Introduction to ML…
and Ethical Problems

Slides inspired in the materials of Susan Athey, Jan Spiess, Raj Chetty,…

Contrast with Traditional Econometrics

Economists have focused on the case with substantially more observations than covariates (N>>X)

- In-sample MSE is a good approximation to out-of-sample MSE
- OLS is BLUE, and if overfitting is not a problem, then no need to incur bias
- OLS uses all the data and minimizes in-sample MSE

OLS obviously fails due to overfitting when X~N and fails entirely when X>N

- ML methods generally work when X>N

Economists worry about estimating causal effects and identification

- Causal effects
- Counterfactual predictions
- Separating correlation from causality
- Standard errors
- Structural models incorporating behavioral assns

Identification problems can not be evaluated using a hold-out set

- If joint distance of observable same in training and test, will get the same results in both

Causal methods sacrifice goodness-of-fit to focus only on variation in data that identifies parameters of interest

What We Say v. What We Do (Econometrics)

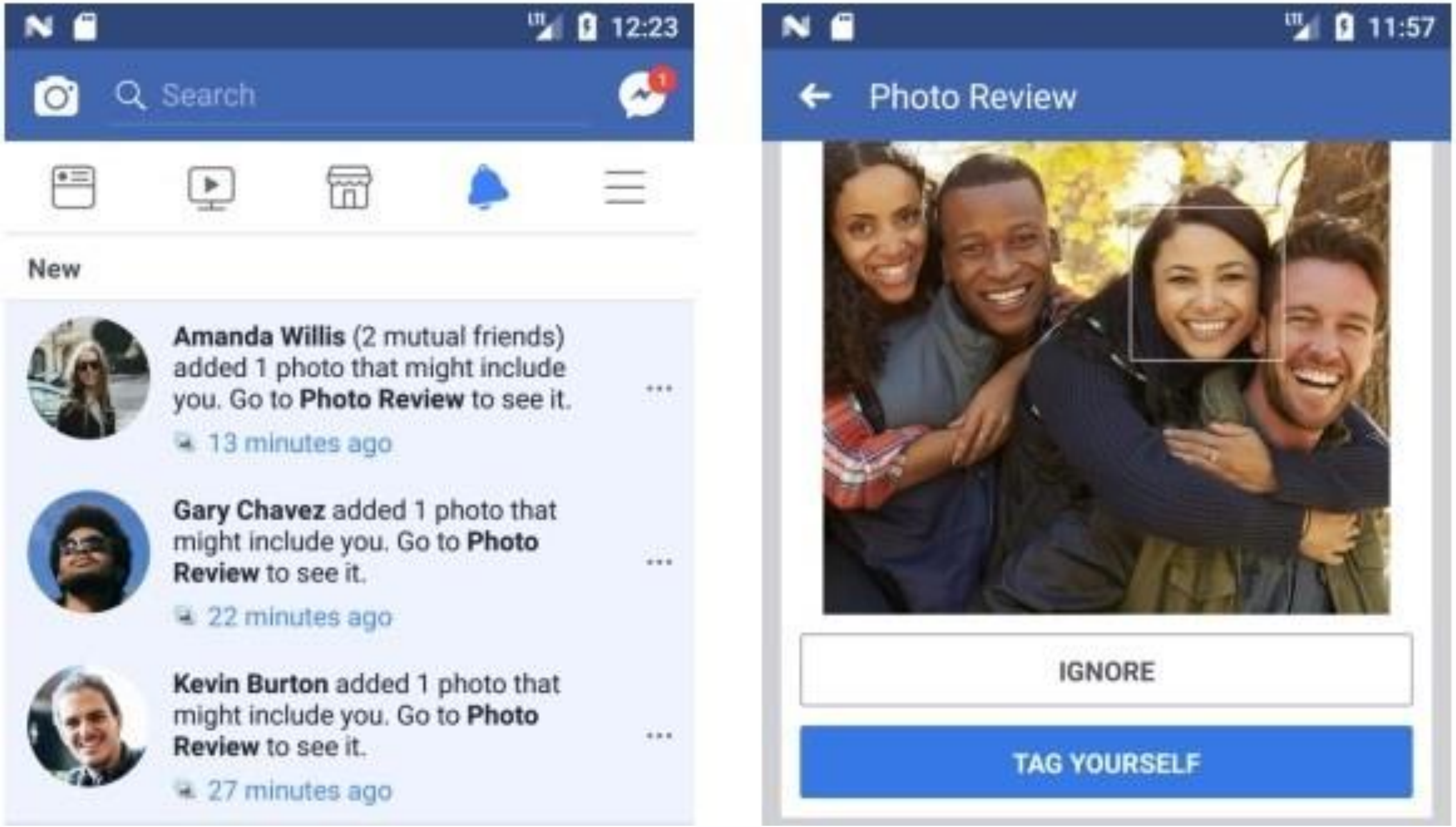
What We Say

- Causal inference and counterfactuals
- God gave us the model
- We report estimated causal effects and appropriate standard errors
- Plus a few additional specifications for robustness

What we do

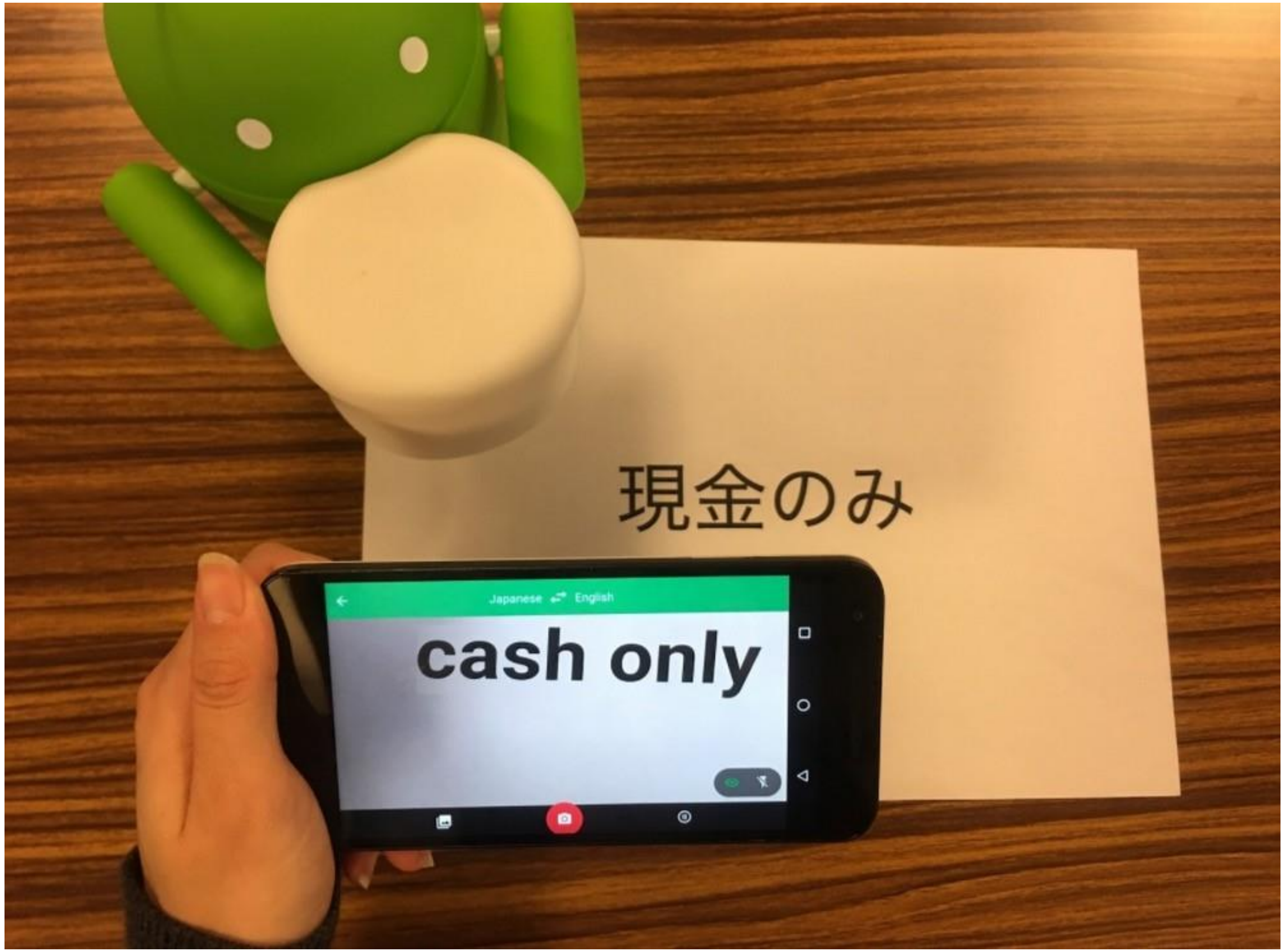
- Run OLS or IV regressions
  - Try a lot of functional forms
  - Report standard errors as if we ran only one model
  - Have research assistants run hundreds of regressions and pick a few “representative” ones
- Use complex structural models
  - Make a lot of assumptions without a great way to test them

