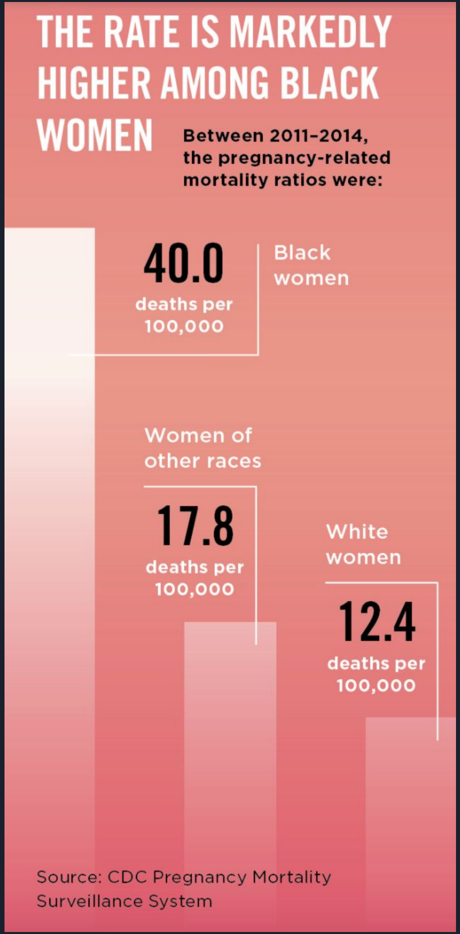
Defund Police

Chasity Savella & Aidan Wolfe

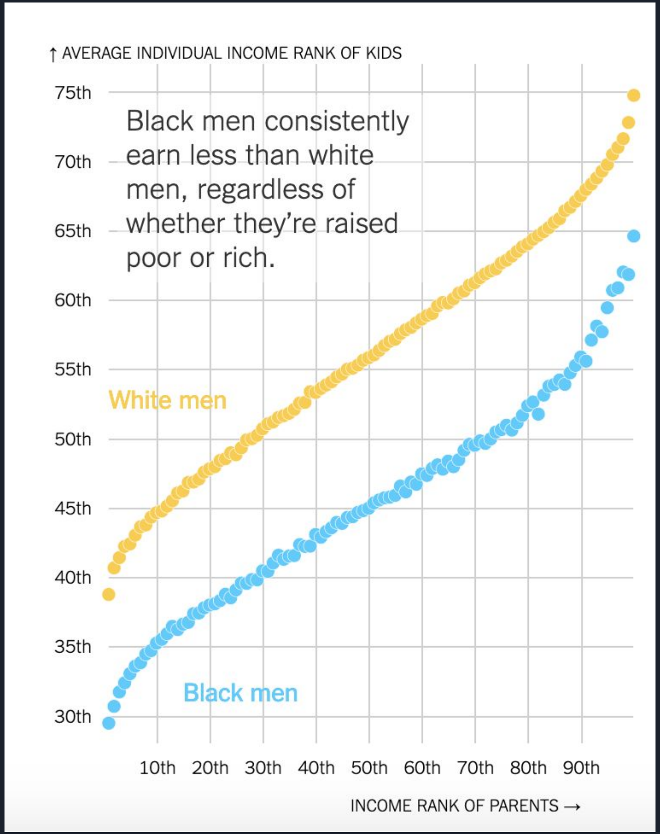
Racial Disparity in the US

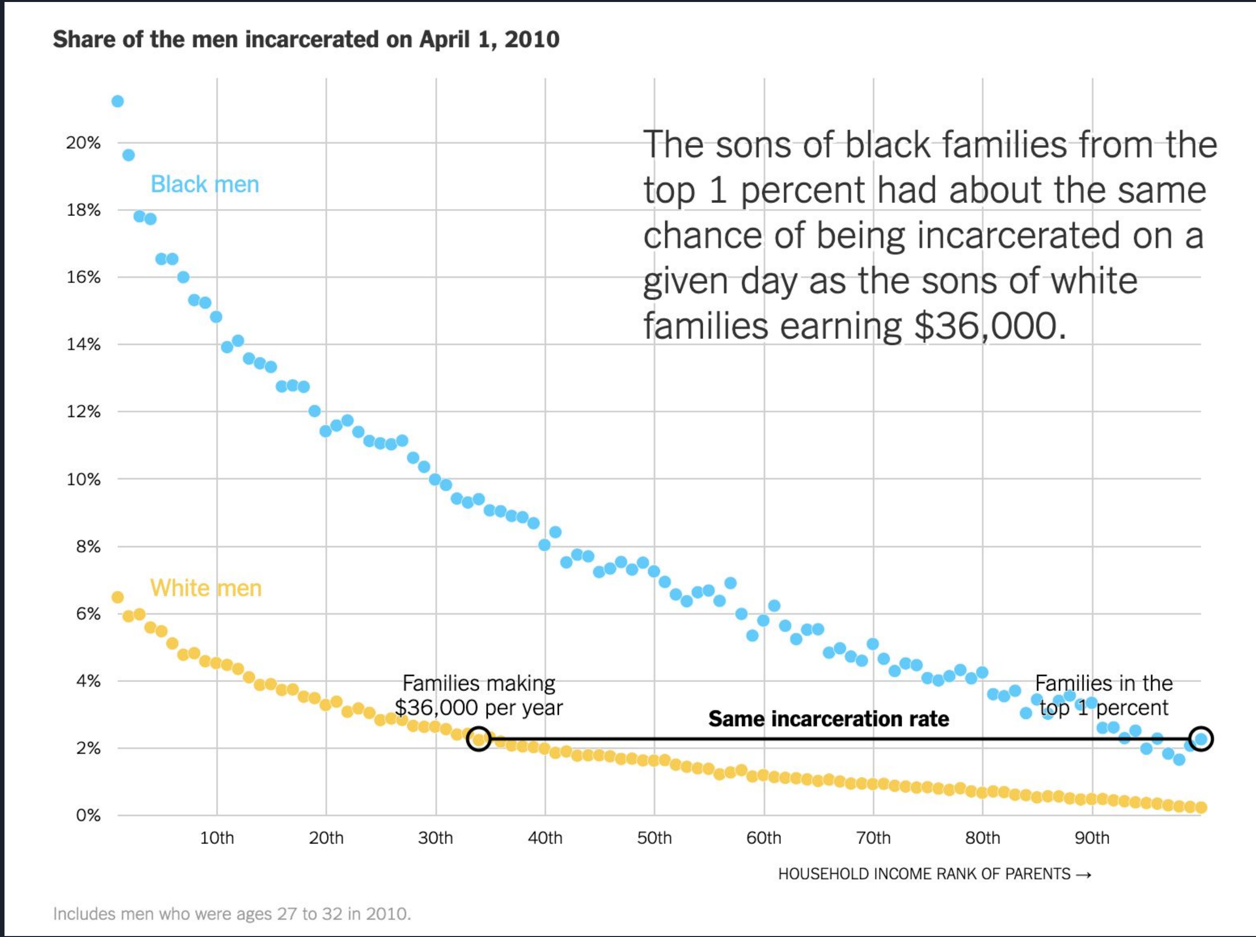
- The wealth inequality and achievement gap between Black and white people in the United States is undeniable.
- If one does not believe that this is because of systemic forces, then one must ask themselves if they believe fundamentally that all races are equal.
- If they say to themselves or others that they believe this to be true, then they must ask themselves what explains this disparity between the living quality of Black and white people in the United States?
- It is either that they think Black people are not capable of achieving as much as white people given the same opportunities, or that there is in fact greater opportunity afforded to white people because of structural forces and conversely that there are greater barriers to success facing Black people.

