Computational Sociology

Online experiments and surveys

Dr. Thomas Davidson

Rutgers University

February 22, 2024

Plan

1. Course updates
2. Online experiments
3. Online surveys

Course updates

- Homework 2 released
  - Due 3/1 at 5pm EST
  - Spotify developer API credentials required

Course updates

- Project proposal due 3/8 at 5pm
  - *Schedule meeting to discuss your project idea before submitting*
  - Submission instructions to follow
- 3-4 pages double-spaced
  - Introduction/motivation
  - Relevant literature
  - Data and collection strategy
  - Tentative methodology
  - References

Online experiments and surveys

Adapting methods to the digital era

- Experiments and surveys are two mainstays of the social sciences
- Both were developed to study people in labs and in the field
- How does the internet and technological advancement create new opportunities for methodological innovation?

Online experiments

Motivation for online experiments

- Lab experiments provide control but little realism (*low external validity*)
  - e.g. Undergraduate students do not represent wider populations
- Field experiments provided realism but little control (*low internal validity*)
  - e.g. Many factors may affect internal validity
- Digital field experiments can provided both, at scale

Online experiments

Methods: Internal experiments

- Companies and other actors experiment internally
  - A/B tests used to test different user-interface and product differences
- The vast majority of these experiments are private, but some are published by researchers.
- Researchers recently made an entire archive of thousands of experiments available, see the Upworthy Research Archive

Online experiments

The Emotional Contagion Study

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer[a,1], Jamie E. Guillory[b,2], and Jeffrey T. Hancock[b,c]

[a]Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of [b]Communication and [c]Information Science, Cornell University, Ithaca, NY 14853

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

Emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. Emotional contagion is well established in laboratory experiments, with people transferring positive and negative emotions to others. Data from a large real-world social network, collected over a 20-y period suggests that longer-lasting moods (e.g., depression, happiness) can be transferred through networks [Fowler JH, Christakis NA (2008) *BMJ* 337:a2338], although the results are controversial. In an experiment with people who use Facebook, we test whether emotional contagion occurs outside of in-person interaction between individuals by reducing the amount of emotional content in the News Feed. When positive expressions were reduced, people produced fewer positive posts and more negative posts; when negative expressions were reduced, the opposite pattern occurred. These results indicate that emotions expressed by others on Facebook influence our own emotions, constituting experimental evidence for massive-scale contagion via social networks. This work also suggests that, in contrast to prevailing assumptions, in-person interaction and nonverbal cues are not strictly necessary for emotional contagion, and that the observation of others' positive experiences constitutes a positive experience for people.

computer-mediated communication | social media | big data

demonstrated that (*i*) emotional contagion occurs via text-based computer-mediated communication (7); (*ii*) contagion of psychological and physiological qualities has been suggested based on correlational data for social networks generally (7, 8); and (*iii*) people's emotional expressions on Facebook predict friends' emotional expressions, even days later (7) (although some shared experiences may in fact last several days). To date, however, there is no experimental evidence that emotions or moods are contagious in the absence of direct interaction between experiencer and target.

On Facebook, people frequently express emotions, which are later seen by their friends via Facebook's "News Feed" product (8). Because people's friends frequently produce much more content than one person can view, the News Feed filters posts, stories, and activities undertaken by friends. News Feed is the primary manner by which people see content that friends share. Which content is shown or omitted in the News Feed is determined via a ranking algorithm that Facebook continually develops and tests in the interest of showing viewers the content they will find most relevant and engaging. One such test is reported in this study: A test of whether posts with emotional content are more engaging.

The experiment manipulated the extent to which people (N = 689,003) were exposed to emotional expressions in their News Feed. This tested whether exposure to emotions led people to

Online experiments

Design and results

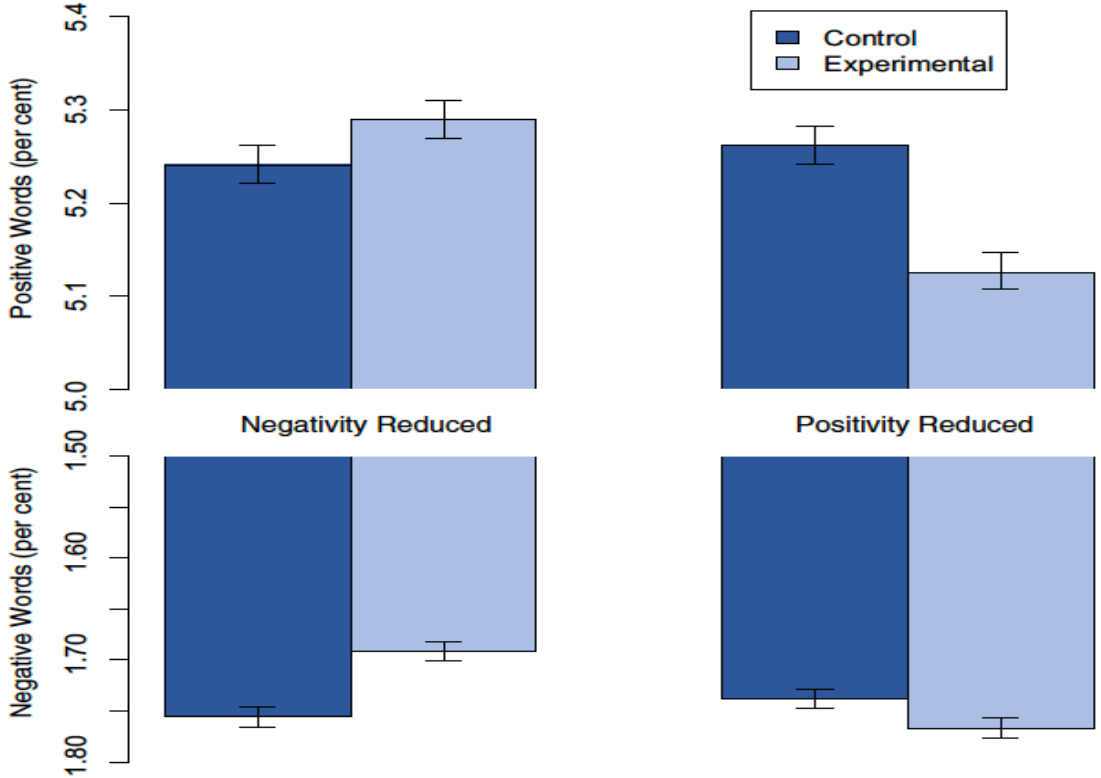


Fig. 1. Mean number of positive (*Upper*) and negative (*Lower*) emotion words (percent) generated people, by condition. Bars represent standard errors.

Online experiments

Reactions

https://www.cbsnews.com/news/controversial-facebook-emotion-study-journal-responds/

Online experiments

Reactions

Editorial Expression of Concern and Correction

PSYCHOLOGICAL AND COGNITIVE SCIENCES

PNAS is publishing an Editorial Expression of Concern regarding the following article: "Experimental evidence of massive-scale emotional contagion through social networks," by Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock, which appeared in issue 24, June 17, 2014, of *Proc Natl Acad Sci USA* (111:8788–8790; first published June 2, 2014; 10.1073/pnas.1320040111). This paper represents an important and emerging area of social science research that needs to be approached with sensitivity and with vigilance regarding personal privacy issues.

Questions have been raised about the principles of informed consent and opportunity to opt out in connection with the research in this paper. The authors noted in their paper, "[The work] was consistent with Facebook's Data Use Policy, to which all users agree prior to creating an account on Facebook, constituting informed consent for this research." When the authors prepared their paper for publication in PNAS, they stated that: "Because this experiment was conducted by Facebook, Inc. for internal purposes, the Cornell University IRB [Institutional Review Board] determined that the project did not fall under Cornell's Human Research Protection Program." This statement has since been confirmed by Cornell University.