Advances in Machine Intelligence



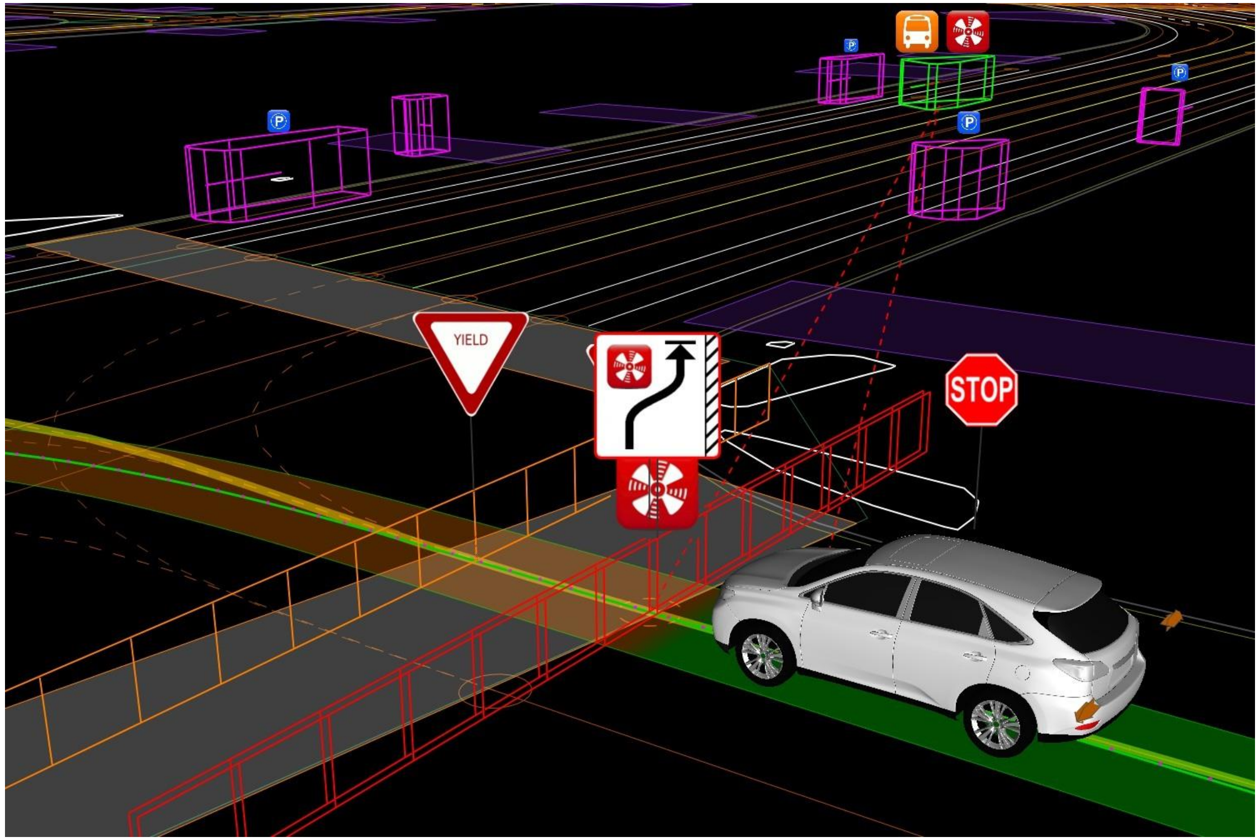
Picture source: Facebook

Advances in Machine Intelligence



Picture source: Google

Advances in Machine Intelligence



Picture source: Waymo

Some Features of Machine Learning (ML)

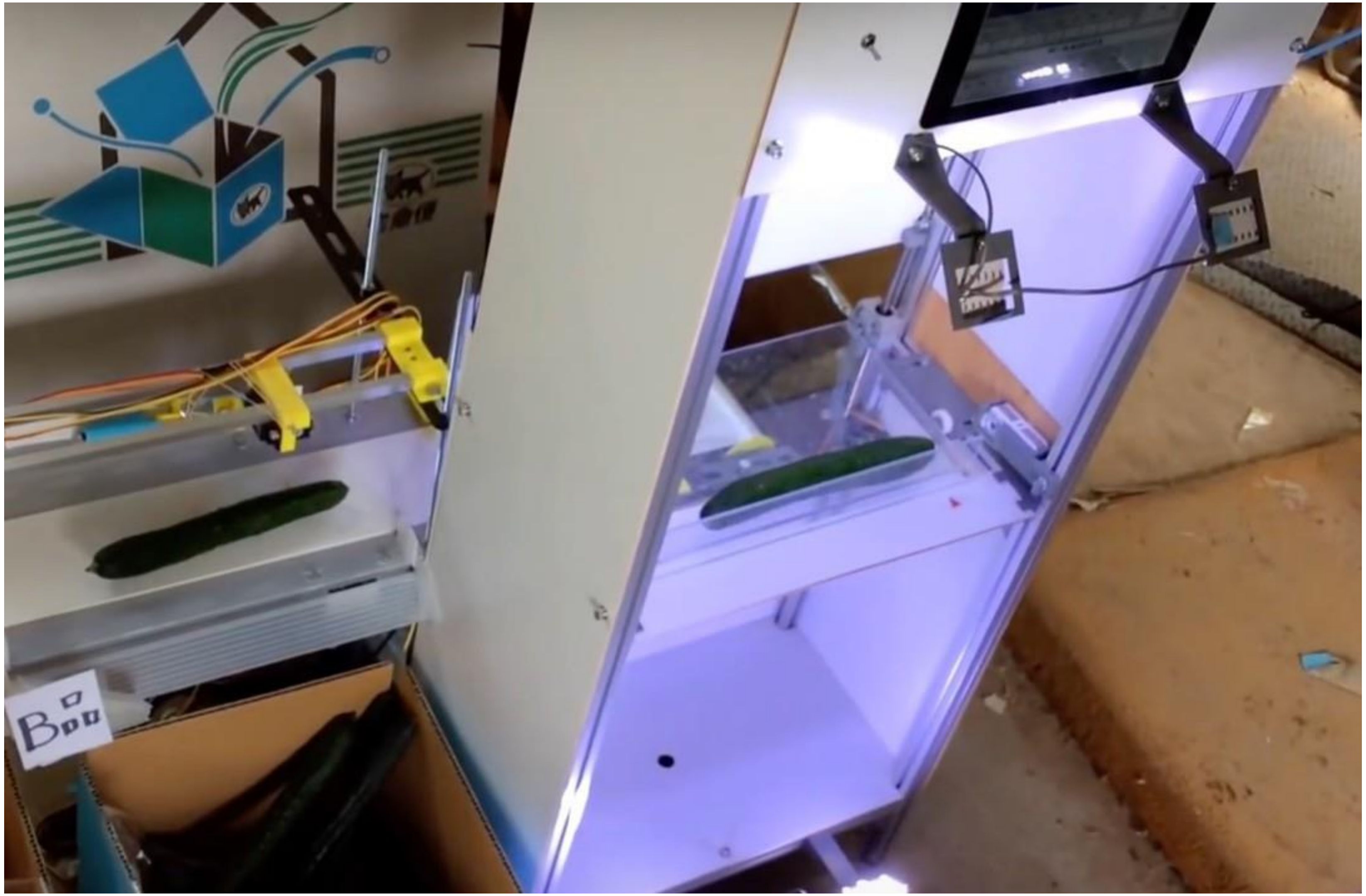
- Flexible, rich, data-driven models
- Can work with very high-dimensional data
- Limit expressiveness to avoid overfit (**regularization**)
- Learn how much expressiveness to allow (**tuning**)
- Industry-strength tools readily available

- Supervised learning: focus on **prediction**

→ Idea: turn intelligence task into supervised-learning problems

  - Bank decides who to give credit to
  - Tax authority decides which returns to audit
  - Image recognition
  - Self-driving cars

