

## Algorithms and Bias

- I. Algorithms
  - a. Historical foundations
  - b. Basic features
    - i. Finite set of data
    - ii. Public and understandable set of rules
    - iii. Not calculable by humans
    - iv. Iterative/Recursive
- II. Where bias and/or ethical and political choices can intervene in the application of algorithms?
  - a. The data set can be constructed and composed of data that is biased/non-randomly selected

*How do policing algorithms, such as Compstat, function? How may bias creep into the original dataset and how might this bias be exacerbated? How can this contribute to social unfairness?*

- b. The algorithm itself will necessarily incorporate ethical decisions into its rule-set.

*What was the purpose of the algorithm that was designed for the Los Angeles Homeless Services Authority? What problem was it trying to solve? What ethical theory did their solution embody and how did this lead to a problem of fairness?*

Possible alternative principles:

- i. Egalitarian
- ii. Sufficientarian
- iii. Prioritarian

- c. Algorithms can be *used* unfairly even if unbiased

- III. How to select amongst competing conceptions of fairness? John Rawls: *A Theory of Justice* (1971)
  - a. The concept of justice
  - b. How to generate theoretical agreement or convergence on principles of justice for society?
    - 1. Original position
    - 2. How to prevent bias and achieve unanimity: chose for everyone
      - 1) Motivational assumptions
      - 2) Veil of Ignorance
        - i. Defined:
        - ii. Categories
          - a) Race
          - b) Gender
          - c) Sexual Orientation

- d) Talent
- e) Conception of the Good

*What principle would you select if you did not know who you would be in this scenario?*