Obtaining informed consent and allowing participants to opt out are best practices in most instances under the US Department of Health and Human Services Policy for the Protection of Human Research Subjects (the "Common Rule"). Adherence to the Common Rule is PNAS policy, but as a private company Facebook was under no obligation to conform to the provisions of the Common Rule when it collected the data used by the authors, and the Common Rule does not preclude their use of the data. Based on the information provided by the authors, PNAS editors deemed it appropriate to publish the paper. It is nevertheless a matter of concern that the collection of the data by Facebook may have involved practices that were not fully consistent with the principles of obtaining informed consent and allowing participants to opt out.

Inder M. Verma
Editor-in-Chief

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Correction for "Experimental evidence of massive-scale emotional contagion through social networks," by Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock, which appeared in issue 24, June 17, 2014, of *Proc Natl Acad Sci USA* (111:8788–8790; first published June 2, 2014; 10.1073/pnas.1320040111).

The authors note that, "At the time of the study, the middle author, Jamie E. Guillory, was a graduate student at Cornell University under the tutelage of senior author Jeffrey T. Hancock, also of Cornell University (Guillory is now a postdoctoral fellow at Center for Tobacco Control Research and Education, University of California, San Francisco, CA 94143)." The author and affiliation lines have been updated to reflect the above changes and a present address footnote has been added. The online version has been corrected.

The corrected author and affiliation lines appear below.

Adam D. I. Kramer[a,1], Jamie E. Guillory[b,2], and Jeffrey T. Hancock[b,c]

[a]Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of [b]Communication and [c]Information Science, Cornell University, Ithaca, NY 14853

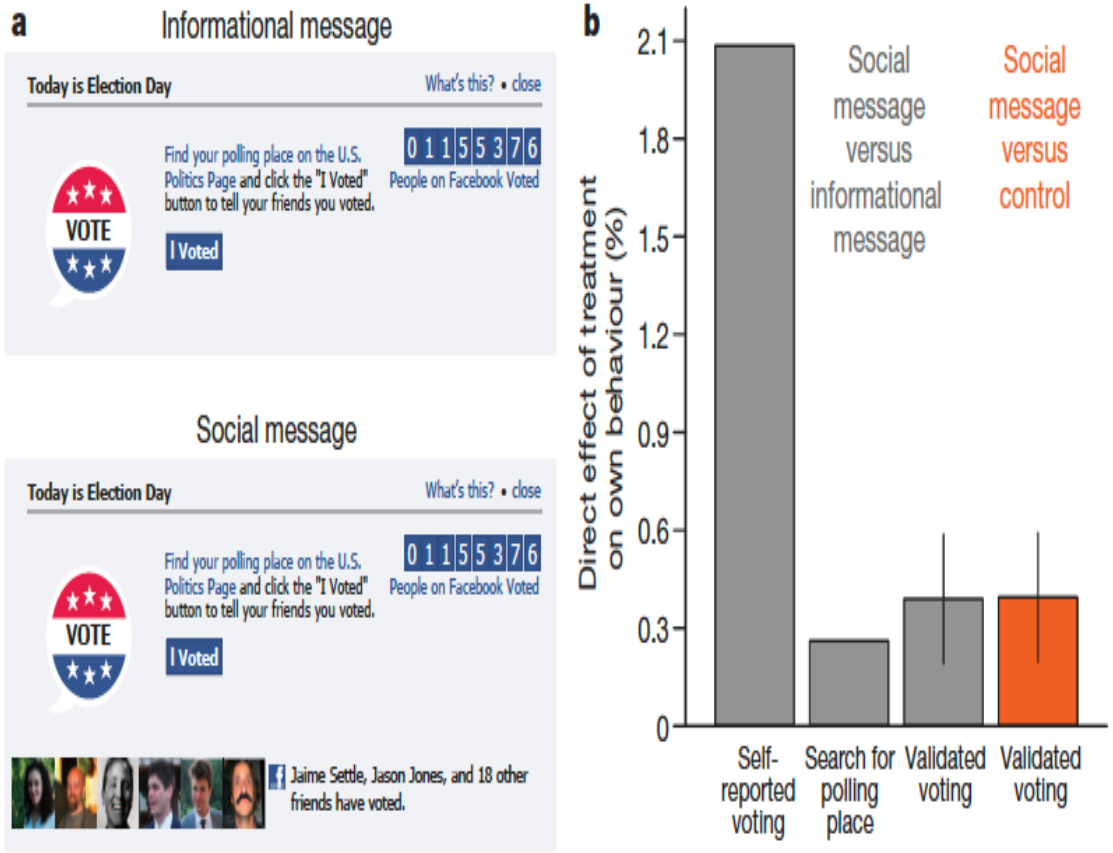
[1]To whom correspondence should be addressed. Email: akramer@fb.com.

[2]Present address: Center for Tobacco Control Research and Education, University of California, San Francisco, CA 94143.

www.pnas.org/cgi/doi/10.1073/pnas.1412583111

Online experiments

Facebook and voter turnout



Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489 (7415): 295–98. https://doi.org/10.1038/nature11421.

Online experiments

New developments in platform research

SOCIAL MEDIA

How do social media feed algorithms affect attitudes and behavior in an election campaign?

Andrew M. Guess[1]*, Neil Malhotra[2], Jennifer Pan[3], Pablo Barberá[4], Hunt Allcott[5], Taylor Brown[4], Adriana Crespo-Tenorio[4], Drew Dimmery[4,6], Deen Freelon[7], Matthew Gentzkow[8], Sandra González-Bailón[9], Edward Kennedy[10], Young Mie Kim[11], David Lazer[12], Devra Moehler[4], Brendan Nyhan[13], Carlos Velasco Rivera[4], Jaime Settle[14], Daniel Robert Thomas[4], Emily Thorson[15], Rebekah Tromble[16], Arjun Wilkins[4], Magdalena Wojcieszak[17,18], Beixian Xiong[4], Chad Kiewiet de Jonge[4], Annie Franco[4], Winter Mason[4], Natalie Jomini Stroud[19], Joshua A. Tucker[20]

We investigated the effects of Facebook's and Instagram's feed algorithms during the 2020 US election. We assigned a sample of consenting users to reverse-chronologically-ordered feeds instead of the default algorithms. Moving users out of algorithmic feeds substantially decreased the time they spent on the platforms and their activity. The chronological feed also affected exposure to content: The amount of political and untrustworthy content they saw increased on both platforms, the amount of content classified as uncivil or containing slur words they saw decreased on Facebook, and the amount of content from moderate friends and sources with ideologically mixed audiences they saw increased on Facebook. Despite these substantial changes in users' on-platform experience, the chronological feed did not significantly alter levels of issue polarization, affective polarization, political knowledge, or other key attitudes during the 3-month study period.