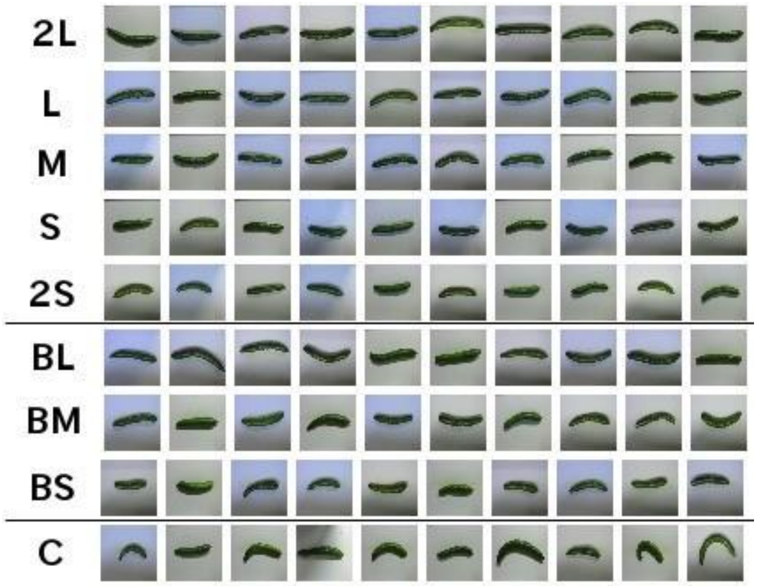
Machine Intelligence in the Field



Images: Makoto Koike (via The New Yorker), Kazunori Sato (via YouTube)

Cucumber Classification Problem

$$\underbrace{y}_{\text{cucumber quality}} = f(\overbrace{x}^{\text{pixels}})$$



- Old-style AI: deduce from human intuition, introspection
- New-style ML: induce from training data
  - Take “labelled” data
  - Fit a function f in the training sample

Image source: Google

Predictive Analytics in Criminal Justice

Motivation: Biases in Human Decision Making

- Growing interest in using machine learning (predictive analytics) tools to aid decision makers, e.g. in the context of criminal justice

- Humans’ decisions often exhibit substantial biases

- Recall the Bertrand and Mullainathan (2004) study biases in hiring by sending out fictitious resumes with identical credentials in response to real job ads

  - Vary name of applicant to be “white-sounding” (e.g., Emily Walsh or Greg Backer) vs. “black-sounding” (e.g., Lakisha Washington or Jamal Jones)

Biases due to Decision Fatigue

- Such biases in decisions are not driven entirely by deep-rooted beliefs

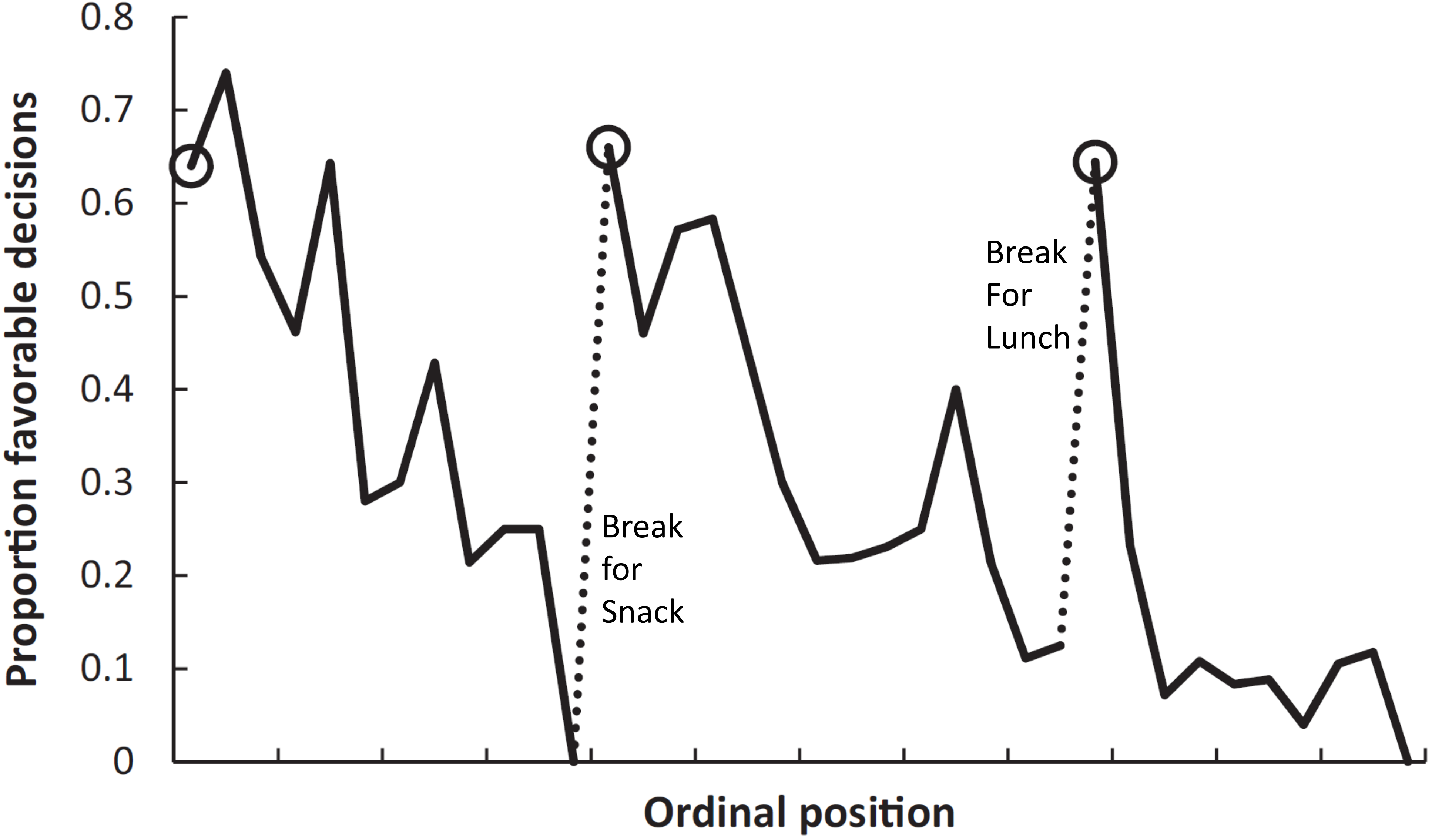
- Decisions also vary greatly based on transitory factors unrelated to substantive features of the issue at hand

- Danziger et al. (2011) demonstrate this by analyzing data on judges’ decisions to grant prisoners parole

Studying Decision Fatigue in Parole Decisions

- Data: 1,100 judicial rulings on parole for prisoners in Israel over 10 months
- Judges review about 20 cases on average each day in succession
  - Ordering of cases depends upon when attorney shows up and is essentially random
- Judges can decide to grant parole or reject (delay to a future hearing, maintaining status quo)
- Key institutional feature: two breaks during the day for meals
  - 10 am for late-morning snack (40 mins)
  - 1 pm for lunch (1 hour)

Proportion of Rulings in Favor of Prisoner by Time of Interview During the Day



Source: Danziger, Levav, Avnaim-Pesso (PNAS 2011)

Can Machines Help Humans Overcome Biases?

- Can machine learning help us reduce such biases?
- Idea: develop algorithms to predict outcomes of interest and use these to guide or replace human decisions
- These algorithms are not necessarily subject to human biases, but do they outperform humans' decisions overall or have other biases/shortcomings?

