

1 Hacking the Electorate

1.1 Motivations: 2 Common Misconceptions

1. The information fallacy: that campaigns have encyclopedic knowledge of every voter
2. The inconsistency between actual and perceived voter: Actual voters have complicated preferences and policy desires, but the perceived voter is essentially just based on simple demographics.

1.2 Policy Behind Public Data

The main data that campaigns try to obtain includes voter registration, census, and licensing data. Prior to 2002 this data was very localized but after the 2000 election mishap regulations were passed to aggregate this information so that it is more easily accessible.

1.3 Some definitions and goals for the author

Perceived Voter Model: Attempts to determine how and why particular data leads to certain political decisions.

The aspects covered include the campaign, the perceived voter, direct voter contact, etc.

Hypothesis for the book: The availability of public records will influence whether campaigns use geographic level strategy or individual level strategy

Mobilization: If they have party info they will focus on getting people within their party to vote.

1.4 Data Accumulation

In order to aggregate public records, the Democratic party in particular uses the platforms Catalist and NGP VAN.

Catalist claims to have data on 240 million voters and allows for microtargeting of particular groups or demographics.

NGP VAN provides a platform for campaigns to track engagement and manage fundraising.

Campaigns use these databases because they can get lots of data at once that would take a long time to collect themselves, plus these companies do some of their own analysis to increase the effectiveness of the data. They get characteristics about many many voters as well as important aspects to target. This lets them group people into categories of perceived voters.

1.5 Results of Tested Hypothesis

The key was focusing on the perceptions that campaigns had based on public data available.

The two main sources of data used when available was party registration data and primary data. For states with these two datasets, it was possible to create fairly accurate voting predictions. Not all states provide this information, however, so one method for states without this data was using geographic info such as past election results by precincts. This was problematic for mixed-partisan areas as it was hard to get a good sense of which people were affiliated with which party.

An alternative strategy was microtargeting models such as those used by Catalist.

In party registration states, where said info is public, the strategy was to target independent voters directly who have a high chance of voting and getting their party mobilized and voting. This led to less accidental interaction with people not in their party who they cant convince.

In non-party registration states, campaigns instead focus more on persuading undecided voters since they dont know partisanship.

Campaigns Strategies led to downstream effects: Non-party registration states tend to not have correlation between party and turnout because campaigns arent targeting particular parties. On the other hand, in party registrations states, party did correlate to turnout due to the geographic strategies employed.

1.6 Takeaways

Public data policies are affecting how campaigns perceive the people and their strategies.

Perceptions varied drastically across states due to different policy with regards to public data.

1.7 Racialized Engineering

Important context for the book that it was written in 2015 and a lot of the info was before Trumps bid, Cambridge Analytica scandal, mainstream coverage of gerrymandering, and DNC email hack. In addition, all of the analysis was based on the 2008 election, where there was no incumbent candidate.

1.8 Use of Public Record

The author identified 8 states where most voters listed their race on voter registration forms. Authors claim is that campaigns pay more attention to race when that data is available. Some states include race (like Alabama), some dont (like North Carolina) when giving out voter data.

For states that dont include Race, Catalist is able to simply use their platform to predict race for each voter based on names, racial composition of neighborhoods, etc., so they get that information rather accurately regardless. They were more accurate at predicting whiteness than other races, which sometimes led to white voters being more accurately targeted compared to other races.

The downstream effects of this is that, for example, Republicans can ignore voters listed as African-American, or Democrats ignore gun owners as those groups of people are heavily weighted towards one particular partisanship. This leads to non-typical voters getting almost no outreach, reducing their turnout. Ultimately, voter turnout was actually higher for voters with listed races than not because campaigns are targeting their efforts by race.

1.9 Other sources besides public data

Commercial Data is generally not very effective, as it only gives around 6 percent of the information on voter turnout. A model is only as strong as the correlations in the data used to train it. This means that generally public data holds all of the power for influencing campaigns.