Data on Racial Disparity in the US



Data on Racial Disparity in the US




Share of the men incarcerated on April 1, 2010
20%
18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
Black men
White men
The sons of black families from the top 1 percent had about the same chance of being incarcerated on a given day as the sons of white families earning $36,000.
Families making $36,000 per year
Same incarceration rate
Families in the top 1 percent
10th
20th
30th
40th
50th
60th
70th
80th
90th
HOUSEHOLD INCOME RANK OF PARENTS →
Includes men who were ages 27 to 32 in 2010.

Background on Police Violence/Prison Industrial Complex

- Black people are incarcerated in state prisons at a rate that is 5.1 times the imprisonment of whites.
    - In five states (Iowa, Minnesota, New Jersey, Vermont, and Wisconsin), the disparity is more than 10 to 1.
    - Black people make up 2% of Oregon's state population and 10% of Oregon's prison/jail population.
    - Indigenous peoples are 1% of the Oregon state population and make up 3% of Oregon's prison/jail population.
- In twelve states, more than half of the prison population is Black: Alabama, Delaware, Georgia, Illinois, Louisiana, Maryland, Michigan, Mississippi, New Jersey, North Carolina, South Carolina, and Virginia. Maryland, whose prison population is 72% Black, tops the nation.
- Over the life course, about 1 in every 1,000 black men can expect to be killed by police.
    - For men of all races, the number is 1 in 2,000

Using Data To Perpetuate Inequality

- Four papers:
    1. Layers of Bias: A Unified Approach for Understanding Problems with Risk Assessment - Eckhouse, Lum, Conti-Cook, and Ciccolini, *Criminal Justice and Behavior* (2019)
    2. Measures of Fairness For New York City's Supervised Release Risk Assessment Tool - Lum and Shah, *Human Rights Data Analysis Group* (2019)
    3. To Predict and Serve? - Lum and Isaac, *Significance Magazine* (2016)
    4. Runaway Feedback Loops in Predictive Policing - Ensign, Friedler, Neville, Scheidegger, and Venkatasubramanian, *Proceedings of Machine Learning Research* (2018)

Bias in Risk Assessment Data Models

- Sentencing and pretrial release decisions have faced scrutiny over disparate outcomes for Black people compared to whites
- In an attempt to remove the personal biases of sentencing judges from the equation, risk assessment tools based on data have been implemented to provide recommendations and information to judges
- “the use of risk-assessment tools has expanded dramatically. In particular, they have come to play a key role in decisions about sentencing, probation, and pretrial detention. Over 20 state courts use some form of risk assessment at sentencing, and a single tool, the Arnold Foundation’s Public Safety Assessment, is in use throughout three states, as well as in more than two dozen local jurisdictions (Schuppe, 2017).” Layers of Bias - p 186

Layers of Bias: A Unified Approach for Understanding Problems With Risk Assessment

- This paper outlines three criteria for which we can assess the fairness of risk assessment tools, and explores whether these criteria are currently and can ever be met within the current state of our criminal justice system
- "...judges, parole boards, and other criminal justice decision- makers (as well as immigration authorities) are increasingly turning to data-driven models to predict who will reoffend. Vendors promote these models to the public and to the agencies that use them as the answer to human bias, arguing that computers cannot harbor personal animus or individual prejudice based on race, gender, or any other legally protected characteristic." p 186
- "In this article, we focus on criminal justice decisions mostly in the pretrial release and sentencing contexts, because we believe the stakes are highest when liberty decisions are involved." p 187

Three Layers of Bias

- Top Layer: Fair Algorithms
    - Does the algorithm make fair decisions?
- Middle Layer: Data Quality
    - Is the data used to calculate scores biased in a fundamental way?
    - “The use of arrest as a measure of criminality fundamentally assumes that people who do the same things are arrested at the same rates.” p 196
- Base Layer: Fundamental Conceptual Problems with the Fairness of Data-Driven Decisions
    - Is it fair to use data about other people to make decisions about another person’s liberty?
    - “All models rely fundamentally on the assumption that we can use the behavior of other people to decide whether a particular defendant is too dangerous to release” p 189

Algorithmic Fairness

- Three Criteria:
  1. Does the score generated by a model mean the same thing across different groups?
     - "...if a tenth of Black defendants with a 30% score commit a crime, while half of White defendants with the same score commit a crime, the test does not exhibit predictive fairness." p 189
  2. Do people who do not commit a later crime get a similar score across groups?
     - Are Black defendants who do not reoffend more likely to get high-risk scores than white defendants who do not reoffend?
     - "False Positive"
  3. Do people who go on to commit crimes get similar scores across groups?
     - "White defendants who ultimately commit a crime might be more likely to make bail than Black defendants who likewise commit a second crime." p 190
     - "False Negative"

Algorithmic Fairness

- "Achieving fairness on all three measures is not just practically difficult; it is conceptually impossible if there are differences in the measured rate of reoffending across different groups (Chouldechova, 2017; kleinberg et al., 2016). This impossibility theorem is crucial to understanding the obstacles to developing a fair model. It is mathematically impossible to develop a model that will be fair in the sense of having equal predictive value across groups, and fair in the sense of treating members of groups similarly in retrospect." p 190
- 2016 ProPublica analysis attempted to study to the fairness of a tool made by a private company (Northpointe) called Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)

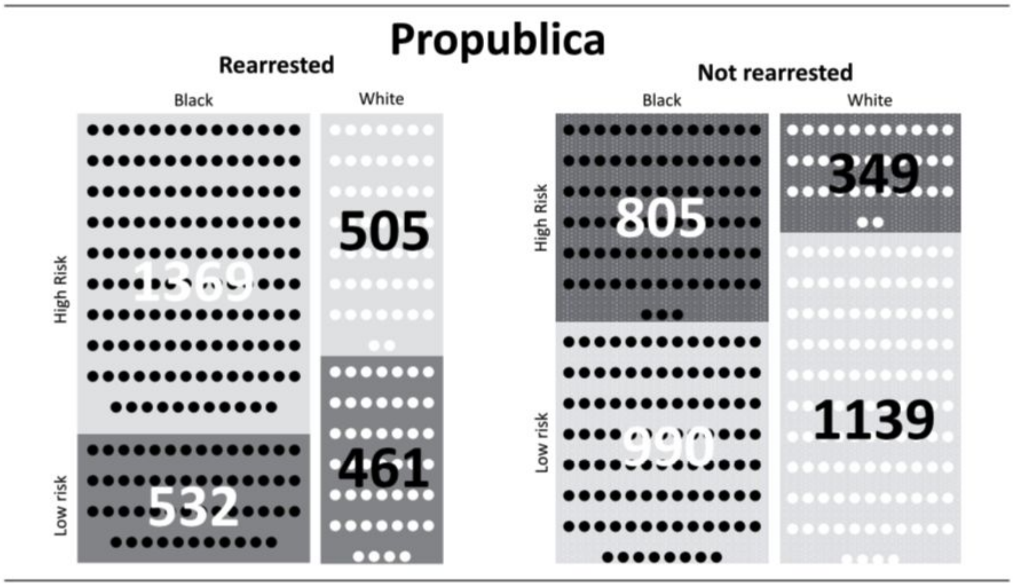


Figure 1: **ProPublica's Analysis Categorized Defendants by Whether They Went on to be Rearrested**
Note. There are clear differences in the "riskiness" categorization for Black and White defendants with the same outcomes, as shown by the unequal height of the bars in this graph.