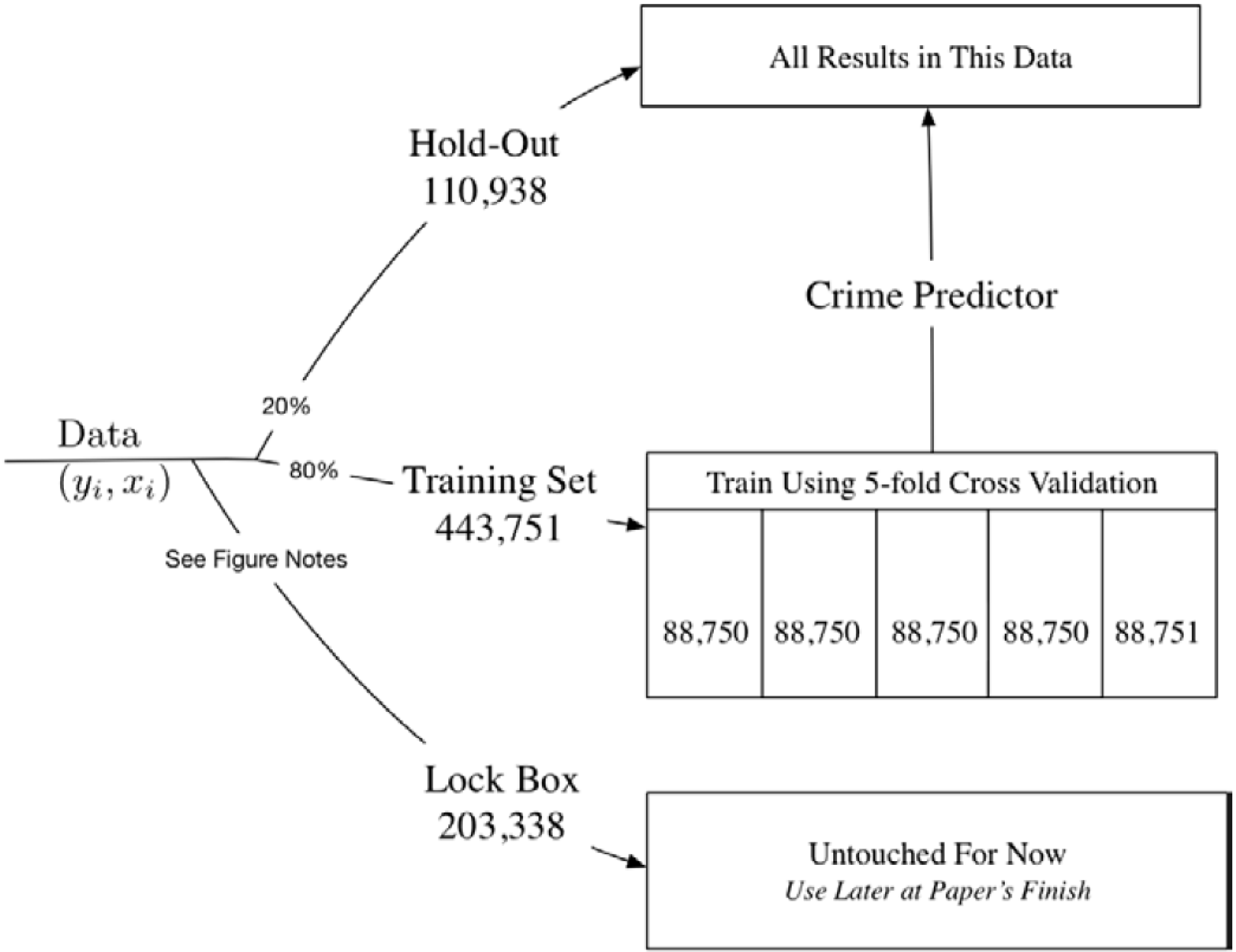
Decisions to Jail vs. Release Defendants

- Kleinberg et al. (2017) compare the accuracy of human decisions and machine predictions in the criminal justice system
- Every year, ~10 million people are arrested in the U.S.
- After arrest, judges decide whether to hold defendants in jail or let them go
- By law, decision is made with the objective of minimizing risk of flight (failure to appear at trial)
- Kleinberg et al. compare machine learning predictions and judges' actual decisions in terms of performance in achieving this objective

Decisions to Jail vs. Release Defendants

- Data: 750,000 individuals arrested in New York City between 2008-2013
  - Same data on prior history that is available to judge (rap sheet, current offense, etc.)
  - Data on subsequent crimes to develop and evaluate performance of algorithm
- Define “crime” as failing to show up at trial; objective is to jail those with highest risk of committing this crime
  - Other definitions of crime (e.g., repeat offenses) yield similar results
- First divide data into three separate samples

Data Used for Empirical Analysis



Source: Kleinberg et al. (2017)

Methodology: Machine Learning Using Decision Trees

- Predict probability of committing a crime using a machine learning method called decision trees

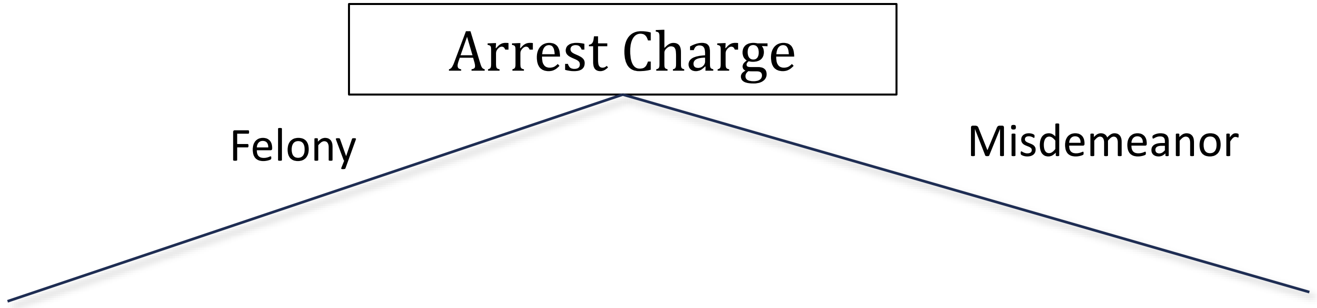
- Main statistical challenge: need to avoid overfitting the data with large number of potential predictors
  - Can get very good in-sample fit but have poor performance out-of-sample
  - Solve this problem using cross-validation, using separate samples for estimation and evaluation of predictions

Methodology: Machine Learning Using Decision Trees

- Three steps to develop predictions using decision trees

1. Split the data based on the variable that is most predictive of differences in crime rates

Hypothetical Decision Tree for Decision to Jail Defendant



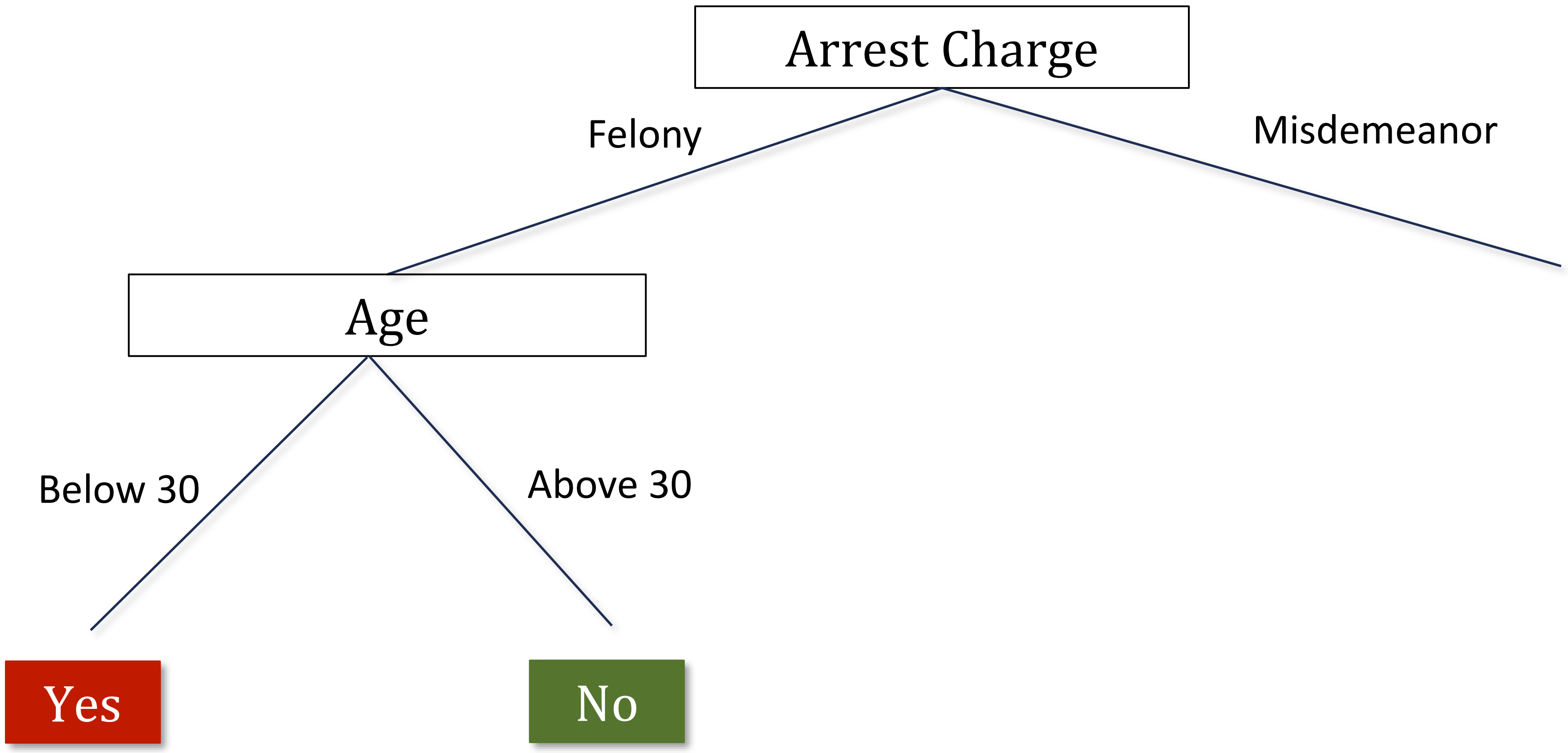
Methodology: Machine Learning Using Decision Trees

