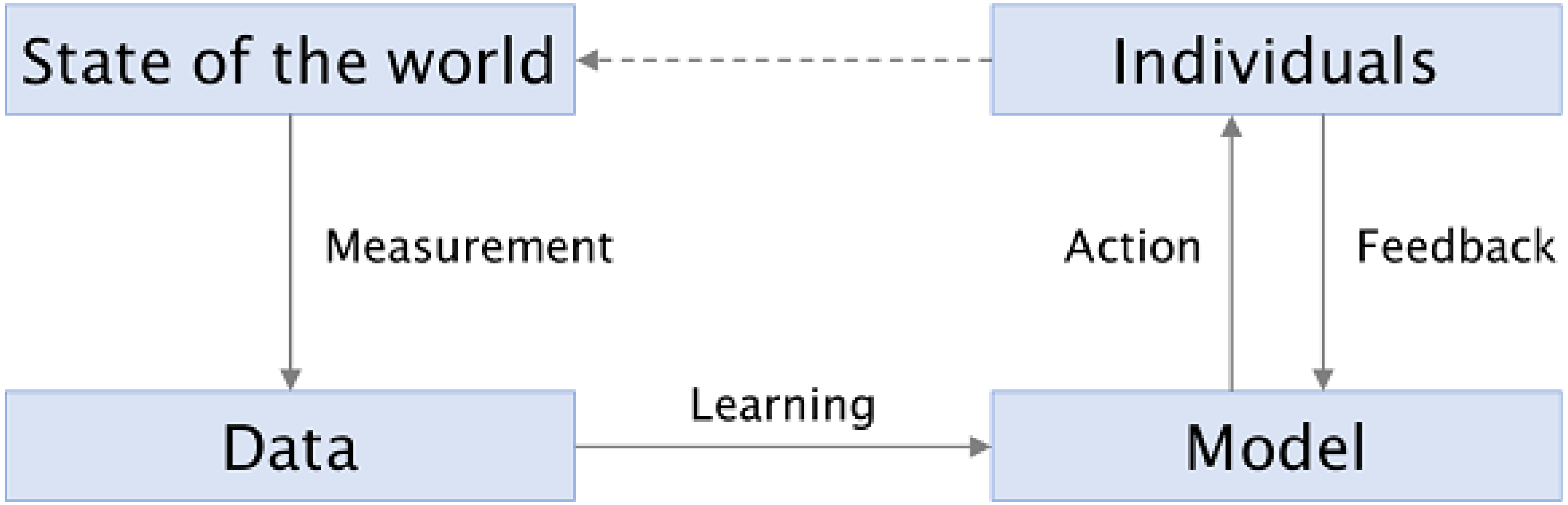
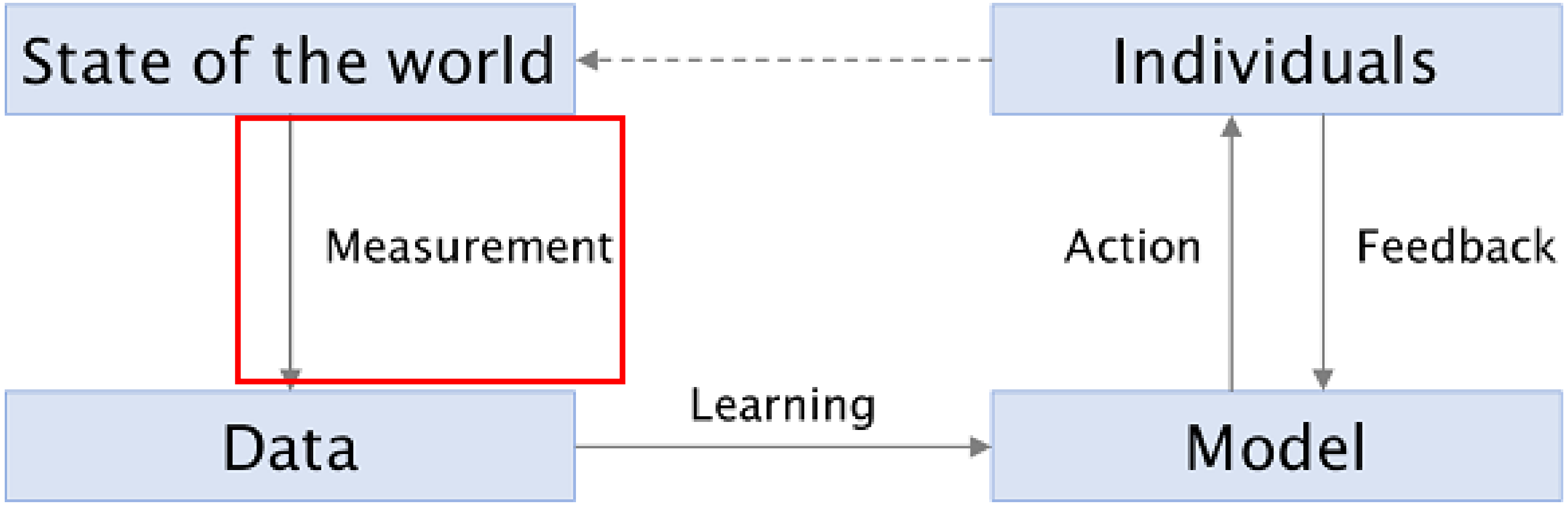
What assumptions do we make when using algorithms for decision-making?