Online experiments

Methods: Using existing environments

- Researchers can use platforms to create their own experiments
  - e.g. Doleac and Stein (2013) used different pictures on Craigslist to measure discrimination
  - e.g. van de Rijt et al. (2014) randomly donated to Kickstarters, upvoted reviews, awarded Wikipedia contributers, and signed petitions to study the Matthew Effect
  - e.g. Munger (2017) used a Twitter "bot" to measure the effect of sanctions on racial harassment

Online experiments

Countering hate speech on Twitter

Polit Behav
DOI 10.1007/s11109-016-9373-5

ORIGINAL PAPER

Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment

Kevin Munger[1]

Online experiments

Design and experimental manipulation

Fig. 3 Treatments. **a** The treatment—black bot. **b** The bot applying the treatment—white bot

Online experiments

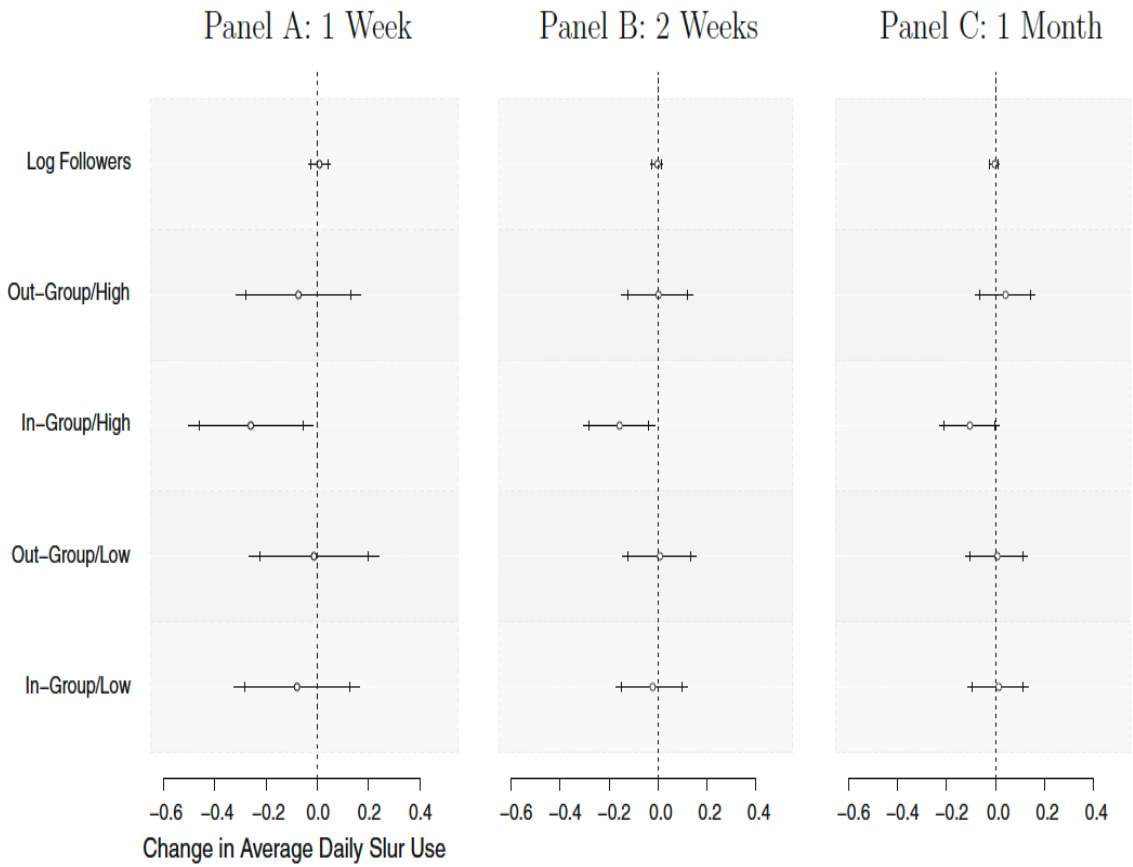
Hypotheses

Table 1 Experimental design and hypothesized effect sizes

	In-group	Out-group
Low followers	Medium effect	Small effect
High followers	Large effect	Medium effect

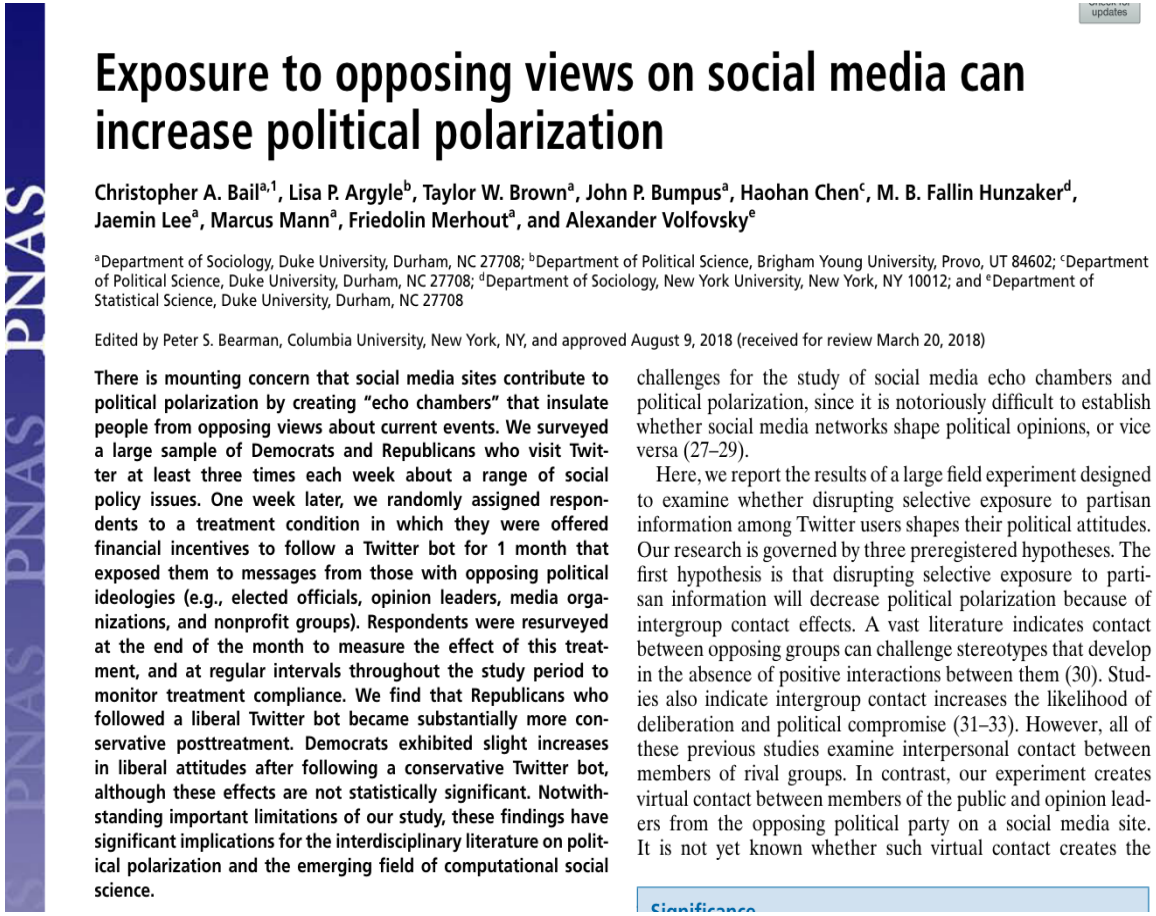
Online experiments

Results



Online experiments

Studying polarization using bots



PNAS PNAS PNAS

Exposure to opposing views on social media can increase political polarization

Christopher A. Bail[a,1], Lisa P. Argyle[b], Taylor W. Brown[a], John P. Bumpus[a], Haohan Chen[c], M. B. Fallin Hunzaker[d], Jaemin Lee[a], Marcus Mann[a], Friedolin Merhout[a], and Alexander Volfovsky[e]

[a]Department of Sociology, Duke University, Durham, NC 27708; [b]Department of Political Science, Brigham Young University, Provo, UT 84602; [c]Department of Political Science, Duke University, Durham, NC 27708; [d]Department of Sociology, New York University, New York, NY 10012; and [e]Department of Statistical Science, Duke University, Durham, NC 27708

Edited by Peter S. Bearman, Columbia University, New York, NY, and approved August 9, 2018 (received for review March 20, 2018)

There is mounting concern that social media sites contribute to political polarization by creating "echo chambers" that insulate people from opposing views about current events. We surveyed a large sample of Democrats and Republicans who visit Twitter at least three times each week about a range of social policy issues. One week later, we randomly assigned respondents to a treatment condition in which they were offered financial incentives to follow a Twitter bot for 1 month that exposed them to messages from those with opposing political ideologies (e.g., elected officials, opinion leaders, media organizations, and nonprofit groups). Respondents were resurveyed at the end of the month to measure the effect of this treatment, and at regular intervals throughout the study period to monitor treatment compliance. We find that Republicans who followed a liberal Twitter bot became substantially more conservative posttreatment. Democrats exhibited slight increases in liberal attitudes after following a conservative Twitter bot, although these effects are not statistically significant. Notwithstanding important limitations of our study, these findings have significant implications for the interdisciplinary literature on political polarization and the emerging field of computational social science.

challenges for the study of social media echo chambers and political polarization, since it is notoriously difficult to establish whether social media networks shape political opinions, or vice versa (27–29).