- Three steps to develop predictions using decision trees

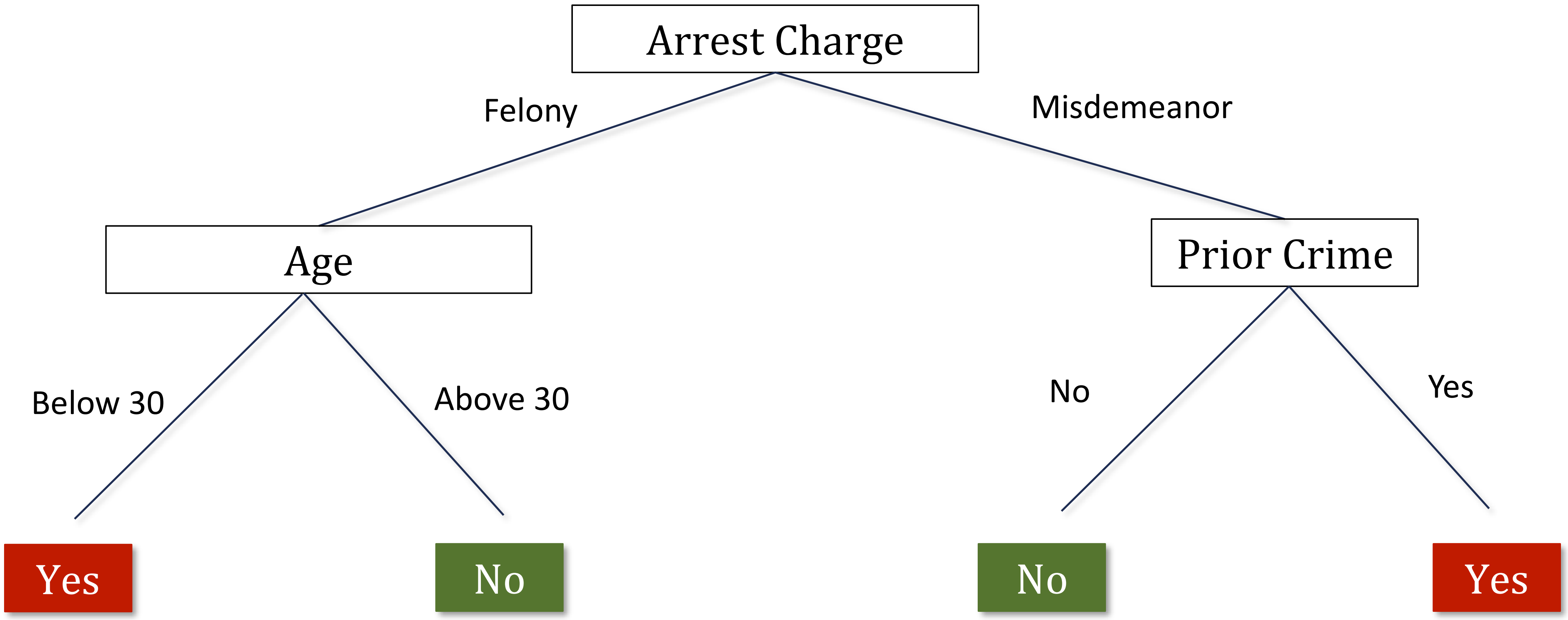
1. Split the data based on the variable that is most predictive of differences in crime rates

2. Grow the tree up to a given number of nodes *N*

Hypothetical Decision Tree for Decision to Jail Defendant



Hypothetical Decision Tree for Decision to Jail Defendant



Methodology: Machine Learning Using Decision Trees

- Three steps to develop predictions using decision trees

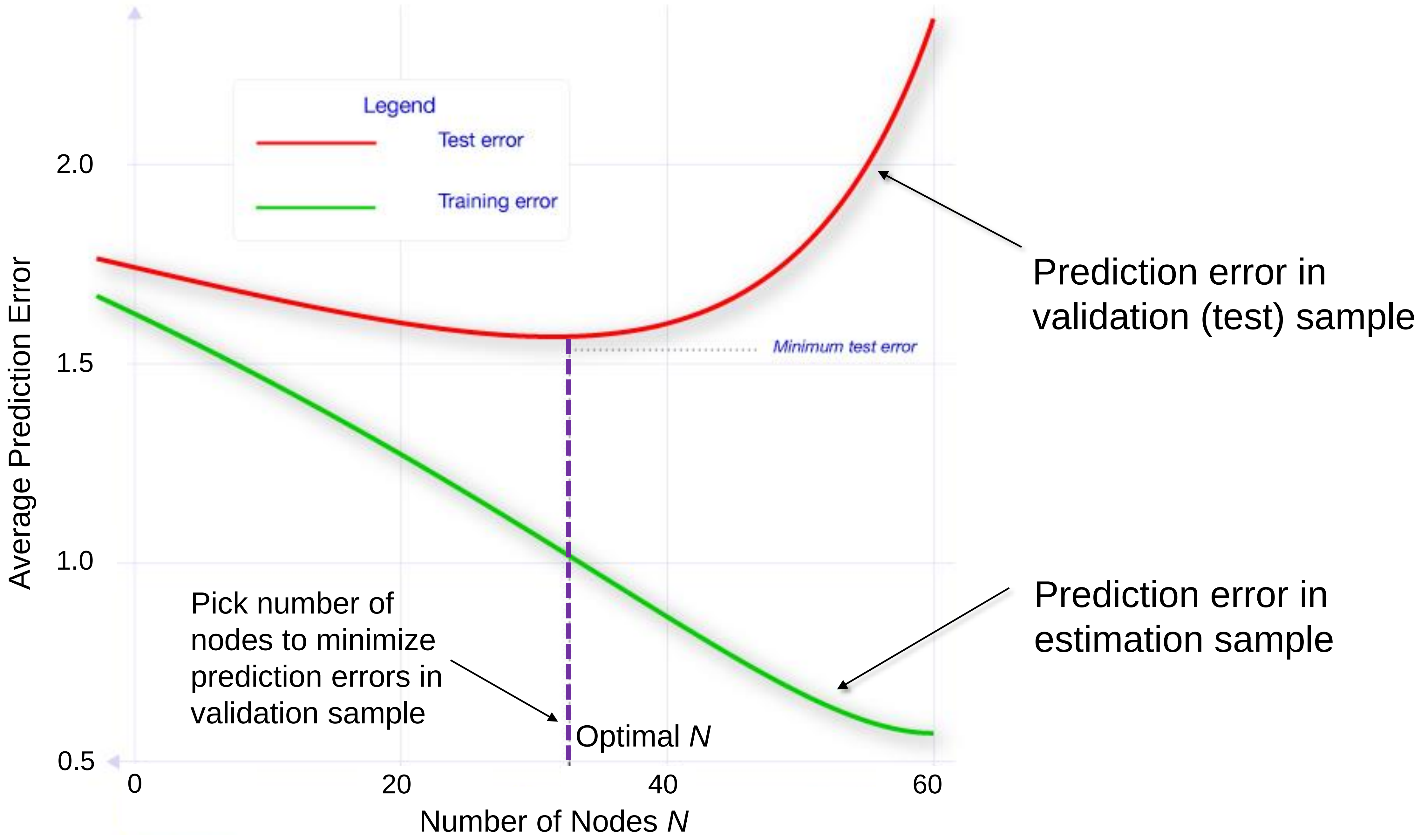
1. Split the data based on the variable that is most predictive of differences in crime rates

2. Grow the tree up to a given number of nodes *N*

3. Use separate validation sample to evaluate accuracy of predictions based on a tree of size *N*

- Repeat steps 1-3 varying *N* and choose tree-size *N* that minimizes average prediction errors