Social Networks are actually also found not to be very effective either because most people either don't want to campaign a dissenting opinion to their social groups, or most social groups are homogenous in opinions so campaigning won't have much of an effect. Again we see that public data is the most important factor.

1.10 Normative Questions

Is it good that campaigns collect micro targeting data from administrative databases?
Is micro targeting bad for democracy?

Some notable downsides: Constituent databases are merged with campaign information, so when members of Congress interact with their constituents they are biased in caring more about the needs of constituents that vote for them.

Additionally, politicians are incentivized to pass laws that make it easier for them to do electioneering rather than laws that are the best for the country.

Some pros of microtargeting: Campaigns can connect with the electorate and know what voters care about most and encourage them to go vote.

Some additional cons: Politicians can flat out ignore large portions of the electorate that will not be convinced to vote for them. This creates a divide in the country as everyone only talks to people of the same political preferences resulting in political echo chambers.

Ultimately, microtargeting has its benefits but there are a lot of problems with it.

1.11 Solutions

Campaigns should release their databases semipublicly so that people can decide what information campaigns have on them. This helps increase the benefits and decrease the downsides.

They can't release the whole database because then there is no incentive to make a good model because releasing it would just give your opponents the same info, hence why people should only be able to look up their own data.

This would give more transparency between electorate and campaigns and allow for more voter interaction with the campaign process.

2 Automating Inequality

2.1 About the Author

Virginia Eubanks is an associate professor of political science at the University of Albany, SUNY

She has written 3 books including Automating Inequality.

2.2 Automating Inequality Introduction

Focuses on the transition of various governmental assistance programs from solely human decision making to using predictive models and algorithms.

These algorithms are less flexible.

2.3 History

In the Mid 1800s many cities constructed poorhouses that were meant to house the elderly, disabled, mentally ill, and other disadvantaged peoples. She calls the data stores of individuals who cannot access welfare as being put in "digital poorhouses"

Throughout history of welfare there are attempts to separate the deserving and undeserving poor. She makes the case that algorithmic models however amplify the issues already existing with human decision making.

2.4 Case Study 1: Automating Public Benefits enrollment in Indiana

The government teamed up with IBM to create an algorithm to automate benefit enrollments. The algorithm had a 12.2 percent false negative rate, up from human decision making.

Stipes family - Child with developmental delays gets denied medicaid after years of care.

Lindsay Kidwell - Mother who submitted all documents appealed decision and won

Omega Young - Missed appointment to recertify Medicaid, received 10,000 dollar medical bill, won appeal the day after she died.

This case study highlights failures in algorithmic decision making due to lack of human interaction, inaccessibility to computers, and a high error rate.

IBM was ultimately fined 70MM+

2.5 Case Study 2: Coordinated Entry System

In 2013 the city of Los Angeles launched Home for Good which aimed to address the mismatch between housing supply and demand. Its underlying philosophy is that it prioritizes chronic homelessness over crisis homelessness. It ranked the homeless on a scale from 1 to 17 based on "vulnerability."

The issues with CES was that it didn't do anything to address the underlying issue of not having enough housing being available compared to demand. Additionally, there was low test-retest reliability. Finally, politicians and homeowners did not want to build more housing despite wanting the homeless problem solved.

2.6 Case Study 3: Predicting Risk for Child Mistreatment in Allegheny County, PA

In the original system, a human case worker would be called and attempt to assess the situation based on public records. Model was created to help decided whether or not the family should or should not be investigated, after which a case worker would step in. This was made in order to expedite the process.

In the model, the poor are more likely to be surveilled than the middle class. As a result, the standard for their parenting is much higher. This is because the poor are more likely to go to public services rather than private providers. There is also evidence to suggest that the model's risk scores started dictating worker actions rather than being used for its original purpose.

2.7 Conclusion

Dr. Eubanks concludes by saying that while we have made strides in many other areas of human endeavor such as technological advancement, we have lagged behind in the area of inequality and human rights. And in some ways, this aforementioned technological advancement has deepened inequality through automation. Finally, she closes by offering areas where the conversation should move to in the future such as UBI and other public assistance programs.