Here, we report the results of a large field experiment designed to examine whether disrupting selective exposure to partisan information among Twitter users shapes their political attitudes. Our research is governed by three preregistered hypotheses. The first hypothesis is that disrupting selective exposure to partisan information will decrease political polarization because of intergroup contact effects. A vast literature indicates contact between opposing groups can challenge stereotypes that develop in the absence of positive interactions between them (30). Studies also indicate intergroup contact increases the likelihood of deliberation and political compromise (31–33). However, all of these previous studies examine interpersonal contact between members of rival groups. In contrast, our experiment creates virtual contact between members of the public and opinion leaders from the opposing political party on a social media site. It is not yet known whether such virtual contact creates the

Significance

Online experiments

Studying polarization using bots

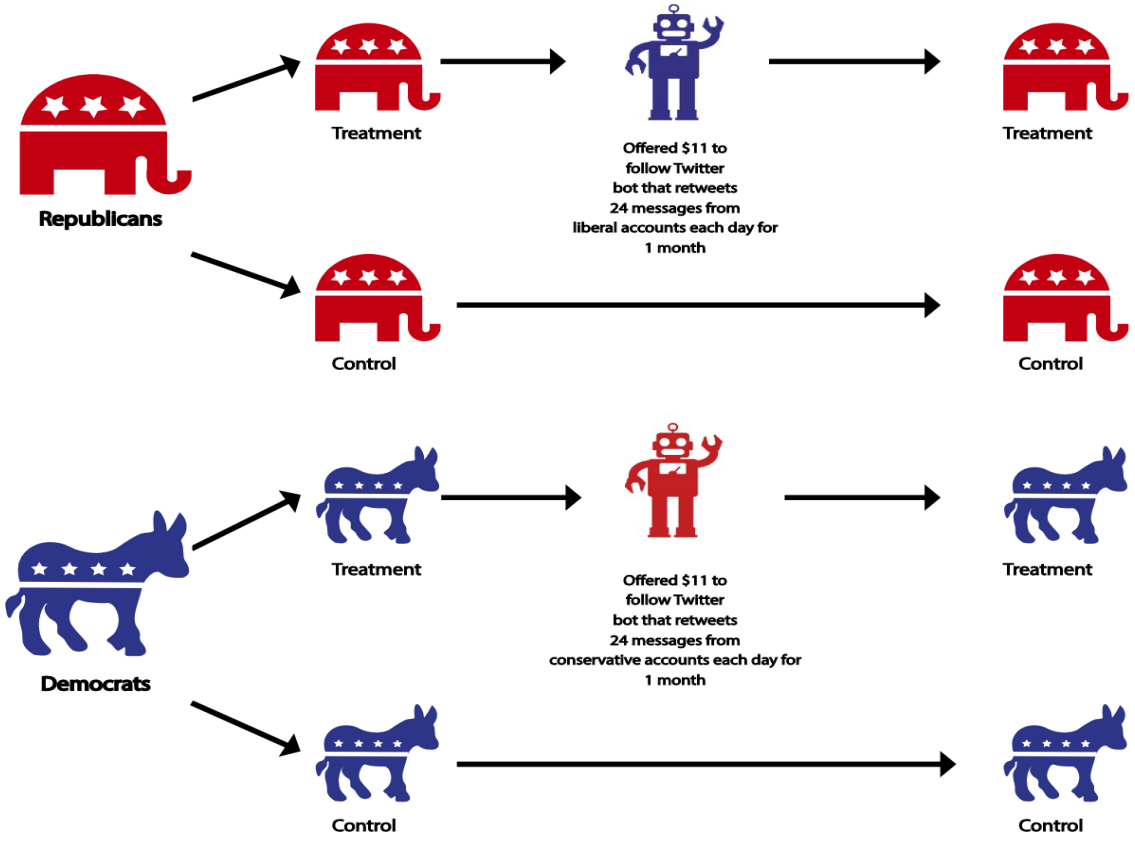


Fig. 1. Overview of research design.