Prediction Errors vs. Size of Decision Tree



Comparing Machine Predictions to Human Predictions

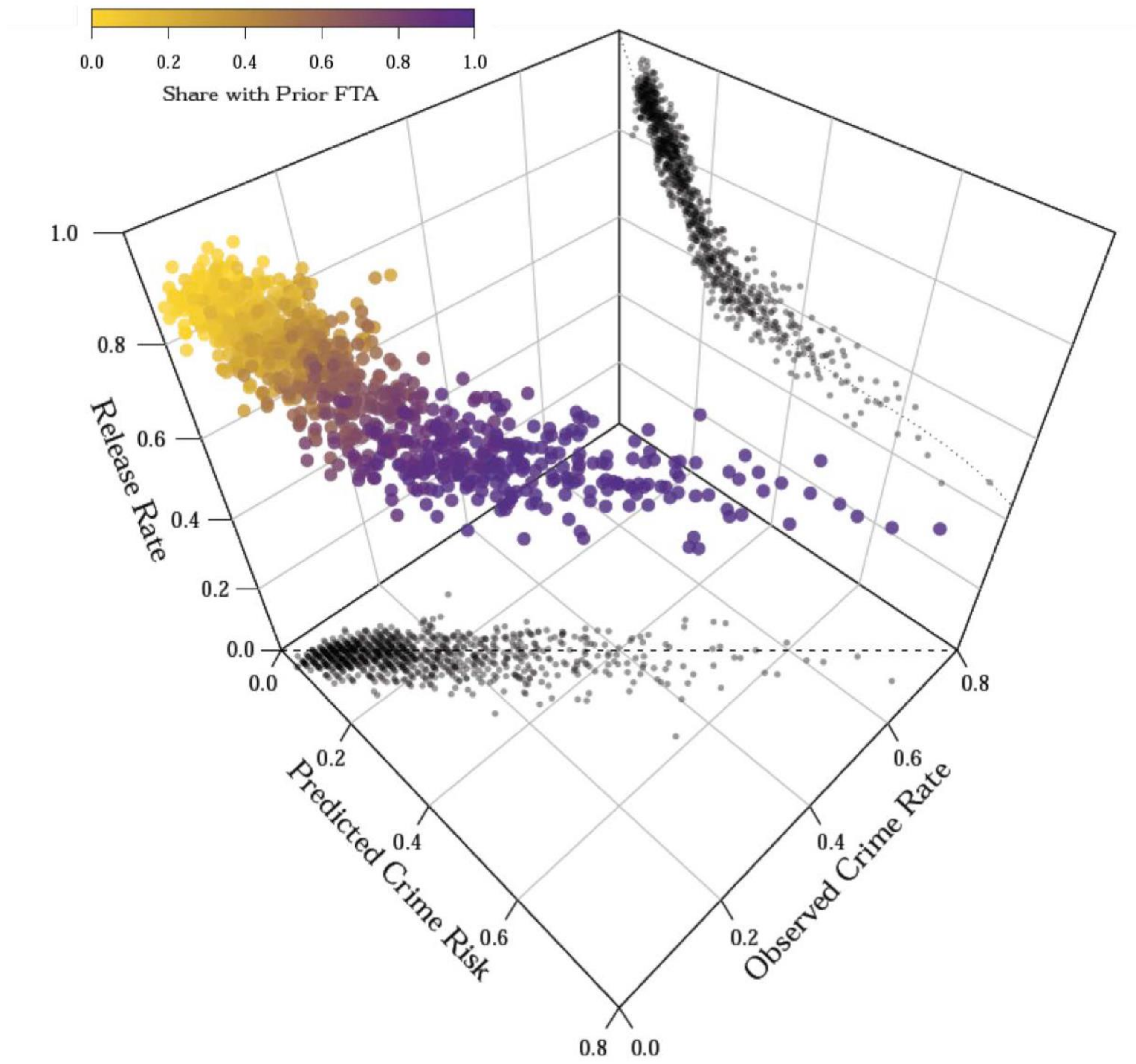
- Applying this method yields predictions of crime rates for each defendant

- Machine-based decision rule: jail the defendants who have the highest predicted risk

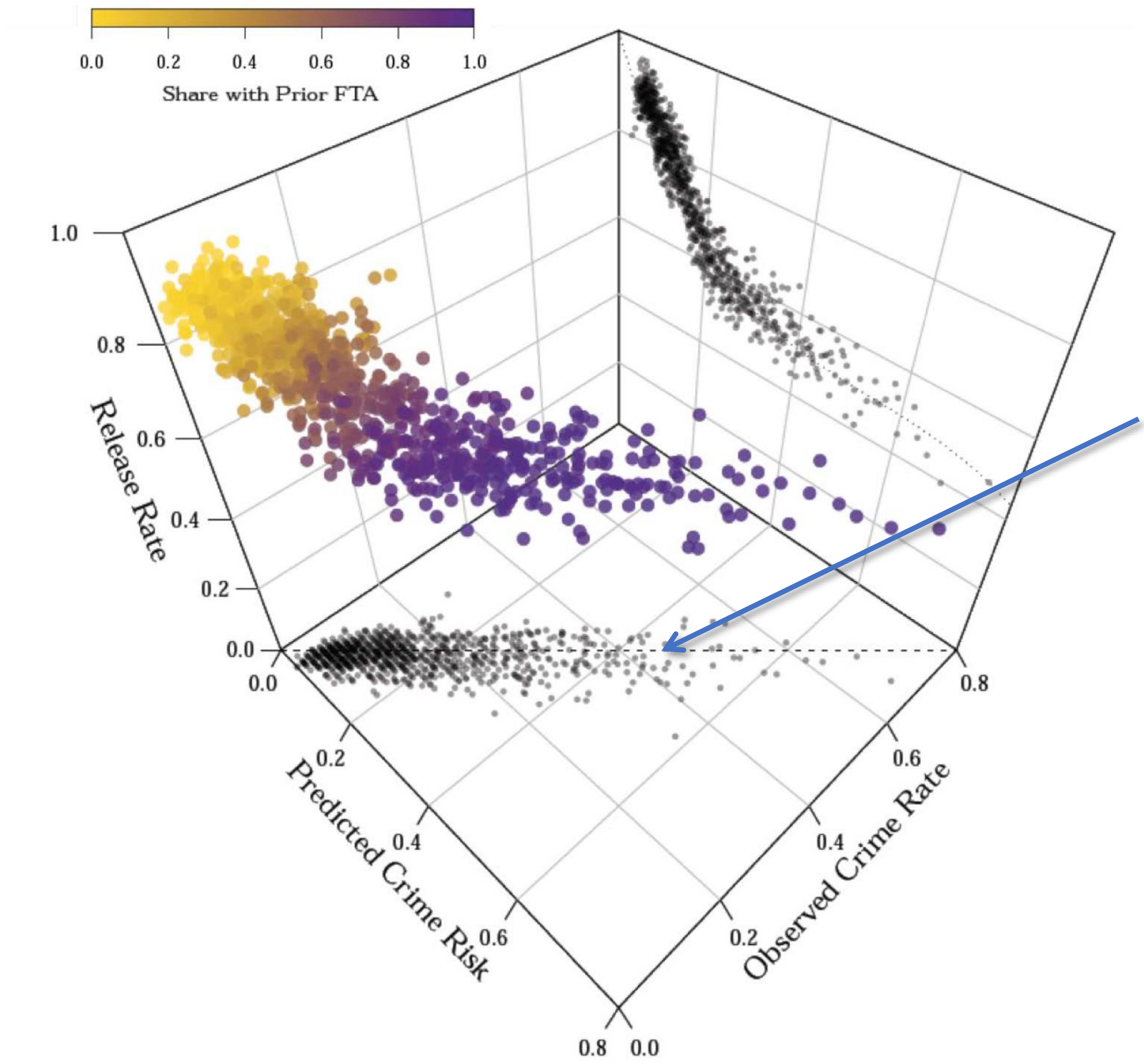
- How does this machine-based rule compare to what judges actually do in terms of crime rates it produces?

