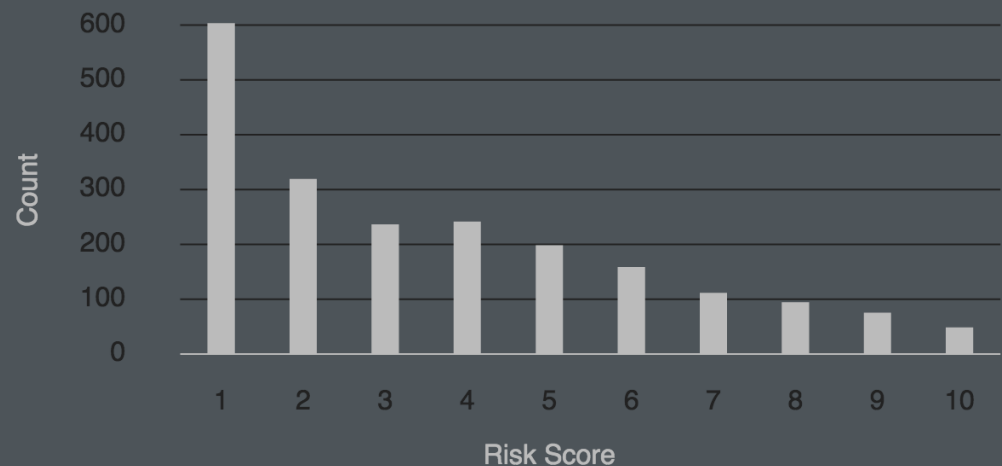
Risk scores for two races

Black Defendants' Risk Scores



“These charts show that scores for white defendants were skewed toward lower-risk categories. Scores for black defendants were not.”

White Defendants' Risk Scores



Racial bias

Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

Classification association rules (CARs)

D: database of individuals

UID	sex	age	score
Ann	F	31	low
Bob	M	27	high
Cate	F	55	high
Dave	M	43	low

TD: database of individuals that looks like transactions

UID	attributes
Ann	gender=F, age \in [30,35), score=low
Bob	gender=M, age \in [25,30), score=high
Cate	gender=F, age \in [55, 60), score=high
Dave	gender=M, age \in [40, 45), score=low

$S \ X \rightarrow C$ X is a set of attribute-value pairs, and $c \in C$ is a (binary) outcome

in our example, *score* is the outcome (low or high), also called the class label

continuous attribute values must be discretized (mapped to buckets)
as part of the transformation - age in our example

Potentially discriminatory rules (PD-CARs)

D: database of individuals

UID	gender (S)	age (X1)	edu (X2)	score (C)
Ann	F	[30,35)	BS	low
Bob	M	[25,30)	MS	high
Cate	F	[55, 60)	PhD	high
Dave	M	[40, 45)	BS	low

$S X \rightarrow C$ S is a binary attribute-value assignment -
membership in a protected group (gender in our example)

X is a set of “regular” attribute-value pairs
(age and edu in our example)

C is a binary attribute-value assignment -
classification outcome (score in our example)

Potentially discriminatory rules (PD-CARs)

R: $S X \rightarrow C$

UID	gender (S)	age (X1)	edu (X2)	score (C)
Ann	F	[30,35)	BS	low

S binary membership in a protected group (gender)

X “regular” attribute-value pairs (age and edu)

C binary classification outcome (score)

support $(S X \rightarrow C) = \% D$ that assigns the same attribute values for S , X and C

confidence $(S X \rightarrow C) = \text{support}(S X \rightarrow C) / \text{support}(S X)$

α -protection $(S X \rightarrow C) = \text{confidence}(S X \rightarrow C) / \text{confidence}(X \rightarrow C)$

PD-CARs for the ProPublica dataset

Get the data (PP.csv) and the loading script for PostgreSQL (Load_PP.sql) from <https://drive.google.com/open?id=0Bz83LL5KZNx3NXZIMUtZWDBPV00>

The post-processed Pro Publica dataset consist of one table: Pro_Publica (uid, attr, val). Check that the table has 29375 rows.

Compute PD-CARs with vdecile as the target attribute, race as the protected attribute, and age, gender and marriage as regular attributes.

A stub for the solution (Compute.sql) is also available at <https://drive.google.com/open?id=0Bz83LL5KZNx3NXZIMUtZWDBPV00>.

For a more exhaustive description see https://www.cs.drexel.edu/~julia/cs500/documents/assignments/CS500_HW6_Winter2017.pdf