The machine learning loop



The machine learning loop

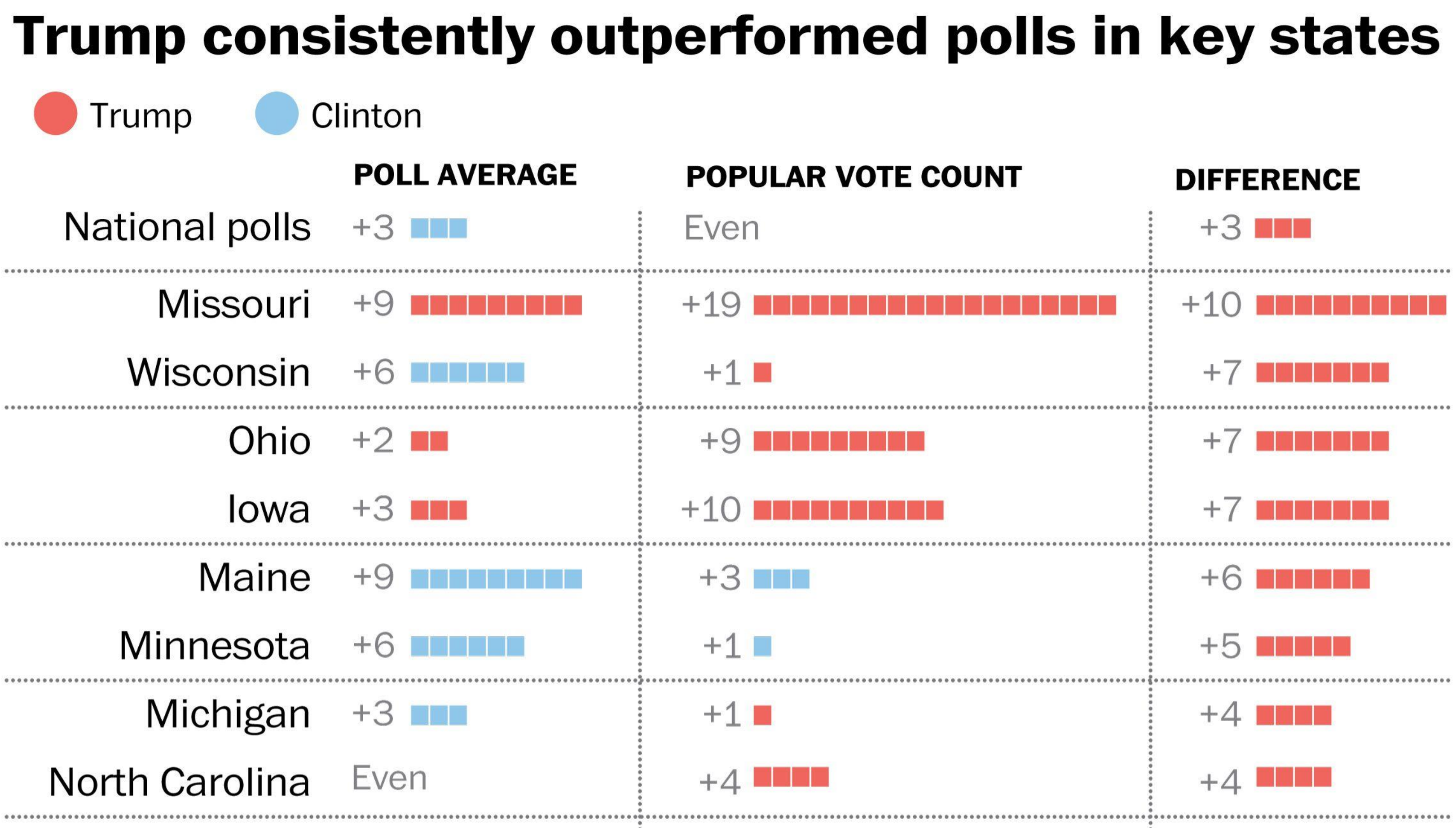


The machine learning loop: measurement

Issue: measurement is often subjective

The machine learning loop: measurement

Issue: measurement is often subjective. Bias in data collection itself:

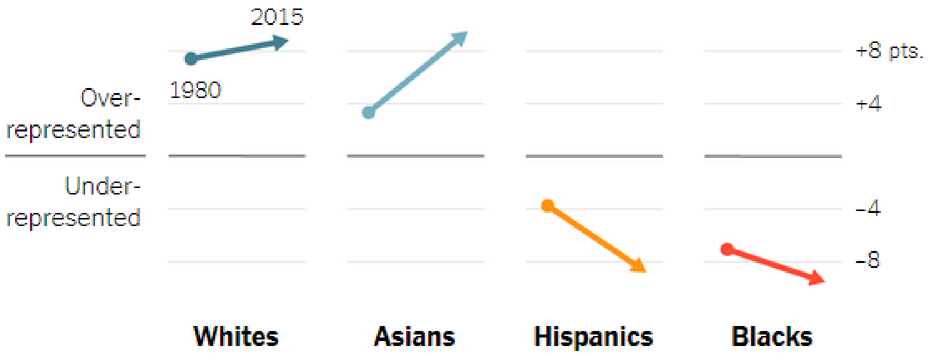


The machine learning loop: measurement

Issue: measurement is often subjective. Data is not stable over time.

Even With Affirmative Action, Blacks and Hispanics Are More Underrepresented at Top Colleges Than 35 Years Ago

By JEREMY ASHKENAS, HAEYOUN PARK and ADAM PEARCE AUG. 24, 2017



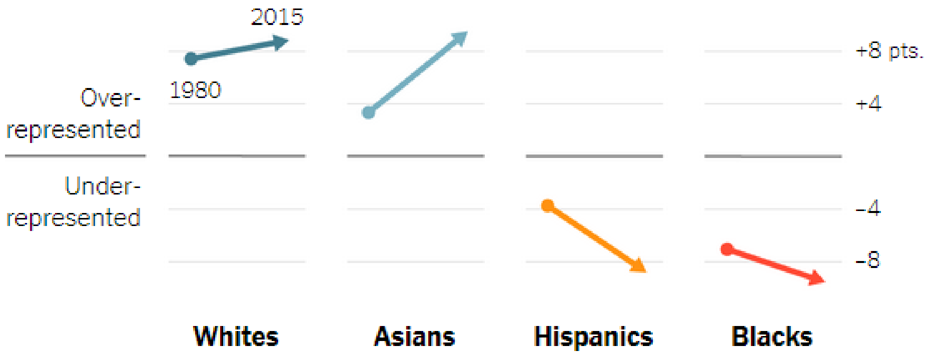
How much more or less each group is represented among freshmen at top colleges relative to the U.S. population

The machine learning loop: measurement

Issue: measurement is often subjective. Data is not stable over time.

Even With Affirmative Action, Blacks and Hispanics Are More Underrepresented at Top Colleges Than 35 Years Ago

By JEREMY ASHKENAS, HAEYOUN PARK and ADAM PEARCE AUG. 24, 2017



How much more or less each group is represented among freshmen at top colleges relative to the U.S. population

But:

- Definitions of race in the U.S. have been subject to change
- U.S census has had varying race categories over time
- This data: from 1980 to 2015, but multiracial category introduced in 2008

The machine learning loop: measurement

Issue: measurement is often subjective. What features to collect?

The machine learning loop: measurement

Issue: measurement is often subjective. What features to collect?

University of California Drops SAT Scores for Admission

The University of California won't consider SAT and ACT scores that are submitted with admission and scholarship applications under a settlement of a student lawsuit.