"By analyzing the accuracy of the predictions by race, ProPublica concluded that COMPAS was nearly twice as likely to inaccurately predict that a Black defendant was at high risk for rearrest as a White defendant."

"of those who were rearrested, White defendants were more likely to have been classified as low risk than Black defendants. Similarly, of those who were not rearrested, Black defendants were more likely to have been classified as high risk than White defendants." p 190

"Therefore, the algorithm violated the second and third criteria described above, providing false positive and false negative rates that differed by race"

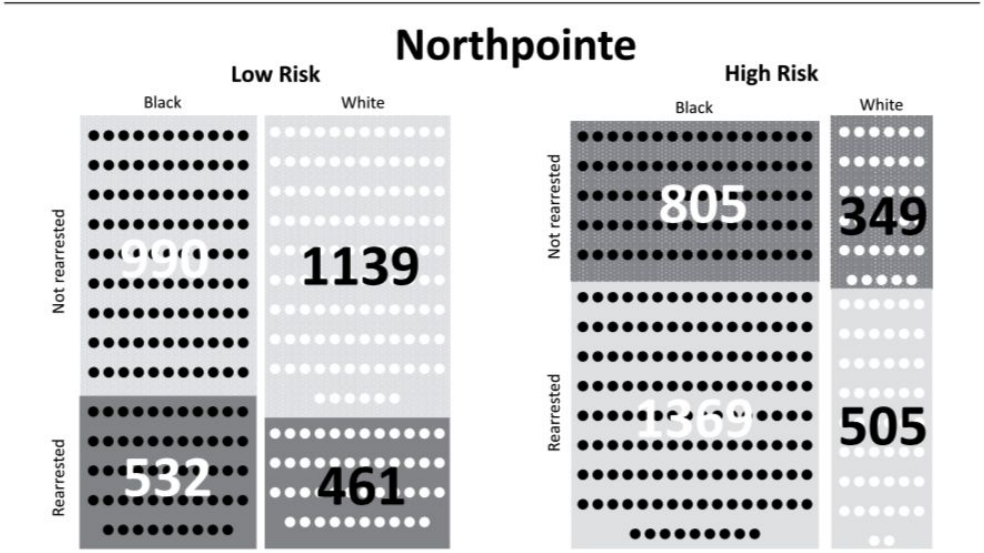


Figure 2: Northpointe's Analysis Categorized Defendants by Their Risk Score, Then Measured How Many of Those in Each Category Reoffended

Note. The share of Black and White defendants with the same risk score who go on to be rearrested is similar and thus, the bars are of similar heights.

- Northpointe argues "of those classified as high risk, the proportion who were not actually rearrested was roughly equivalent between White and Black populations. Similarly, they found that of those who were classified as low or medium risk, Blacks and Whites had a roughly equal chance of being rearrested." p 191
- However, "note the substantial difference in the overall rate of rearrest for Black and White defendants in this analysis: 51% of Black defendants were rearrested versus 39% of White defendants. This difference means that models will predict a greater proportion of Black defendants will be rearrested than White defendants, because the models *assume that the future will be like the past.*" p191-192

Algorithmic Fairness

- Proponents of these types of data models argue they are “race neutral” if they do not include race as a variable
- However, many variables act as proxies for race
  - “In a society structured by racism and segregation, many variables commonly included in models, from location to employment to prior police encounters, will be correlated with race.” p193
- “There is no clear standard for the level of predictive accuracy needed to justify using models for high-stakes questions like liberty decisions.” p.194

Data Bias

- “The use of arrest as a measure of criminality fundamentally assumes that people who do the same things are arrested at the same rates.” p196
- “Black Americans are disproportionately likely to be stopped and searched by police, whether they are driving or walking … White and Black Americans use marijuana and other drugs at similar rates, but Black Americans are much more likely to be arrested for drug possession... *This is a problem in the criminal justice system, but it is also a problem with criminal justice data.*”
- “The model is not predicting individual behavior, but an event influenced by police decision-making.” p197
- “Using arrests and other criminal justice data as an unbiased source of information on individual behavior would require us first to build a racially unbiased criminal justice system.” p197

Measures of Fairness for Supervised Release Risk Assessment Tool

Notes on data used in the following discussions:

- Combined data from Criminal Justice Agency database and NY Division of Criminal Justice Services database
- Risk assessment tool to screen defendants suitability for acceptance into a pretrial supervised release program
  - Created statistical model to predict likelihood that a defendant will be re-arrested for a felony during the pre-trial period
- Takes demographic and criminal history and outputs a risk category.
  - Range: low - high risk
- Predictive accuracy of about 0.67
- Data collected to train model was collected during peak years of Stop, Question, and Frisk
  - 87% of criminals caught by SQF were Blacks and Latinos

Measures of Fairness for Supervised Release Risk Assessment Tool

	race	Brooklyn	Manhattan	Queens	Staten Island	Bronx
1	WHITE	0.30	0.29	0.38	0.63	0.98
2	BLACK	1.62	4.51	1.12	2.26	2.67
3	HISPANIC	1.21	2.26	0.54	0.65	2.11

TABLE 2. Drug-related felony arrests per 1000 people

MEASURES OF FAIRNESS FOR NYC'S SUPERVISED RELEASE RISK MODEL

	race	Brooklyn	Manhattan	Queens	Staten Island	Bronx
1	WHITE	0.06	0.09	0.08	0.08	0.08
2	BLACK	0.69	1.43	0.85	0.83	0.64
3	HISPANIC	0.32	0.64	0.23	0.28	0.28

TABLE 3. Weapons possession felony arrests per 1000 people

- The number of felony arrests for each borough and racial group is divided by the total number of individuals with the same racial designation
- Insufficient sample sizes for other ethnic groups
- Disparities are partially the result of racially biased enforcement
- Significantly impacts the development of the data model

Measures of Fairness for Supervised Release Risk Assessment Tool

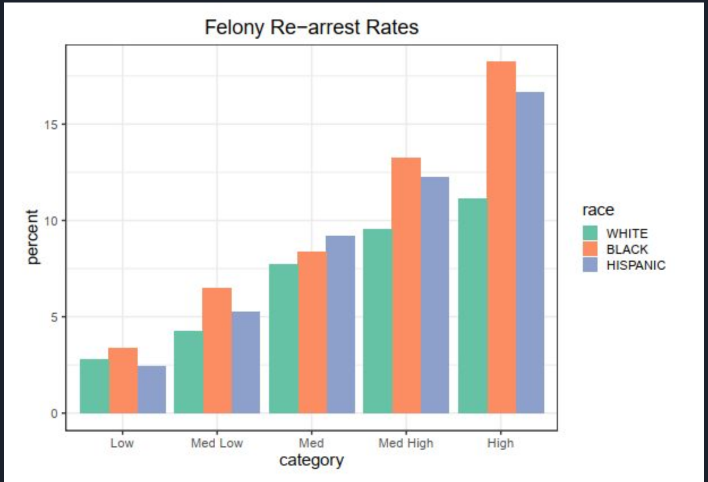


FIGURE 1. Percent of defendants with a pre-trial felony re-arrest for each risk category and race group

- Predictive parity: equal re-arrest rates by race within each risk category
- Figure 1 shows re-arrest per one hundred defendants by race and risk score
- This data model does not meet the ‘predictive parity’ criterion

Measures of Fairness for Supervised Release Risk Assessment Tool

race	FPR (high)	FPR (medium-high +)
WHITE	0.03	0.11
BLACK	0.07	0.22
HISPANIC	0.05	0.17

TABLE 5. False positive rates by race.