- Answer is not obvious: judges can see things that are not in the case file, such as defendant’s demeanor in courtroom

Judges' Release Decisions vs. Machine Predictions and Crime Risk

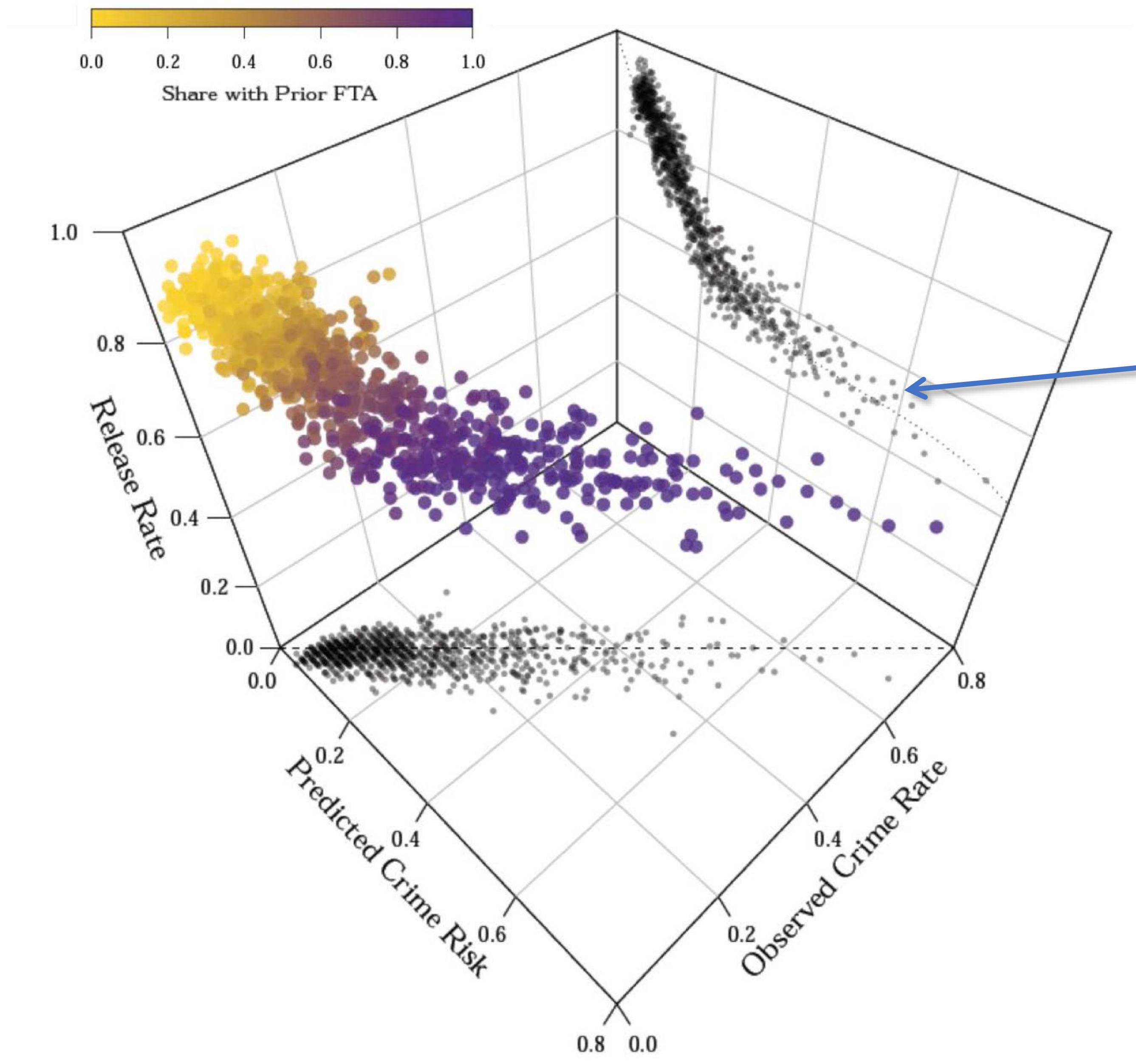


Judges' Release Decisions vs. Machine Predictions and Crime Risk



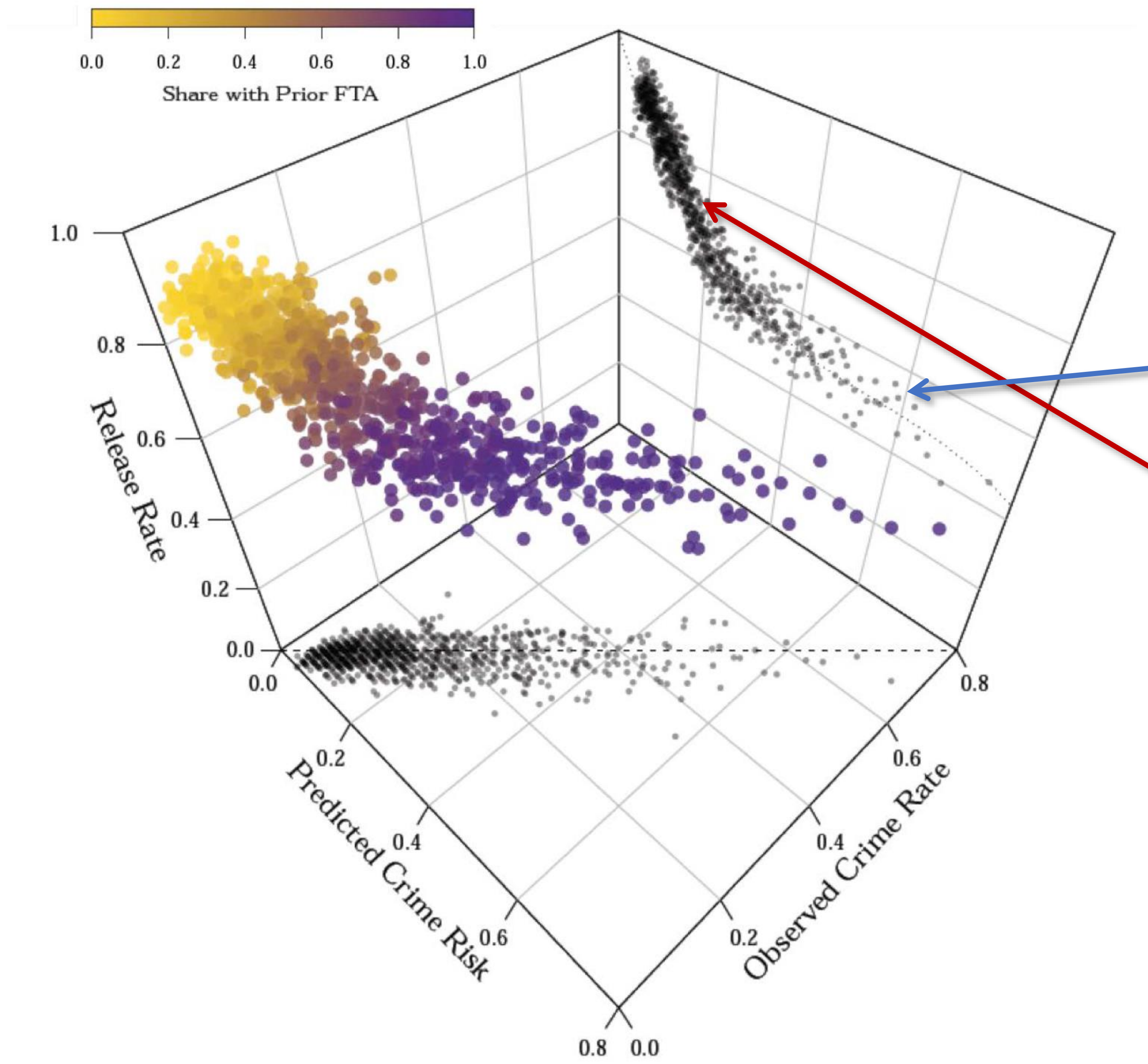
Predicted crime risk is highly correlated with observed crime rates (for defendants who are released)