Online experiments

Studying polarization using bots

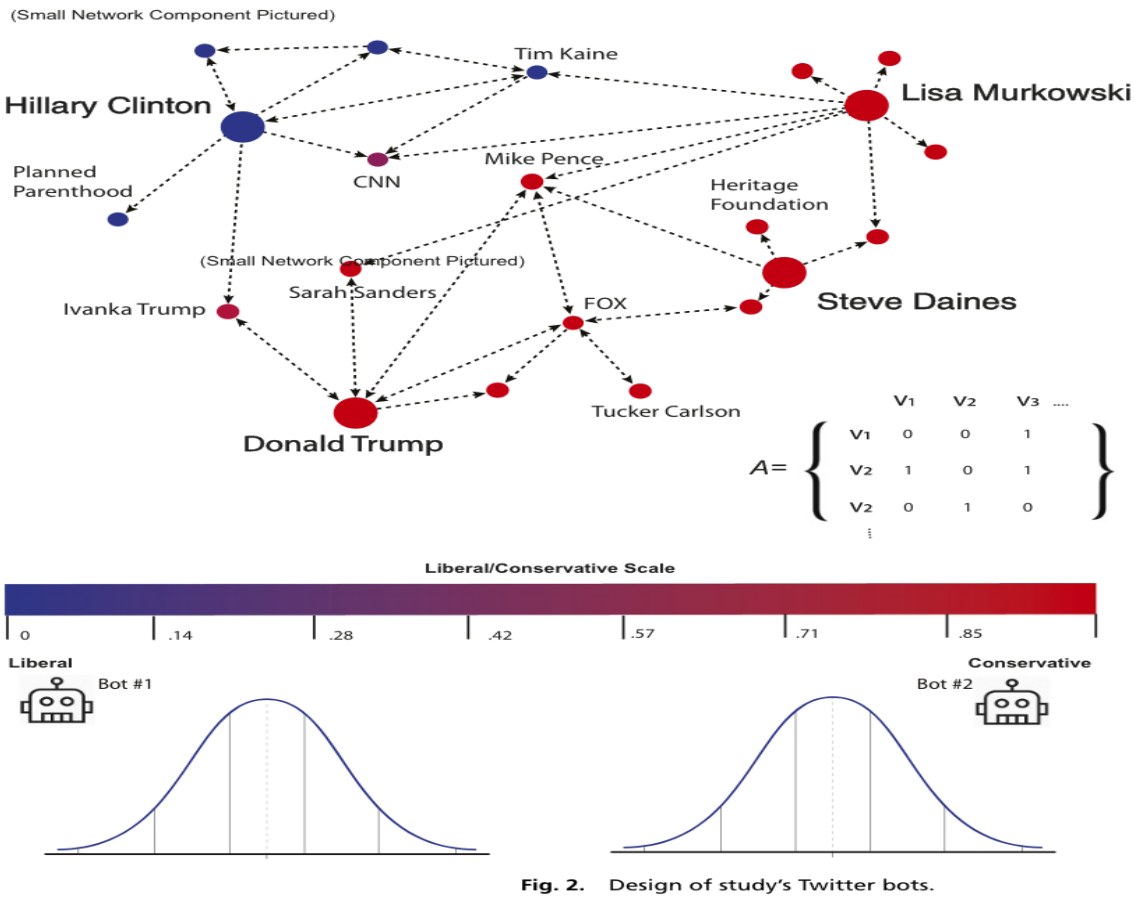
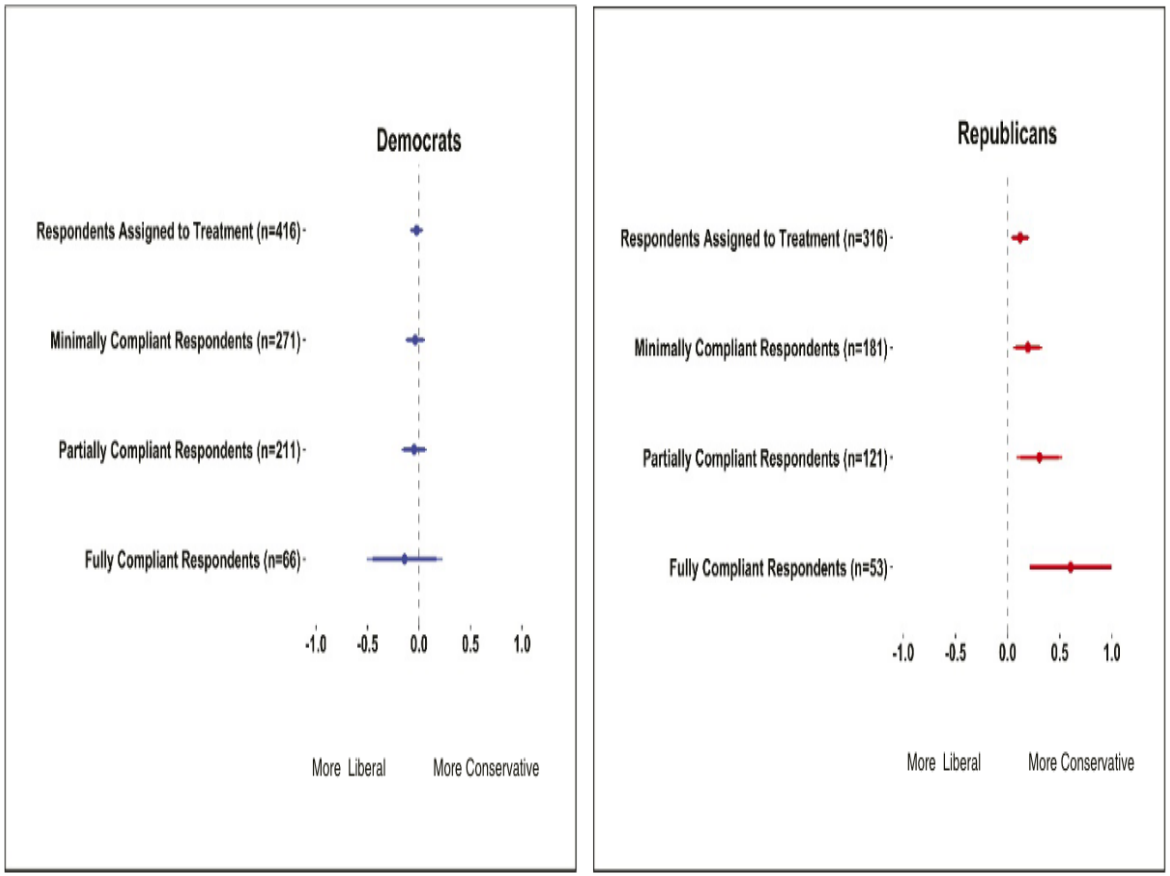


Fig. 2. Design of study's Twitter bots.

Online experiments

Studying polarization using bots



Online experiments

Methods: Digital labs

- Create a virtual environment, fully controlled by the researcher
- High-cost fixed costs associated with developing a platform
- Zero variable cost experiments
  - But incentivizing participation is challenging

Online experiments

The Music Lab Study

Two on Culture

Social Psychology Quarterly
2008, Vol. 71, No. 4, 338–355

Leading the Herd Astray: An Experimental Study of Self-fulfilling Prophecies in an Artificial Cultural Market

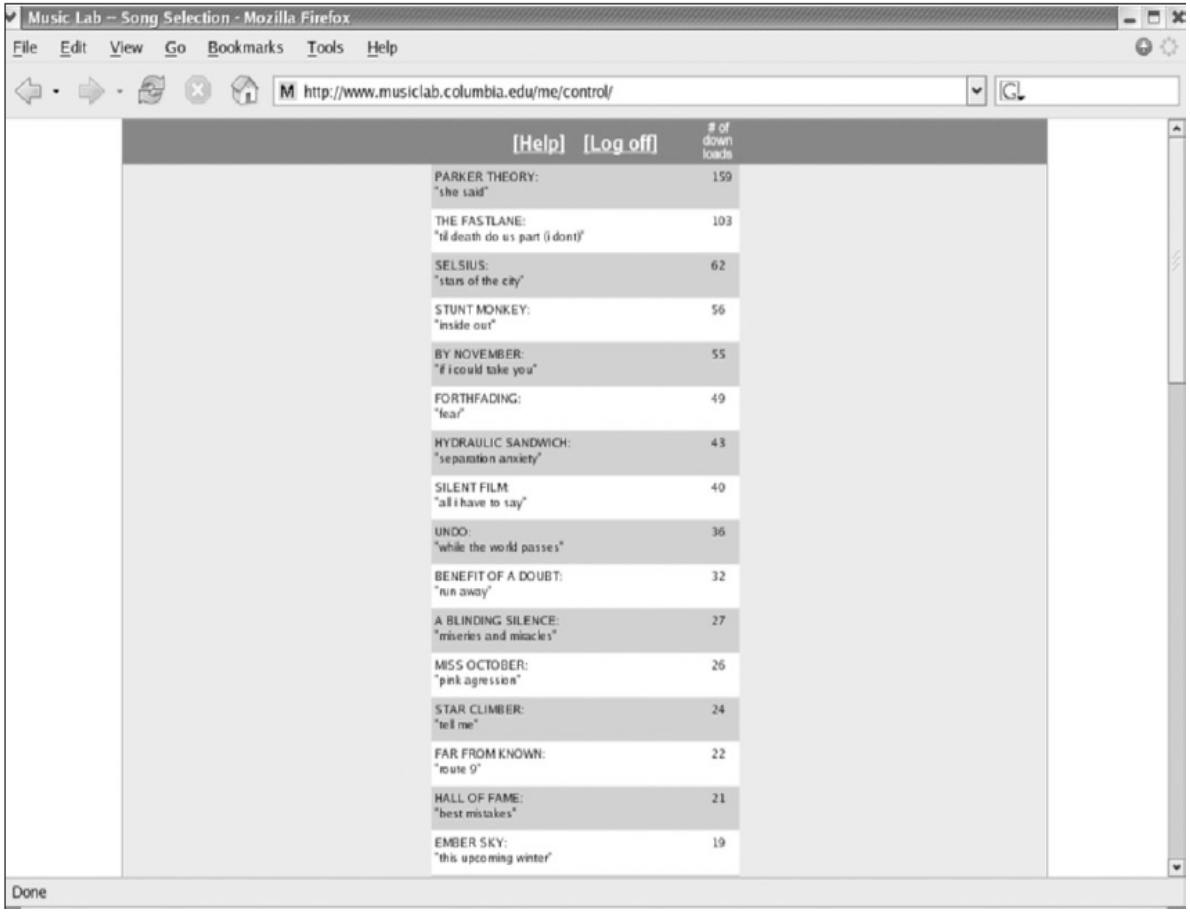
MATTHEW J. SALGANIK
Princeton University

DUNCAN J. WATTS
Yahoo! Research and Columbia University

Individuals influence each others' decisions about cultural products such as songs, books, and movies; but to what extent can the perception of success become a "self-fulfilling prophecy"? We have explored this question experimentally by artificially inverting the true popularity of songs in an online "music market," in which 12,207 participants listened to and downloaded songs by unknown bands. We found that most songs experienced self-fulfilling prophecies, in which perceived—but initially false—popularity became real over time. We also found, however, that the inversion was not self-fulfilling for the market as a whole, in part because the very best songs recovered their popularity in the long run. Moreover, the distortion of market information reduced the correlation between appeal and popularity, and led to fewer overall downloads. These results, although partial and speculative, suggest a new approach to the study of cultural markets, and indicate the potential of web-based experiments to explore the social psychological origin of other macrosociological phenomena.