- False positive parity: measuring false positive rates per race
- False positive: any person who had a positive prediction for felony re-arrest, but was not re-arrested
- This data model does not meet the 'false positive parity' criterion

Measures of Fairness for Supervised Release Risk Assessment Tool

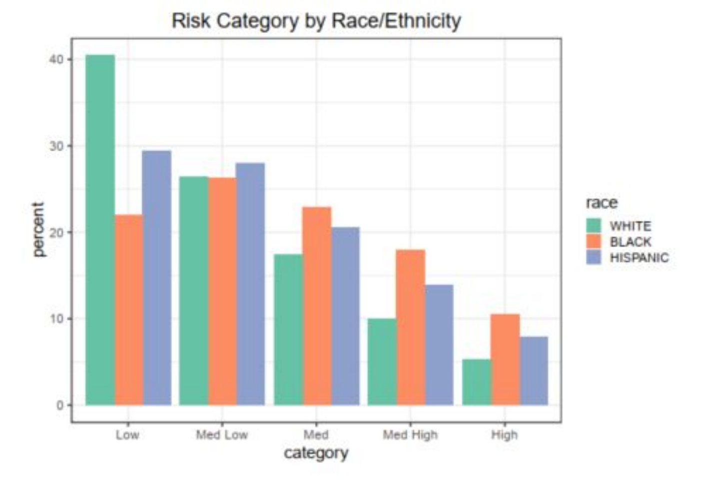


FIGURE 2. Percent of defendants, by race, assigned to each risk group.

- Demographic parity: the rate at which defendants are classified into different risk categories should be equal between all races
- The risk assessment also does not achieve statistical parity by race

What is Predictive Policing

- Broadly: Using mathematical models and other analytical techniques to predict where criminal activity may occur
  - In theory: If it is predicted that 60% of crime will occur in a certain area, a police department is interested in deploying 60% of their active police force to that area to prevent that expected criminal activity
- Machine Learning: algorithms designed to use existing and collected data to make predictions of where crime will occur and where police should be sent on behalf of a police department
  - "...mostly use traditional batch-mode machine learning, where decisions are made and observed results supplement the training data for the next batch."
  - "However, the problem of feedback makes traditional batch learning frameworks both inappropriate and... incorrect." (Runaway Feedback Loops)
  - "Decisions made by the system influence the data that is fed to it in the future. For example, once a decision has been made to patrol a certain neighborhood, crime discovered in that neighborhood will be fed into the training apparatus for the next round of decision-making."

Predictive Policing and Bias in Police-Recorded Data

Notes on data used in the following discussions:

- Synthetic population data collected from US Census
  - Synthetic population: demographically accurate individual level-representation of a real population
- Drug use data collected from National Survey on Drug Use and Health
  - Accurate representation - no incentive for concealment
- The following discussion investigates the effect of police-recorded data on PredPol
  - PredPol - most widely used predictive policing algorithm in the US
  - Algorithm publicly released
  - "While we use the PredPol algorithm in the following demonstration, the broad conclusions we draw are applicable to any predictive policing algorithm that uses unadjusted police records to predict future crime."

Predictive Policing and Bias in Police-Recorded Data

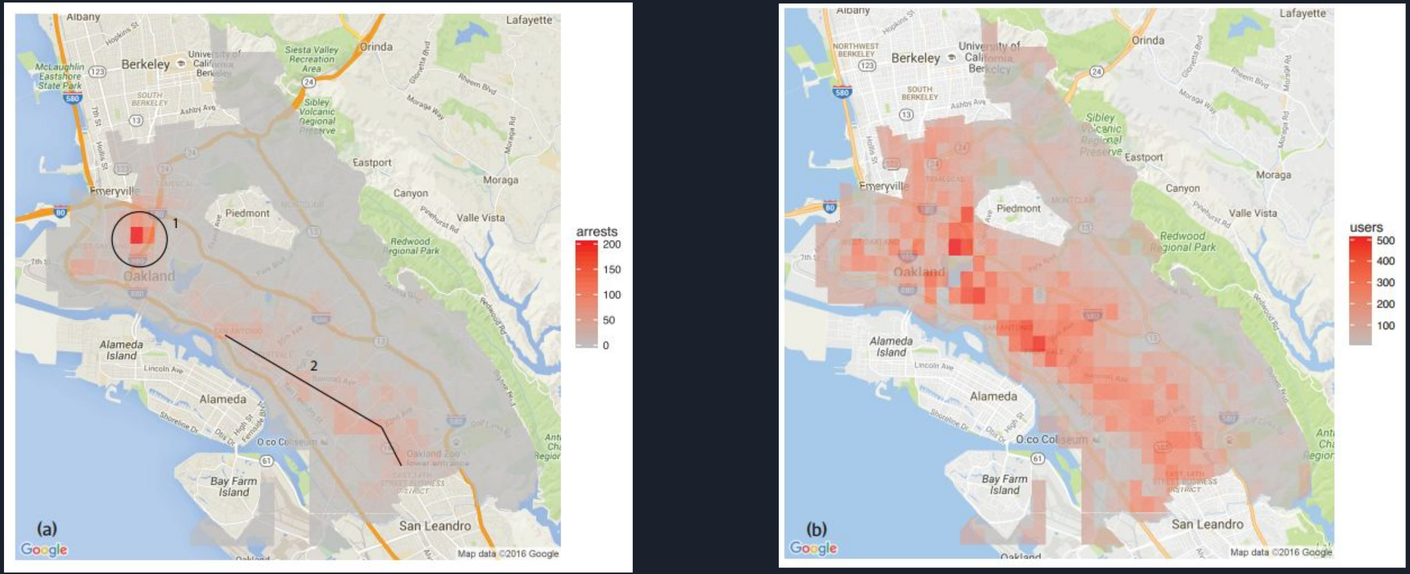


FIGURE 1 (a) Number of drug arrests made by Oakland police department, 2010. (1) West Oakland, (2) International Boulevard. (b) Estimated number of drug users, based on 2011 National Survey on Drug Use and Health

