Differences that make a difference

Statistical profiling and fairness to individuals

Wilfried Hinsch, Köln

07.12.20

Algorithms of Oppression

”[...] I am bringing to light various forms of technological redlining that are on the rise. The near-ubiquitous use of algorithmically driven software, both visible and invisible […] demands a closer inspection […] We are only beginning to understand the long-term consequences of these decision-making tools in both masking and deepening social inequality. […] mathematical formulations to drive automated decisions are made by human beings. While we often think of terms such as “big data” and “algorithms” as being benign, neutral, or objective, they are anything but. The people who make these decisions hold all types of values, many of which openly promote racism, sexism, and false notions of meritocracy [...].”

Safiya Umoja Noble, 2018

Statistical discrimination

- Statistical *representations* of injustice.
- Statistical *evidence* for unfair discrimination.
- Statistical reasoning as a *mechanism* of discrimination.

A profile is …

- a set of manifest personal characteristics which in conjunction support a prediction that a person who fits the profile also has another characteristic which is non-tangible or not yet manifest.
- A *statistically reliable profile* is a profile that supports the prediction of a non-tangible or not yet manifest personal characteristic by sound inductive-statistical reasoning.

… in a word …

profiles are conditional probabilities

$$p\ (f \mid F_{1,}\ F_{2,}\ \ldots\ F_k)$$

In the absence of perfect knowledge ...

- all rational actions and prudential ways of conduct hinge upon profiles and probability estimates about objects, persons, and events.
- all moral evaluation is based on probabilistic conclusions about not directly observable characteristics (dispositions and future actions of agents, beliefs, intentions) from sets of tangible characteristics.

Unfair discrimination is ...

- an objectionable deviation from a norm of equal treatment which imposes a burden or disadvantage on a person *and* which, from a moral point of view, cannot be properly justified with reference to characteristics of the impaired person.

The idea of equality

- *Equality for equals, inequality for unequals!*
- The notion of *relevant* similarities/differences
- Equal treatment vs. treatment as equals

Common problems of statistical reasoning

- Insufficiently specific profiles

  → Inclusion of cases that share the relevant profile but actually do not have the feature which the profile is meant to predict.

  → Exclusion of of cases which do not share the profile but do have the feature which the profile is meant to predict.

- Ignorance of base-rates

  → Distorted probability estimates for small groups with many members who fit a certain profile

Carnap’s requirement of total evidence

$p_1\ (f \mid F_{1\ \&}\ F_2) \qquad = 0.9$

$p_2\ (\text{not-}f \mid G) \qquad = 0.9$

$p^*\ (f \mid F_1\ \&\ F_2\ \&\ G) \qquad = ???$

Costs & benefits of total evidence

$$p_1\ (f \mid F_{1\ \&}\ F_2) \qquad = 0.9$$

$$p_2\ (\text{not-}f \mid G_1?,\ G_1?,\ G_1?,\ \ldots\ ) \qquad = ???$$

$$p^*\ (f \mid\ F_1\ \&\ F_2\ \ \&\ G_1\ \ldots\ G_k) \qquad = ???$$

- Investigation costs
- Loss of coordination
- Prevention of personal impairment,

Fairness to individuals

Can it ever be fair to discriminate between individuals on the basis of a statistically reliable profile, when

- the profile is known to be over-inclusive in a particular case?
- the profile *may* not correctly predict in a particular *case*?

Misgivings about statistical discrimination

- Lack of concern for the individual.
- Discrimination for inappropriate reasons.

Don’t judge me by my group!

- Statistical profiling considers individuals not *as individuals* but only as members of groups of people with whom they share a certain profile.
- it is wrong because it judges people on the basis of probability estimates that relate to features which they may actually not have.

Don’t judge me by my colour!

- If the manifest personal characteristics that function as *predictors* or *proxies* in statistical profiling do not relate to *causes* or *symptoms* of the relevant non-tangible feature, they must be deemed irrelevant from a moral point of view.

➢ Therefore, they cannot justify statistical discrimination.

Evasive ”relevance”

- Varieties of context & purpose
- Preferential hiring

→ Female medics and lawyers

→ Sexy personal assistants

What’s wrong, then, ...

... with statistical profiling & discrimination?

Nothing in principle, but much in practice.

So where do we go?

- No guidance from the Equal Treatment Principle.
- No way to rely on a substantive & and generally valid understanding of “relevant” personal characteristics.
- Two levels of moral argument
  - The evaluation & critique of social practices on the basis of substantive anti-discrimination norms.
  - The specification & justification of anti-discrimination norms.

➢ A pull towards a value-based and consequentialist understanding of distributive justice as the normative basis for a critique of unfair (statistical) discrimination.

- Implications for “Responsible AI”