Online experiments

The Music Lab Study



Online experiments

The Music Lab Study

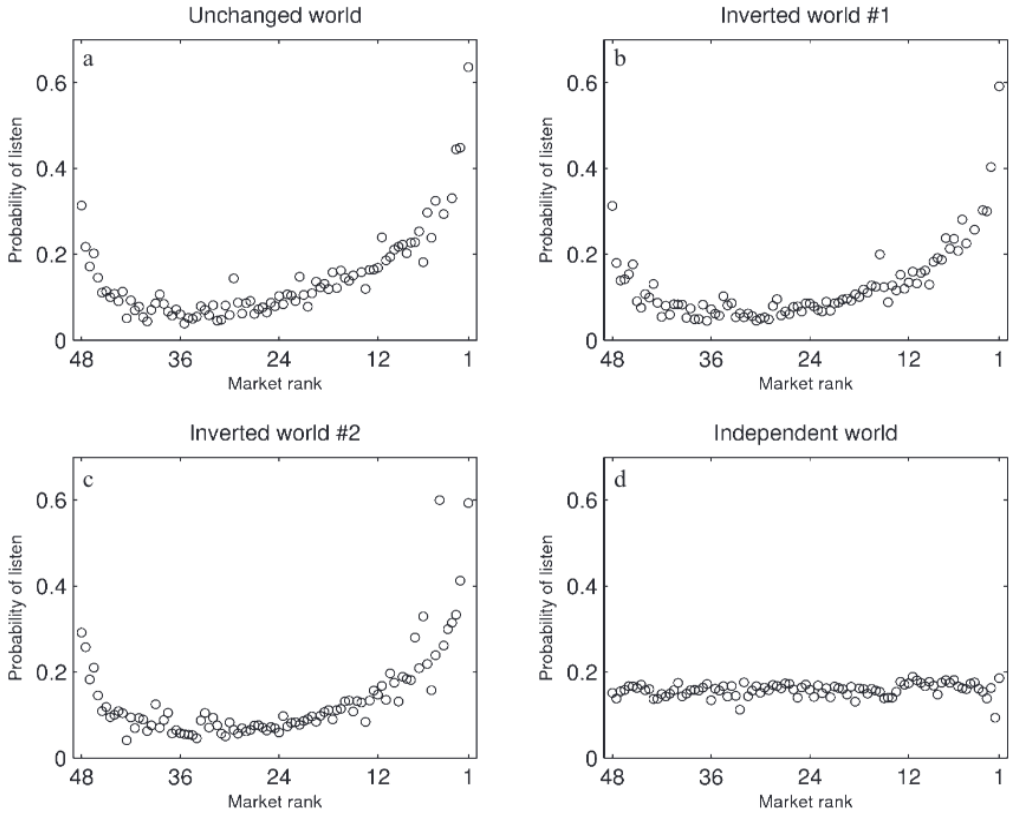


Figure 4. Probability of Listening to a Song as a Function of its Popularity in the Four Worlds

Online experiments

The Music Lab Study

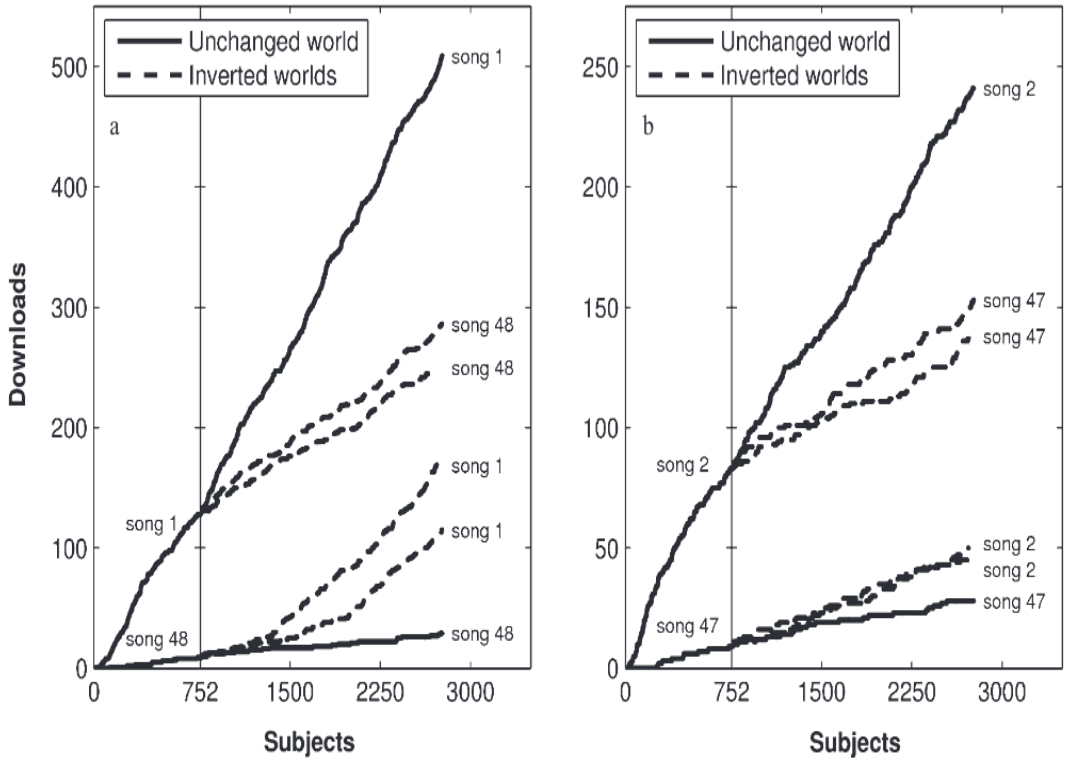
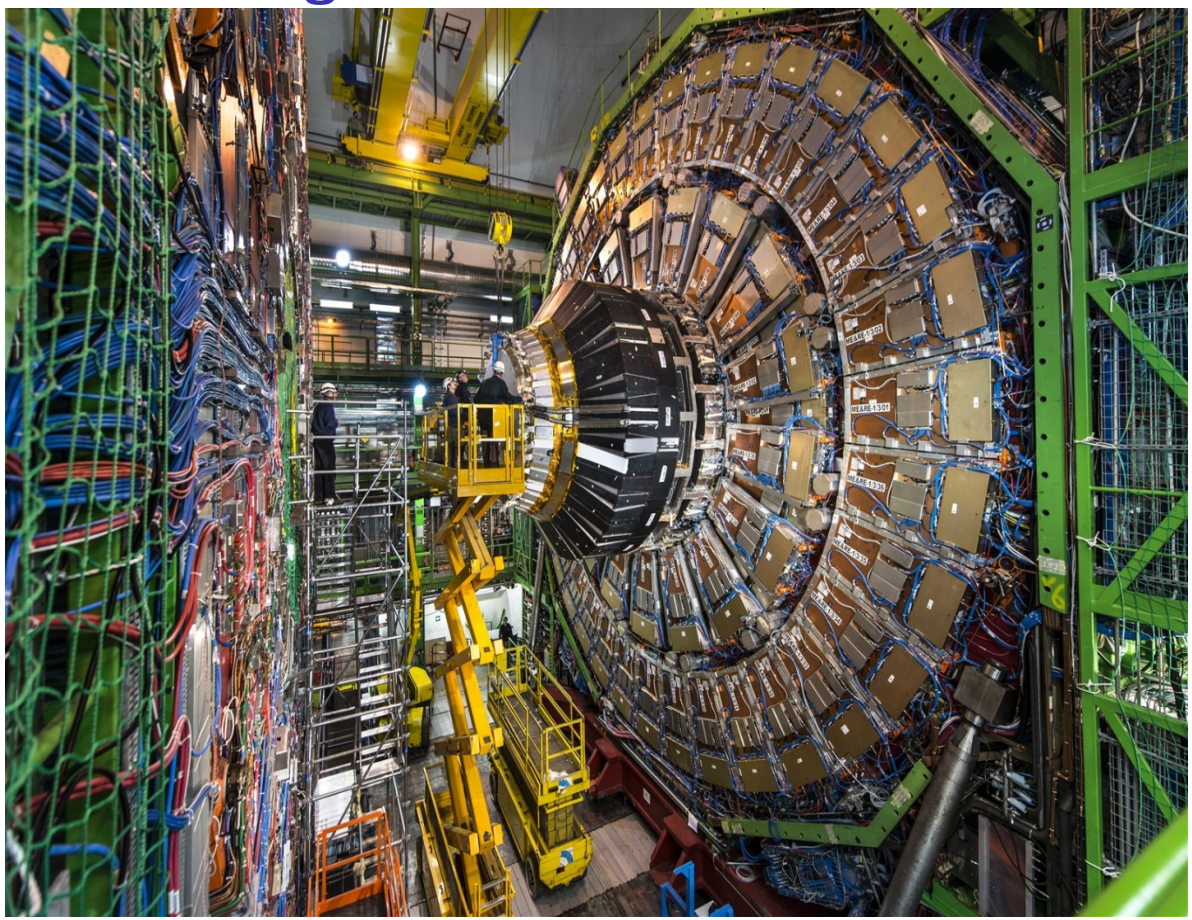