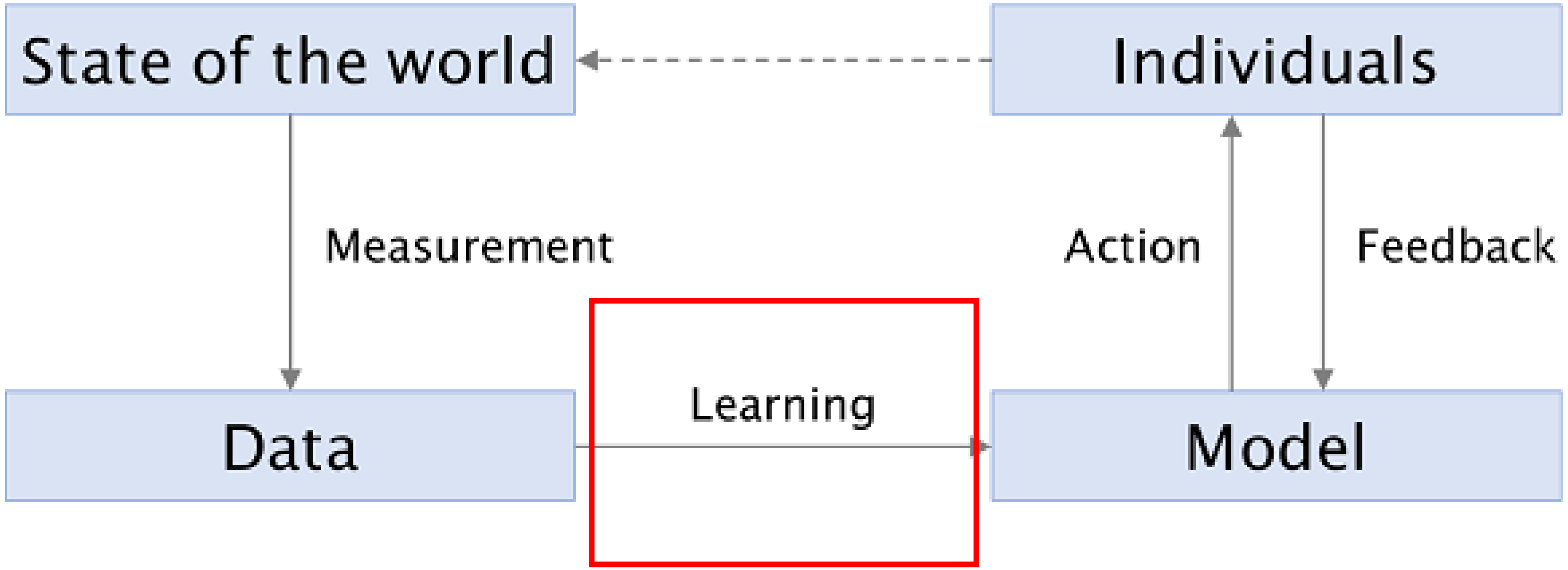
The machine learning loop: measurement

Issue: measurement is often subjective. What’s the right target variable?

The machine learning loop: measurement

Issue: measurement is often subjective. What's the right target variable?

The machine learning loop



The machine learning loop: learning

Issue #1: mimicking bias in data

The machine learning loop: learning

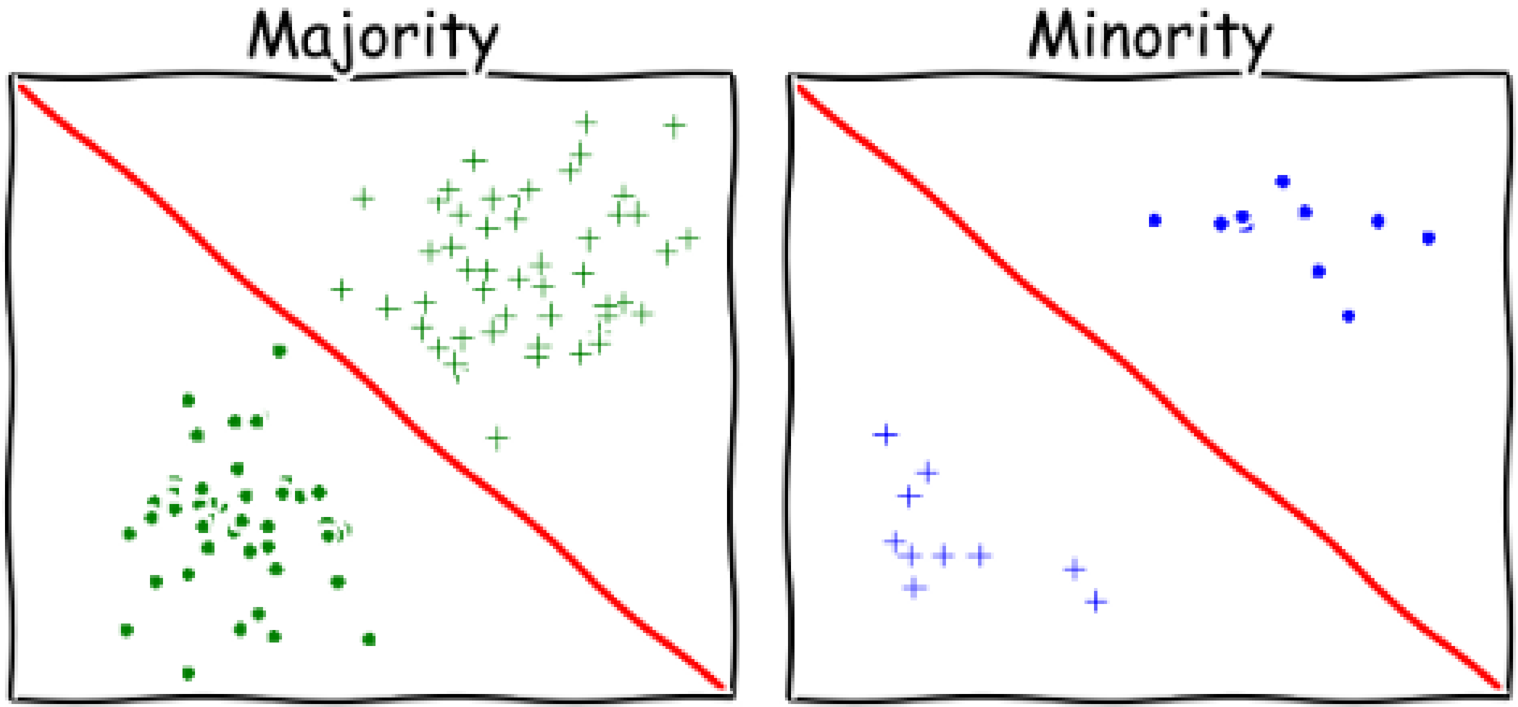
Issue #1: mimicking bias in data

Issue #2: introducing new bias not present in the data

Examples:

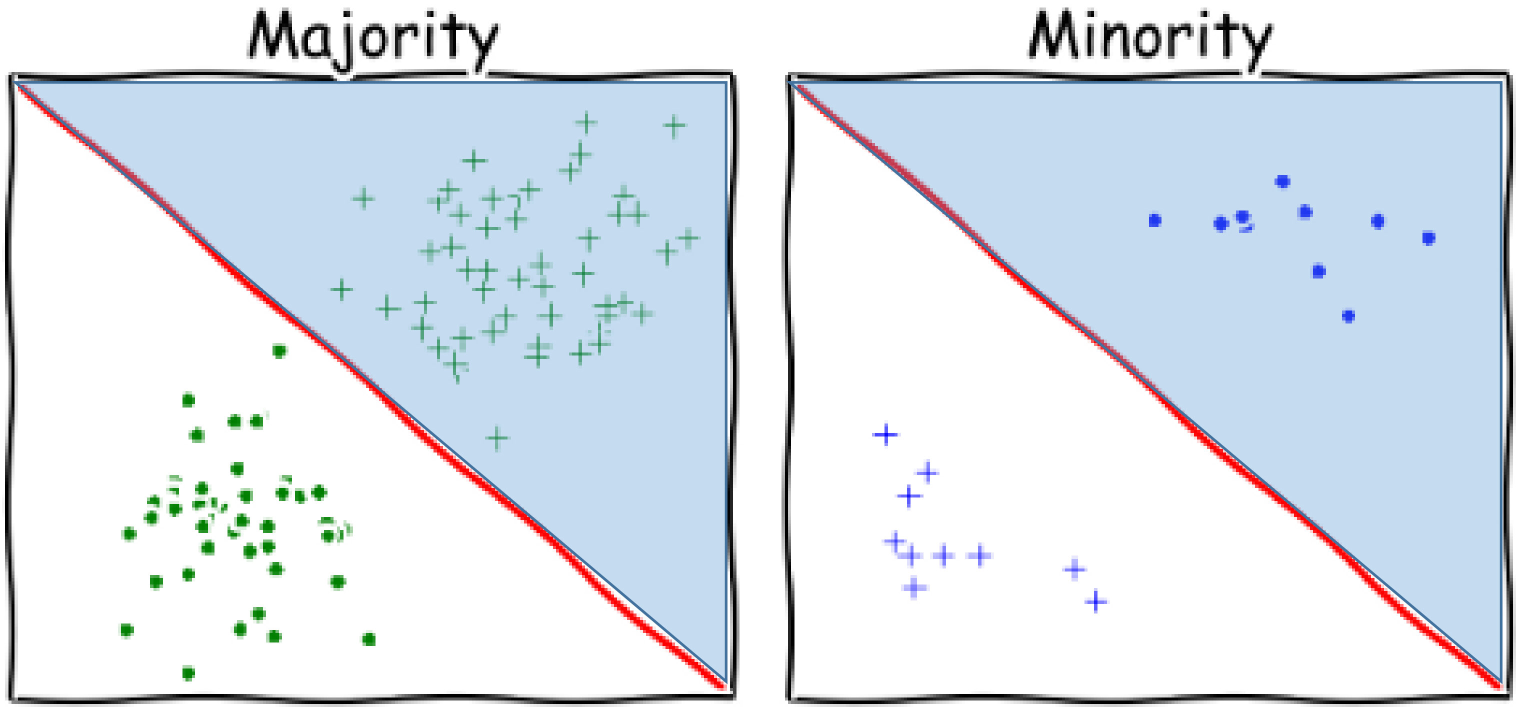
- Ad auctions from last lecture
- Sample size disparity

