

# Hazarding A Guess: The Dangers of Mining Big Data



**OECD TECHNOLOGY FORESIGHT FORUM:  
“EXPLOITING THE DATA-DRIVEN ECONOMY”  
OCTOBER 22, 2012**

**SOLON BAROCAS**  
FELLOW, INTERNET SOCIETY  
DOCTORAL CANDIDATE, MEDIA, CULTURE, AND COMMUNICATION, NEW YORK UNIVERSITY



# Tail Wagging the Dog



- Evidence-based policy vs. policy-making based on evidence
  - A bias toward policy-making for which there is more easily accessible and cheaply obtainable evidence
- More than the rationalization of an existing, more ad hoc procedure; instead, Big Data may change the activity to which it is brought to bear
  - e.g., education in the United States...
  - e.g., the profound change policing experienced with the introduction of profiling: from a general concern with crime reduction (efficacy) to a narrower concern with the likelihood that each police action would result in crime detection (efficiency) (Harcourt 2007)
  - “anyone who is committing the same crime should face the same likelihood of being punished”
    - ✦ Fairness
    - ✦ Rule of law

# Optimization Problems



- Some obvious problems that Big Data pose the FIPPs...
- But some quite technical issues that are not yet well appreciated
  - Tension between fairness and accuracy
    - ✦ Purging illegal—but relevant—features from the model reduces the predictive accuracy of the model
      - Ironically, correcting for latent discrimination requires collecting precisely those pieces of information that are verboten
  - Tension between transparency and accuracy
    - ✦ In many cases, complex models tend to be more accurate than those that are more parsimonious
    - ✦ But the added complexity renders them inscrutable

# The Cost of Non-Erroneous and Lawful Discrimination



- **Discriminatory by design**
  - Non-erroneous discrimination?
  - Lawful discrimination?
    - ✦ Absence of animus
    - ✦ Not involving protected class
- **Fairness (Gandy 2009)**
  - To sort and evaluate to better target and tailor interventions
    - ✦ Produce unequal access to information, goods, and services
- **Equity and Inequality (Gandy 1993)**
  - Constrain life chances
  - Exacerbate historical disparities
- **Ratchet effect (Harcourt 2007)**
  - Stratification
  - Stigmatization

# Substantive Regulations



- Supplement privacy with other normative concerns
  - Autonomy
  - Non-discrimination
  - Fairness
- But this might necessitate a shift from procedural fairness to distributive justice
  - The challenges of developing rigorous technical definitions (Dwork et al. 2012)