Predictive Policing and Bias in Police-Recorded Data

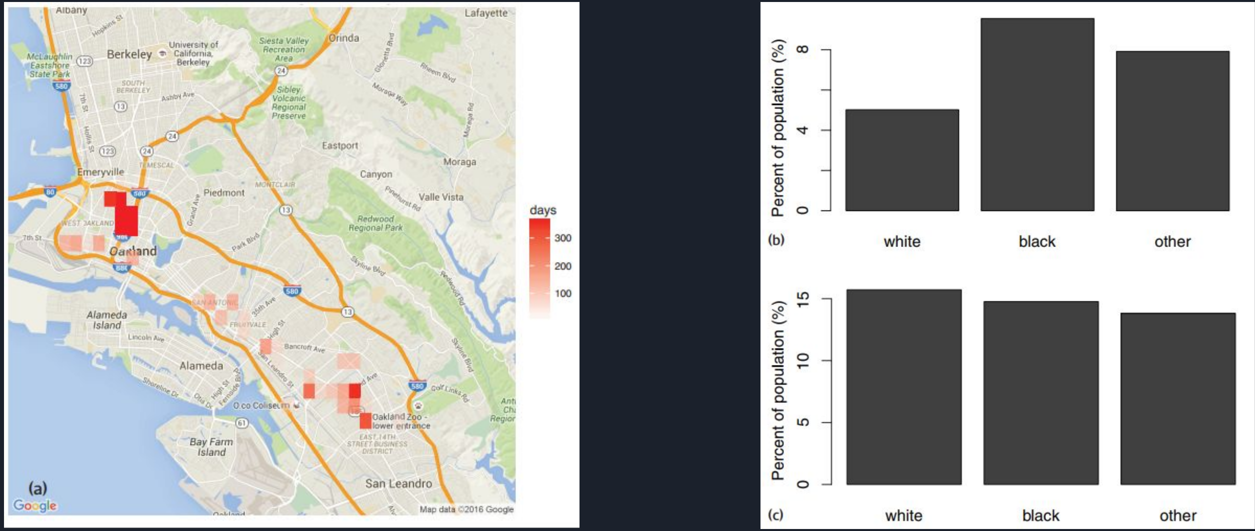


FIGURE 2 (a) Number of days with targeted policing for drug crimes in areas flagged by PredPol analysis of Oakland police data. (b) Targeted policing for drug crimes, by race. (c) Estimated drug use by race

Predictive Policing and Bias in Police-Recorded Data

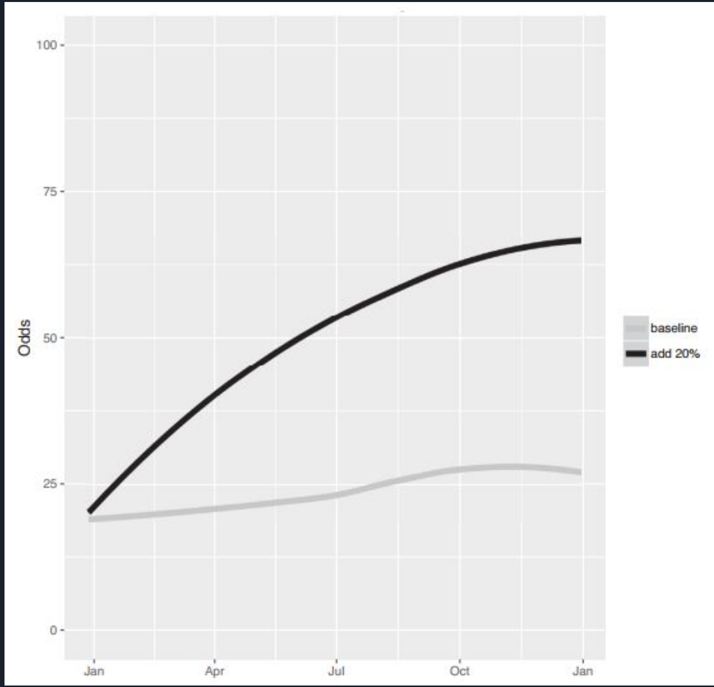


FIGURE 3 Predicted odds of crime in locations targeted by PredPol algorithm, relative to non-targeted locations. 'Baseline' is original Oakland police data. 'Add 20%' simulates the effect of additional crimes being observed in targeted locations

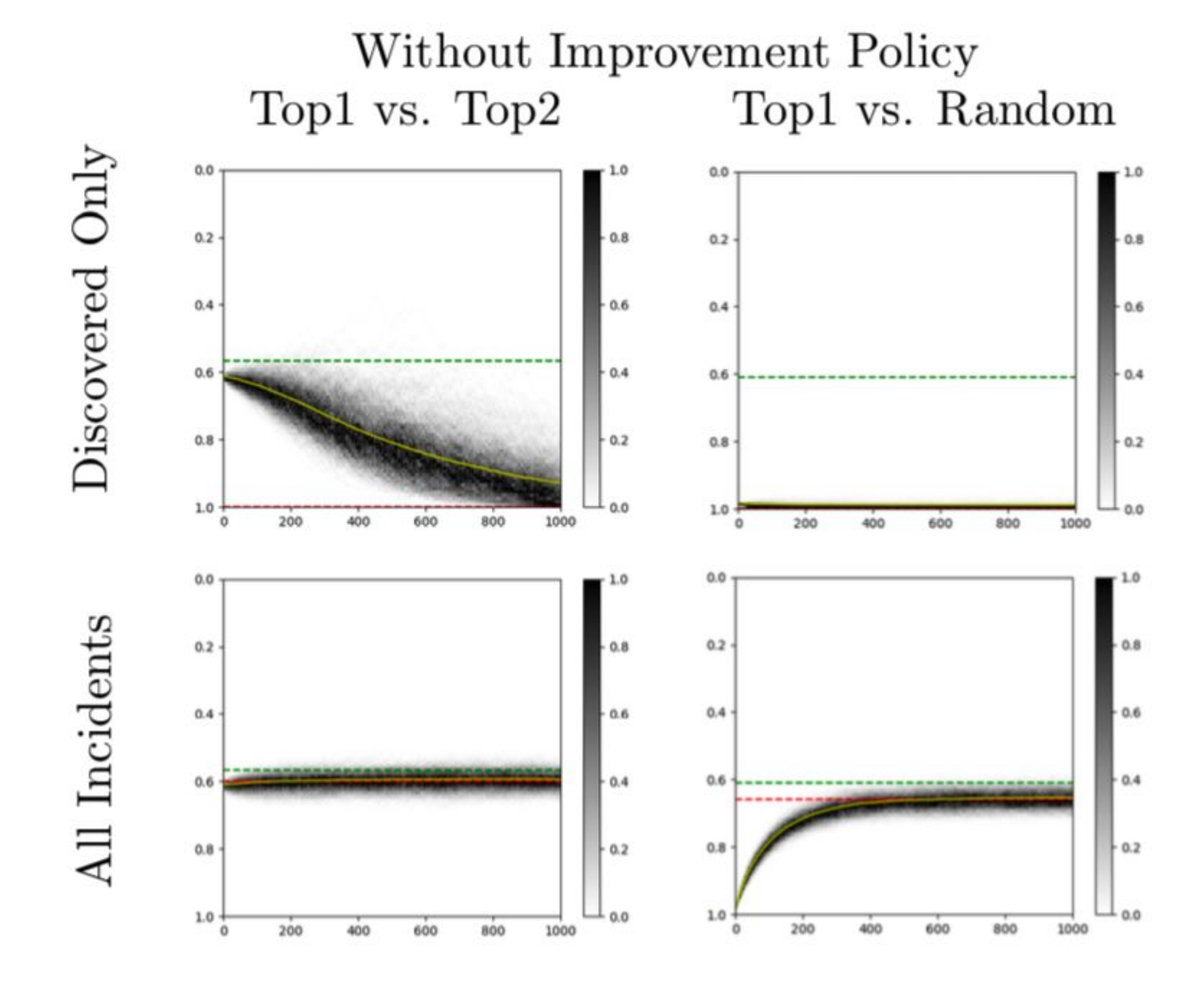
- The more time police spend in a location, the more crime they will find in that location.
- In each location where targeted policing is sent, we increase the number of crimes observed by 20%. These additional simulated crimes then become part of the data set that is fed into PredPol on subsequent days and are factored into future forecasts.
- Results in a runaway feedback loop: the PredPol algorithm becomes increasingly confident that crime is contained in the targeted areas