The machine learning loop: learning



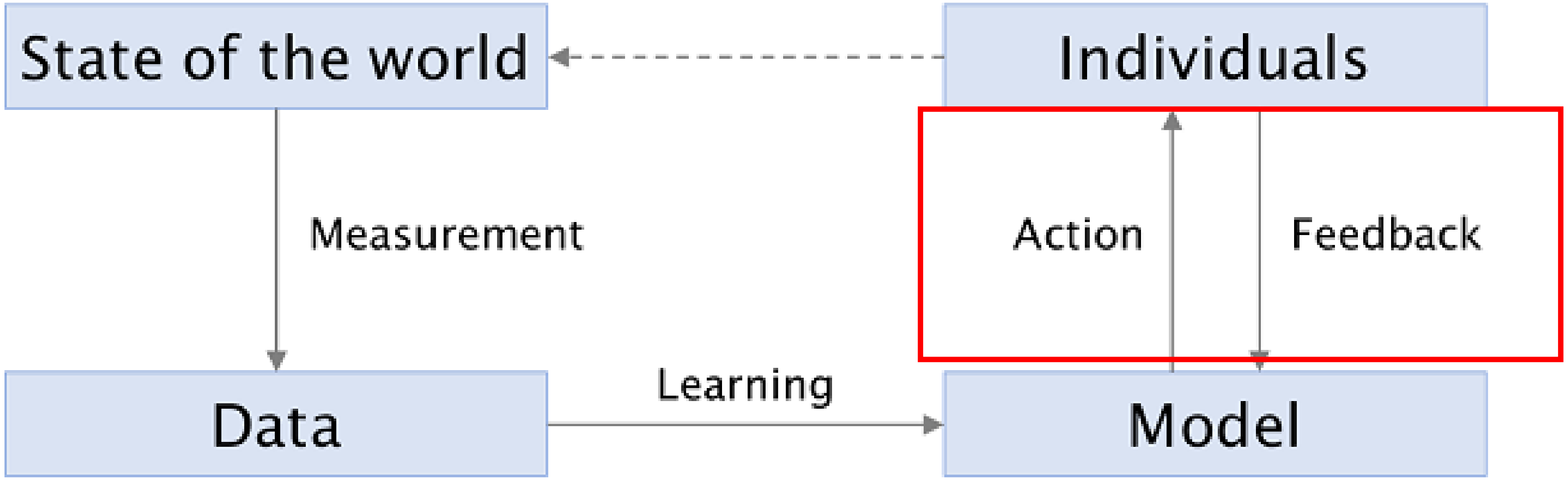
(From “How Big Data is Unfair”, by Moritz Hardt)

The machine learning loop: learning



(From “How Big Data is Unfair”, by Moritz Hardt)

The machine learning loop



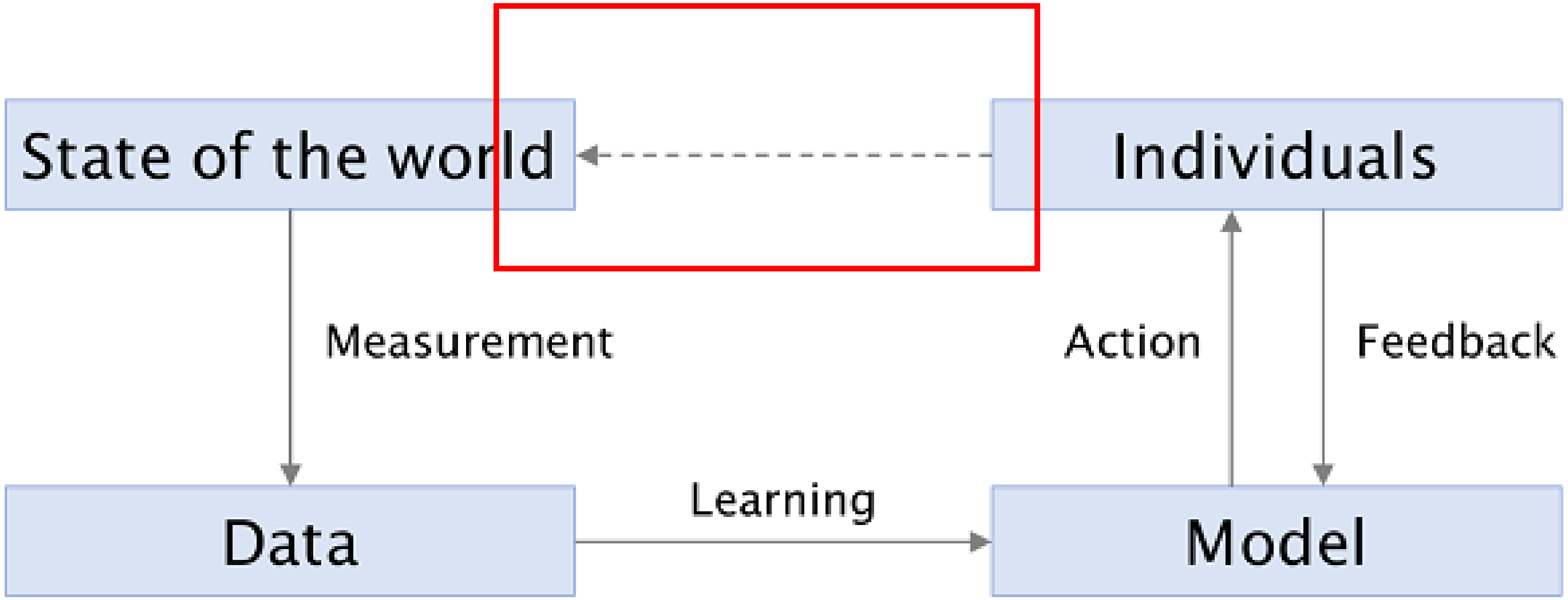
The machine learning loop: individuals react