Judges' Release Decisions vs. Machine Predictions and Crime Risk



Judges release 50% of defendants whose predicted crime risk exceeds 60%

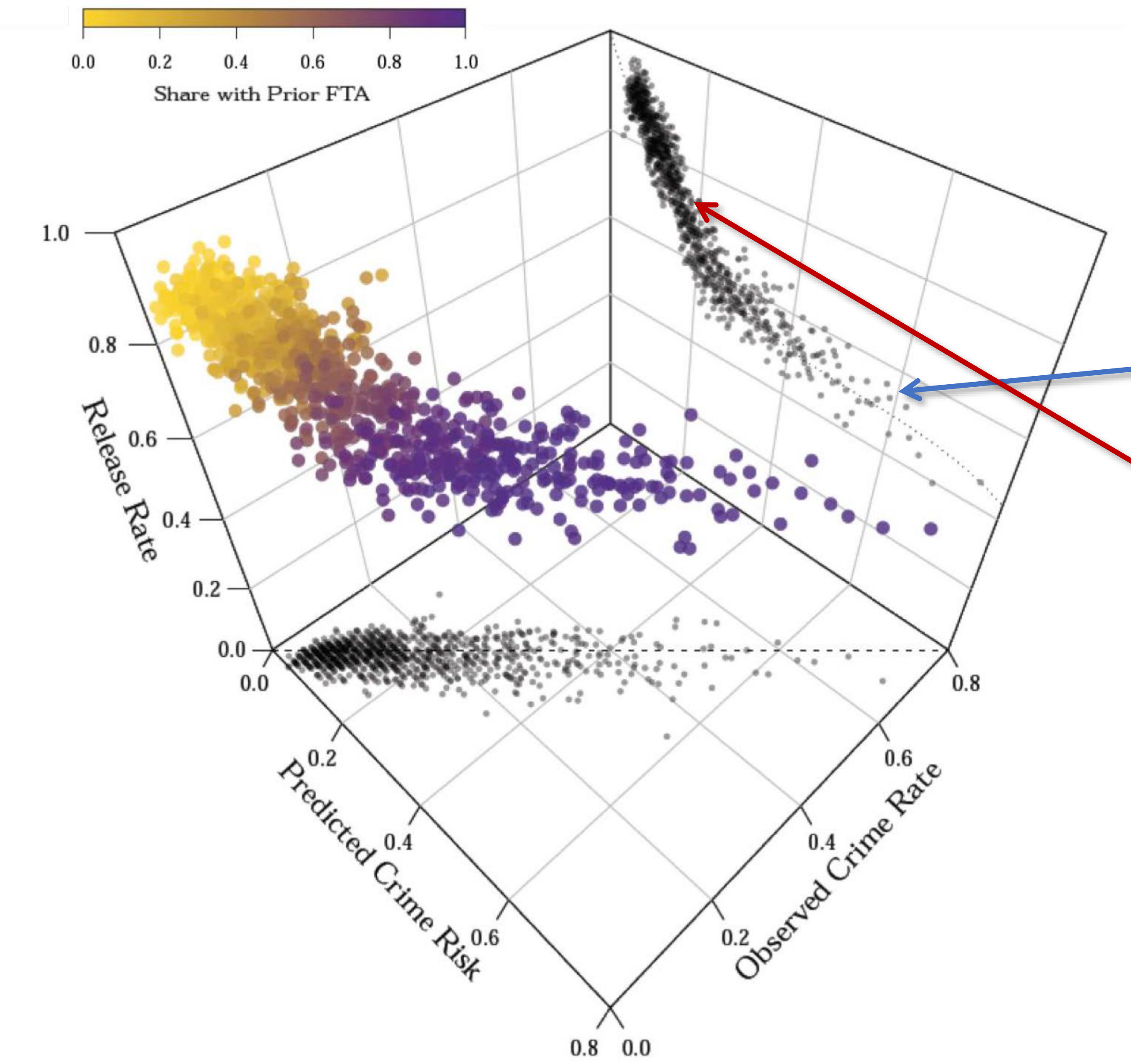
Judges' Release Decisions vs. Machine Predictions and Crime Risk



Judges release 50% of defendants whose predicted crime risk exceeds 60%

Yet they jail 30% of defendants with crime risks of only 20%

Judges' Release Decisions vs. Machine Predictions and Crime Risk



Judges release 50% of defendants whose predicted crime risk exceeds 60%

Yet they jail 30% of defendants with crime risks of only 20%

Swapping low-risk and high-risk defendants would keep jail population fixed while lowering crime rates

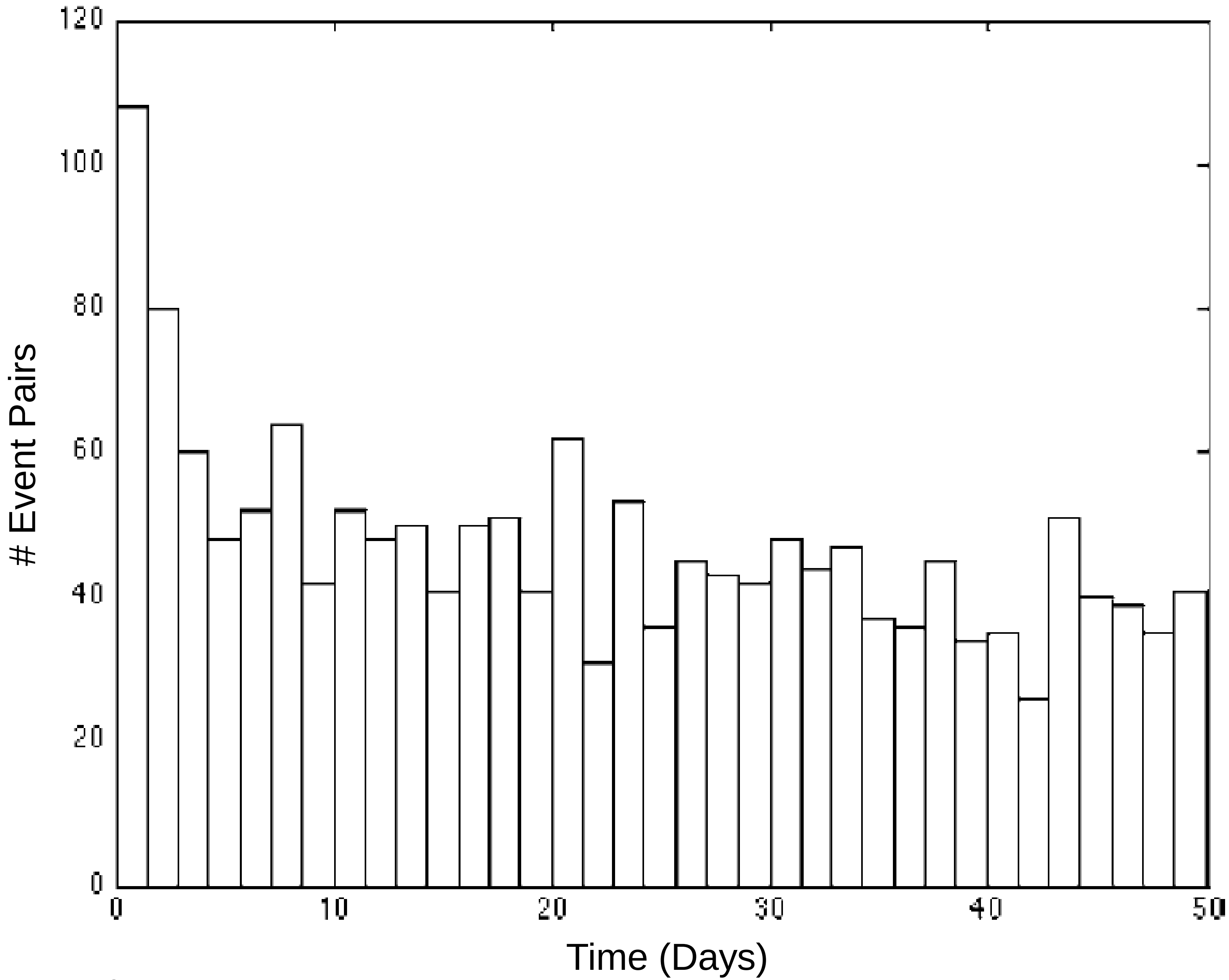
Comparing Machine Predictions to Human Predictions

- How large are the gains from machine prediction?
  - Crime could be reduced by 25% with no change in jailing rates
  - Or jail populations could be reduced by 42% with no change in crime rates
- Why? One explanation: Judges may be affected by cues in the courtroom (e.g., defendant’s demeanor) that do not predict crime rates
- Whatever the reason, gains from machine-based “big data” predictions are substantial in this application

Predictive Policing

- Another active area of research and application of big data in criminology: predictive policing
  - Predict (and prevent) crime before it happens
- Two approaches: spatial and individual
  - Spatial methods rely on clustering of criminal activity by area and time

Times Between Burglary Events Separated by 0.1 Miles or Less



Source: Mohler et al. (JASA 2011)

Predictive Policing

- Another active area of research and application of big data in criminology: predictive policing
    - Predict (and prevent) crime before it happens
- Two approaches: spatial and individual
    - Spatial methods rely on clustering of criminal activity by area and time
    - Individual methods rely on individual characteristics, social networks, or data on behaviors (“profiling”)

Debate Regarding Predictive Analytics

- Use of big data for predictive analytics raises serious ethical concerns, particularly in the context of criminal justice
- Tension between two views:
  - Should a person be treated differently simply because they share attributes with others who have higher risks of crime?
  - Should police/judges/decision makers discard information that could help make society fairer and potentially more just than it is now on average?