Runaway Feedback Loops In Predictive Policing

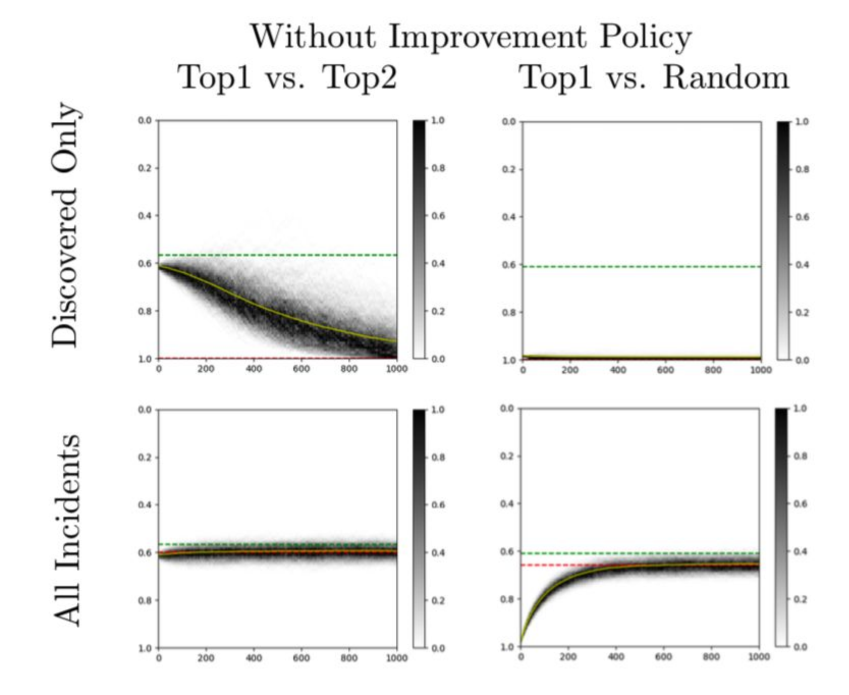
- Paper builds on the work of Lum and Isaac detailed previously
  - Demonstrates the consequences of feedback loops in simulation emulating PredPol
- "PredPol (Mohler et al., 2015) assumes that crimes follow an earthquake aftershock model, so that regions that previously experienced crime are likely to experience crime again, with some decay."
  - This helps explain why the Polya Urn model is representative of the PredPol software
    - "In the basic Pólya urn model, the urn contains x white and y black balls; one ball is drawn randomly from the urn and its color observed; it is then returned in the urn, and an additional ball of the same color is added to the urn, and the selection process is repeated." (https://en.wikipedia.org/wiki/Polya_urn_model)

Runaway Feedback Loops In Predictive Policing

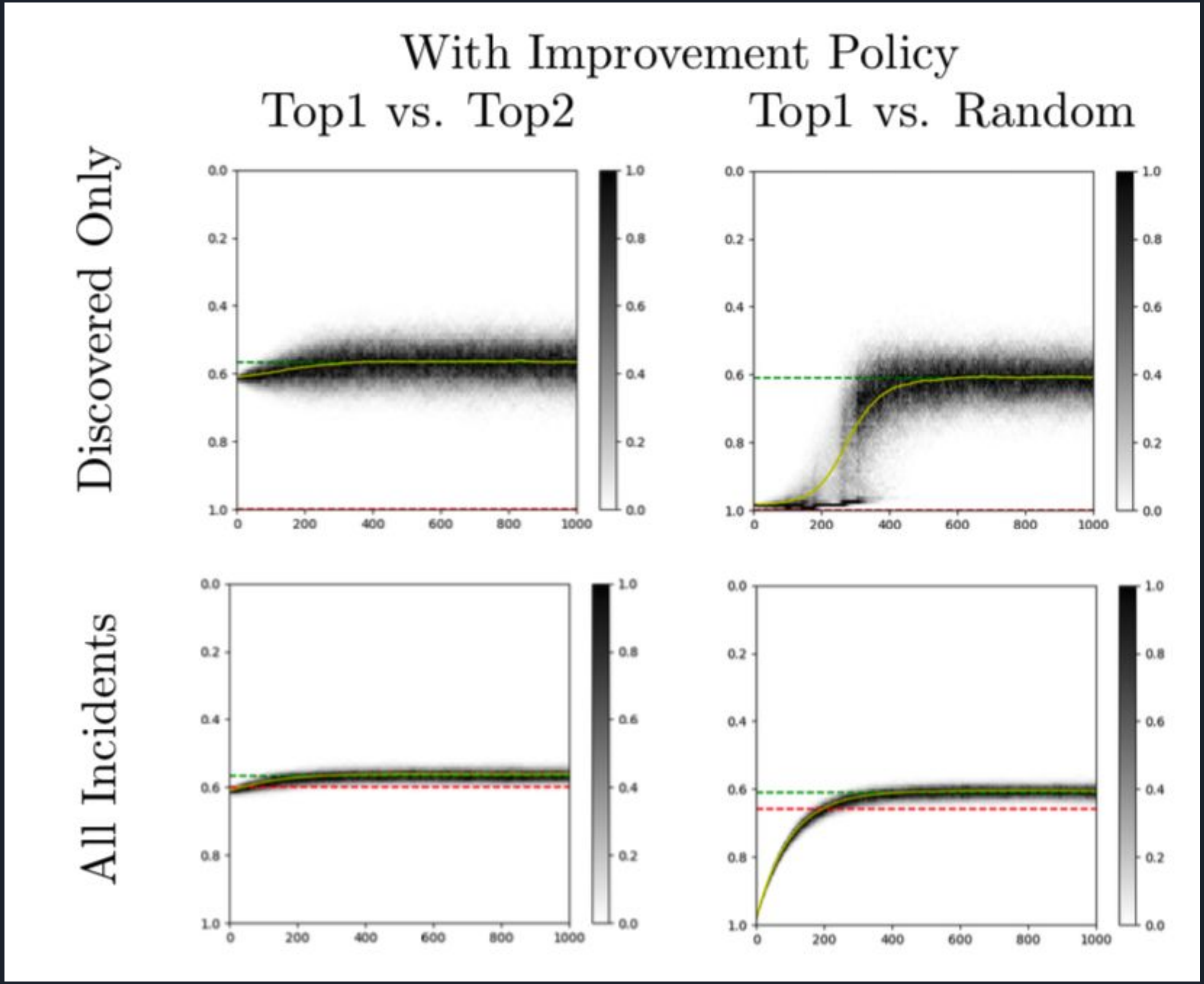
- Developed two urn models based on PredPol
  - No improvement policy
    - Operates as PredPol would operate after training on a data set
  - Improvement policy
    - Authors' attempt to "fix" PredPol and limit feedback loops
      - "Effectively, we want a scheme where as more police are sent, smaller weights are assigned to discovered incidents."