How More Americans Are Getting a Perfect Credit Score

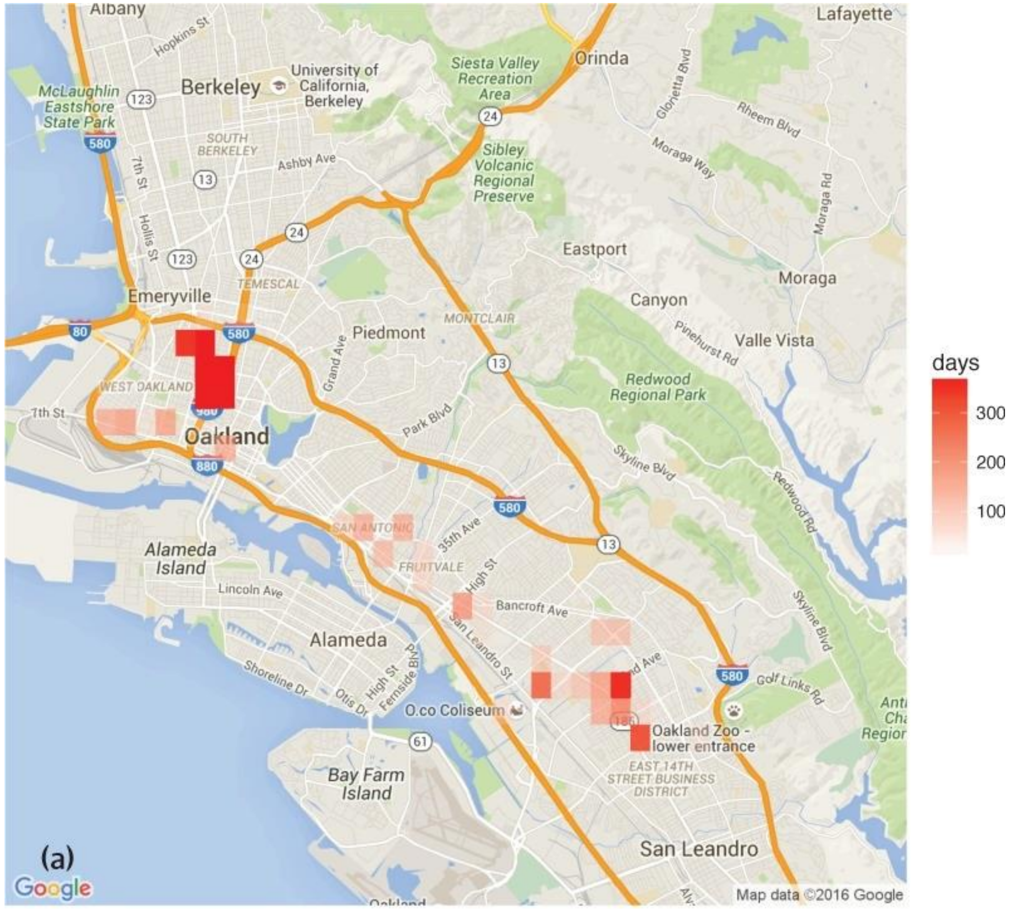
"Super-Prime" consumers are gaming the system in their pursuit of a golden 850 rating.

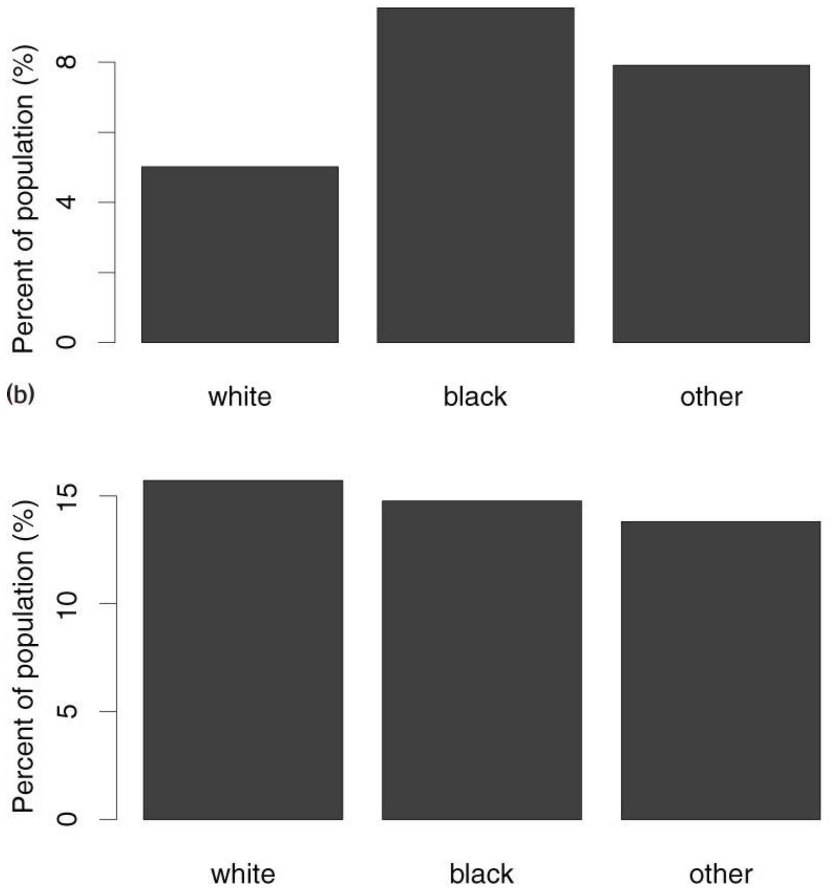
The credit scoring rule you use affects how humans manage their finances and try to improve credit