Figure 5. Popularity Dynamics Before and After the Inversion for the Most and Least Popular Song

Online experiments

New directions in digital labs



Bail, Christopher et al., 2023. "Do We Need a Social Media Accelerator?," *SocArxiv*.

Online experiments

Ethics

- Digital experimentation forces us to pay more attention to ethics
- Salganik proposes the "three R's"
  - *Replace* experiments with less invasive methods, where possible.
  - *Refine* treatment to reduce potential harm.
  - *Reduce* number of participants as much as possible.

Online surveys

Three eras of survey sampling

- Area probability sampling
  - Face-to-face interviews
- Random digit dialling
  - Phone interviews
- Non-probability sampling
  - Online surveys
    - Linked "big data"

Online surveys

Issues with online sampling

- No sampling frame
- Non-representative populations
- Selection bias (i.e. opt-in surveys)
- Violations of IID assumption violations (e.g. snowball sampling)

Online surveys

Forecasting elections with non-representative polls

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting elections with non-representative polls

Wei Wang [a,*], David Rothschild [b], Sharad Goel [b], Andrew Gelman [a,c]

[a] Department of Statistics, Columbia University, New York, NY, USA
[b] Microsoft Research, New York, NY, USA
[c] Department of Political Science, Columbia University, New York, NY, USA

ARTICLE INFO

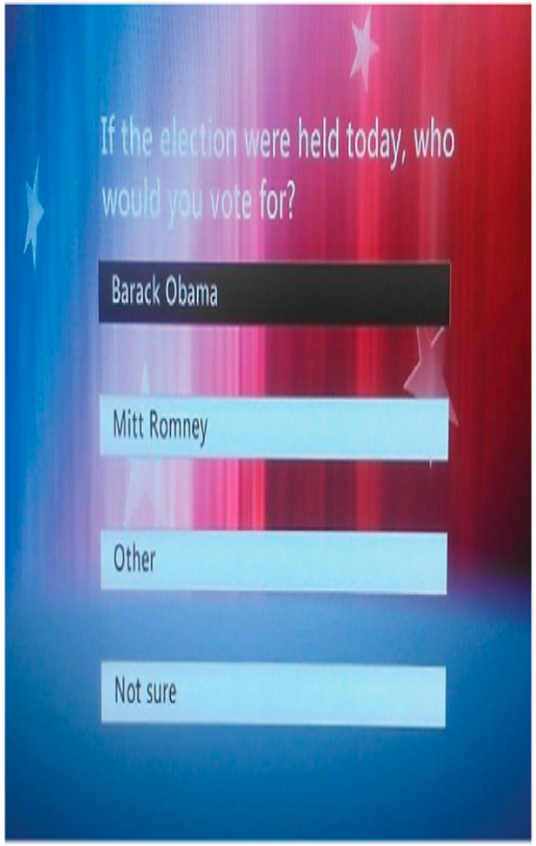
Keywords:
Non-representative polling
Multilevel regression and poststratification
Election forecasting

ABSTRACT

Election forecasts have traditionally been based on representative polls, in which randomly sampled individuals are asked who they intend to vote for. While representative polling has historically proven to be quite effective, it comes at considerable costs of time and money. Moreover, as response rates have declined over the past several decades, the statistical benefits of representative sampling have diminished. In this paper, we show that, with proper statistical adjustment, non-representative polls can be used to generate accurate election forecasts, and that this can often be achieved faster and at a lesser expense than traditional survey methods. We demonstrate this approach by creating forecasts from a novel and highly non-representative survey dataset: a series of daily voter intention polls for the 2012 presidential election conducted on the Xbox gaming platform. After adjusting the Xbox responses via multilevel regression and poststratification, we obtain estimates which are in line with the forecasts from leading poll analysts, which were based on aggregating hundreds of traditional polls conducted during the election cycle. We conclude by arguing that non-representative polling shows promise not only for election forecasting, but also for measuring public opinion on a broad range of social, economic and cultural issues.