Without Improvement Policy
Top1 vs. Top2
Top1 vs. Random
Discovered Only
All Incidents

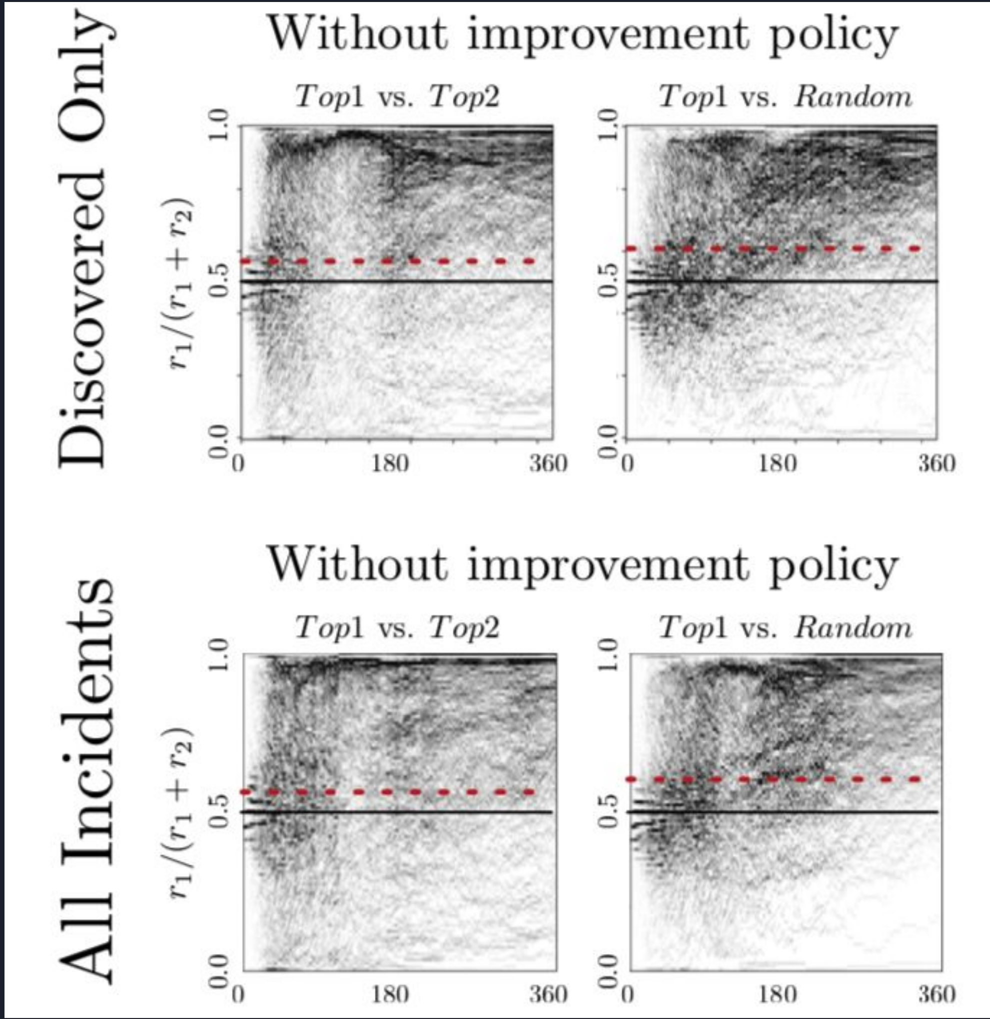
Runaway Feedback Loops In Predictive Policing



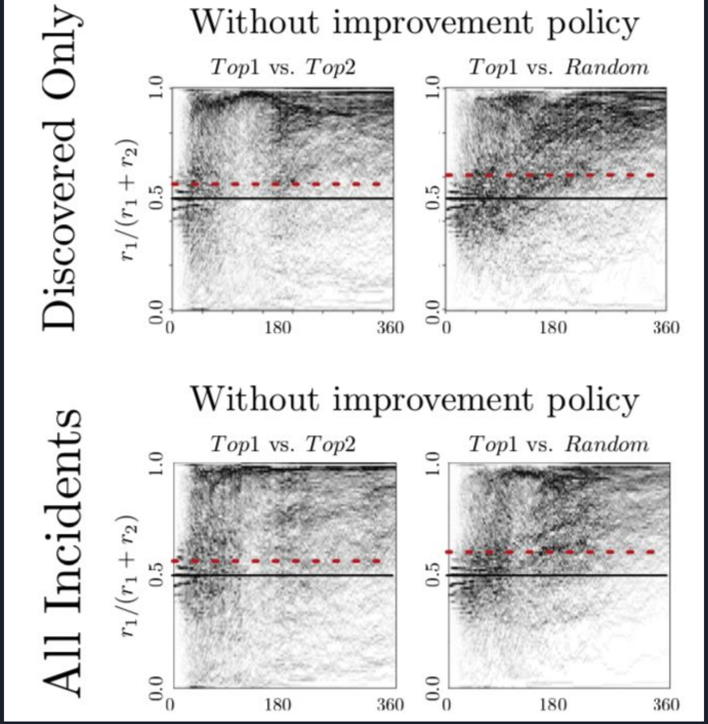
- Distribution over 1000 days vs percentage of balls from region Top1 (based on data from an Oakland neighborhood as in Lum & Isaac) in the urn over 1000 runs
- Historic crime data: Top1 = 609 incidents, Top2 = 379 incidents, Random = 7 incidents
- $\lambda_{Top1} = 3.69$, $\lambda_{Top2} = 2.82$, and $\lambda_{Random} = 2.36$
- Underlying crime rates should see 56.7% of force to Top1 instead of Top2
- 61% to Top1 instead of Random