The machine learning loop: feedback



The machine learning loop: feedback

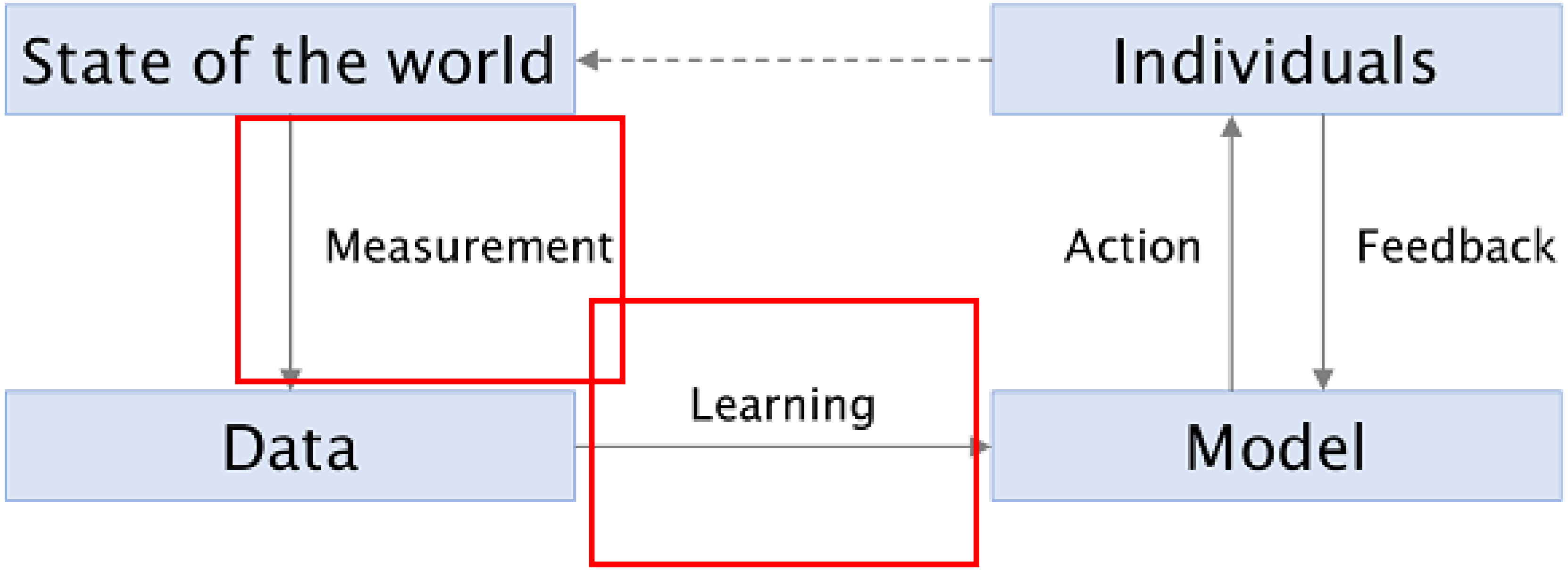




Populations targeted by PredPol

Estimated drug use

The machine learning loop



Our focus for the first few lectures on fairness.
Purely algorithmic problem: how to design algorithms that are as fair as possible

What assumptions do we make…

With respect to measurement?

- How well does data really reflect true attributes/state of the world?
- Do we assume the same data has the same meaning across different groups? Ethnic backgrounds? Socio-economic statuses?

With respect to our learning procedure?

- Does our objective guarantee fairness, even assuming the data is "perfect"? Do we see *direct discrimination?*
- Does our algorithm guarantee fairness, in the presence of biased data? In the presence of data with different meaning for different groups?
  Do we perpetuate *structural bias?*

What assumptions do we make…

“On the (im)possibility of fairness”

Sorelle A. Friedler
Carlos Scheidegger
Suresh Venkatasubramanian