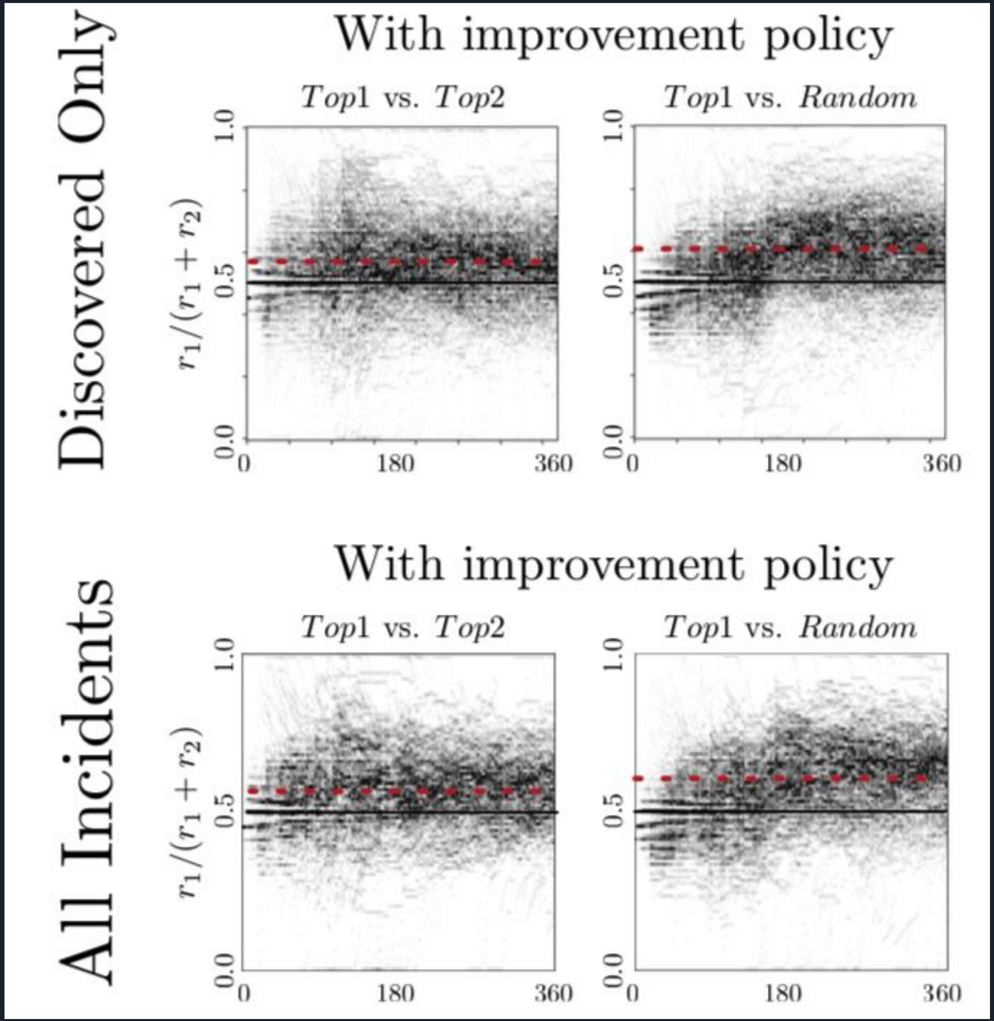
With Improvement Policy
Top1 vs. Top2
Top1 vs. Random
Discovered Only
All Incidents


Discovered Only
Without improvement policy
Top1 vs. Top2
Top1 vs. Random
$r_1/(r_1+r_2)$
1.0
0.5
0.0
0
180
360
All Incidents
Without improvement policy
Top1 vs. Top2
Top1 vs. Random
$r_1/(r_1+r_2)$
1.0
0.5
0.0
0
180
360

Runaway Feedback Loops In Predictive Policing

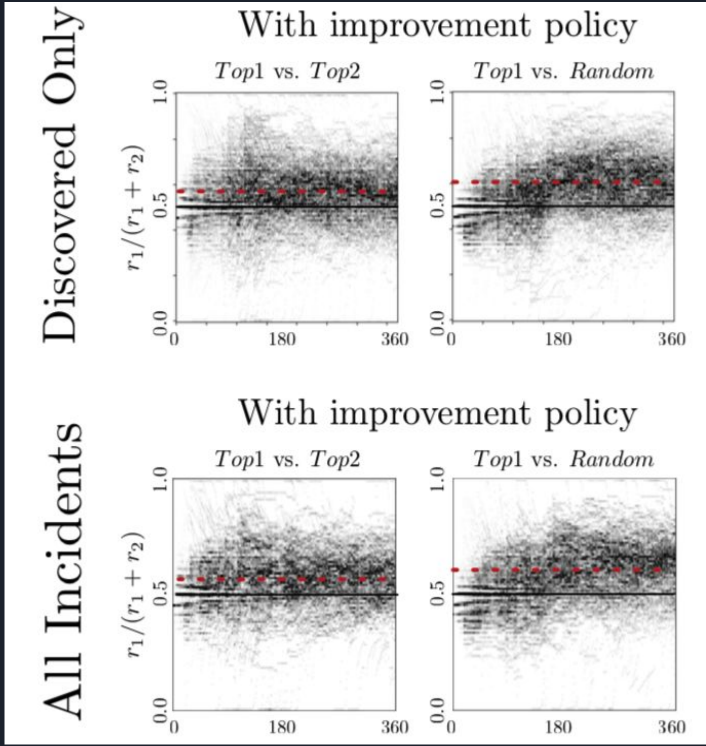


- "At each simulation day, PredPol trains on the previous 180 days of incident data, and pro- duces predicted crime rates r_A and r_B . The decision of where to send police is made probabilistically, by a Bernoulli trial with $p = r_A(r_A+r_B)^{-1}$. This models the targeting effect of sending more police where more crime is expected"



Discovered Only
With improvement policy
$Top1$ vs. $Top2$
$Top1$ vs. $Random$
$r_1/(r_1 + r_2)$
1.0
0.5
0.0
0
180
360
All Incidents
With improvement policy
$Top1$ vs. $Top2$
$Top1$ vs. $Random$
$r_1/(r_1 + r_2)$
1.0
0.5
0.0
0
180
360

Runaway Feedback Loops In Predictive Policing



- The fix is to filter the input as in down-weighting discovered incidents in the simulated urns from before:
  - “Specifically, once we obtain crime report data from the system, we conduct another Bernoulli trial with $p = r_O(r_A + r_B)^{-1}$, where r_O is the predicted rate of the district we did not police that day, and only add the incidents to the training set if the trial succeeds. In other words, the more likely it is that police are sent to a given district, the less likely it is that we should incorporate those discovered incidents.”
- Note that the process is still quite noisy, a further indication that PredPol generates crime rate predictions that are still somewhat unreliable.

Conclusion

- The data used to train the model should not exhibit racial disparity; select training data that is less likely to be generated by biased enforcement
- Runaway Feedback Loops suggests possible fixes to the algorithm
    - Issues remain with the individual decisions of police, and does not address the broader issue of police violence
    - Does not address issues with incarceration, bias in sentencing, or other systemic motivators of crime such as lack of affordable housing or living wage jobs in Black communities
- Larger models still breakdown
    - Not possible in a racist system (concluded in Layers of Bias) - too much bias baked into data

Citations

1. https://www.hsph.harvard.edu/magazine/magazine_article/america-is-failing-its-black-mothers/
2. https://www.nytimes.com/interactive/2018/03/19/upshot/race-class-white-and-black-men.html
3. https://www.sentencingproject.org/publications/color-of-justice-racial-and-ethnic-disparity-in-state-prisons/
4. https://www.pnas.org/content/116/34/16793
5. https://en.wikipedia.org/wiki/Oregon_Department_of_Corrections
6. Layers of Bias: A Unified Approach for Understanding Problems with Risk Assessment - Eckhouse, Lum, Conti-Cook, and Ciccolini, *Criminal Justice and Behavior* (2019)
7. Measures of Fairness For New York City's Supervised Release Risk Assessment Tool - Lum and Shah, *Human Rights Data Analysis Group* (2019)
8. To Predict and Serve? - Lum and Isaac, *Significance Magazine* (2016)
9. Runaway Feedback Loops in Predictive Policing - Ensign, Friedler, Neville, Scheidegger, and Venkatasubramanian, *Proceedings of Machine Learning Research* (2018)