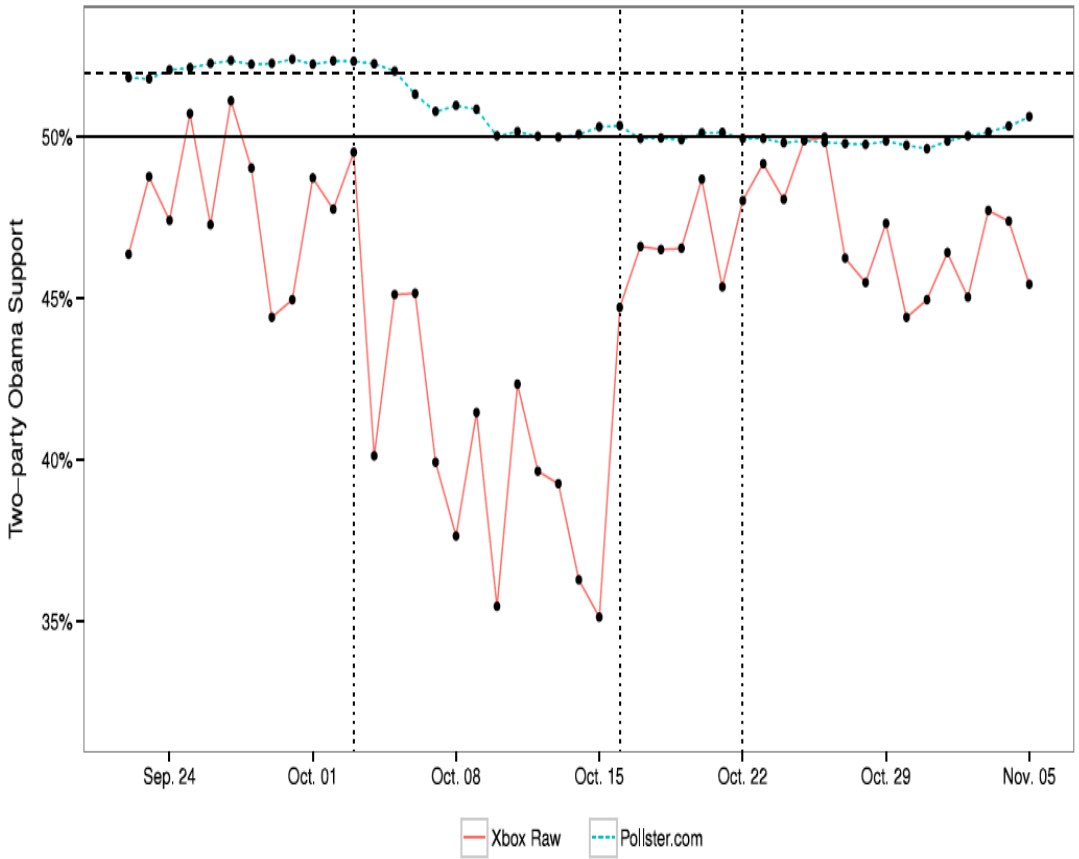
Online surveys

Survey design



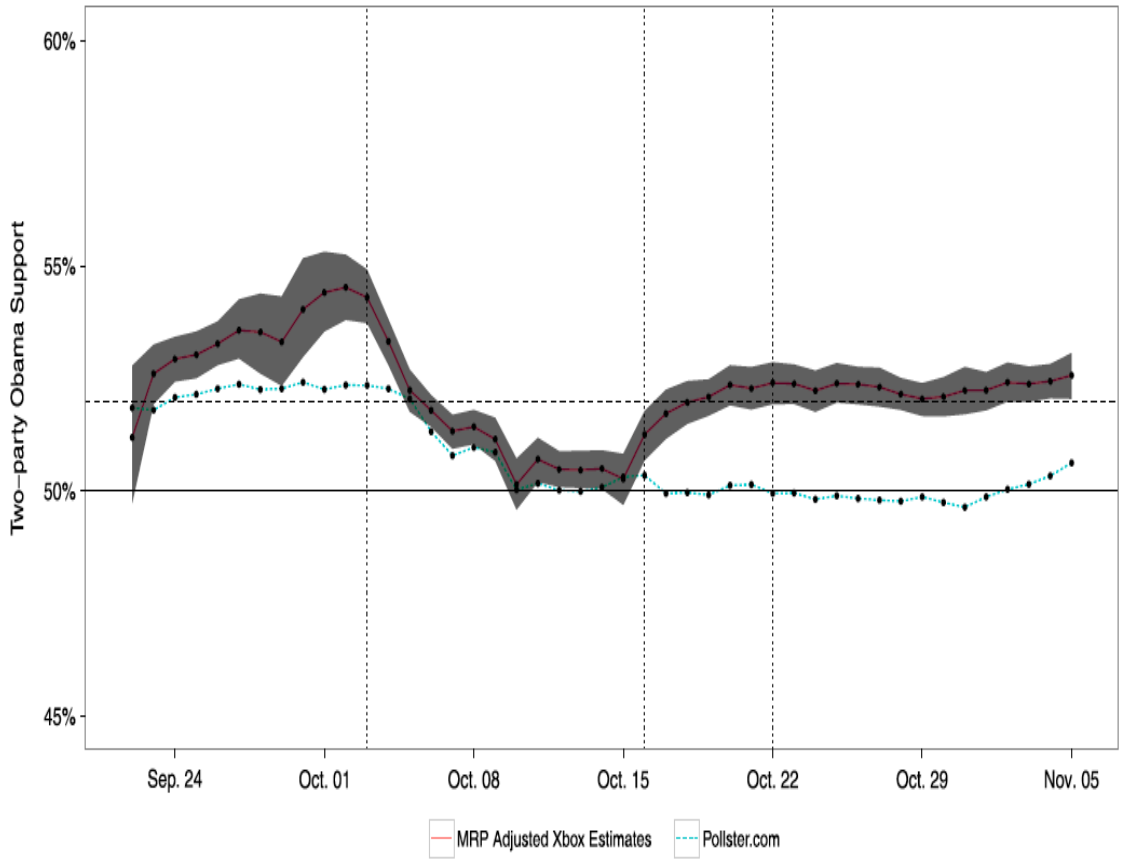
Online surveys

Polls before adjustment



Online surveys

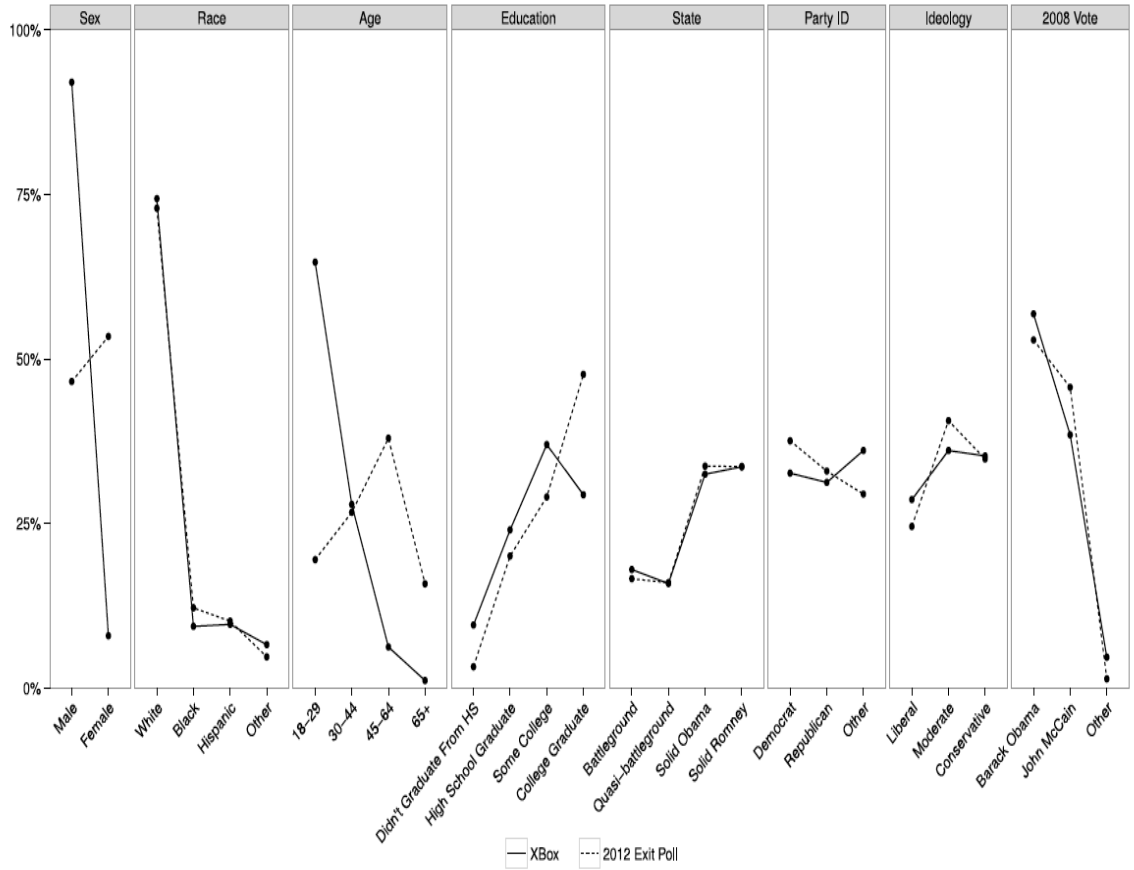
Polls after adjustment



Multilevel regression and post-stratification. See Salganik 130-6 for mathematical intuition. Monica Alexander has a great MRP primer with R code and Rohan Alexander has a more in-depth chapter in *Telling Stories with Data* (2023).

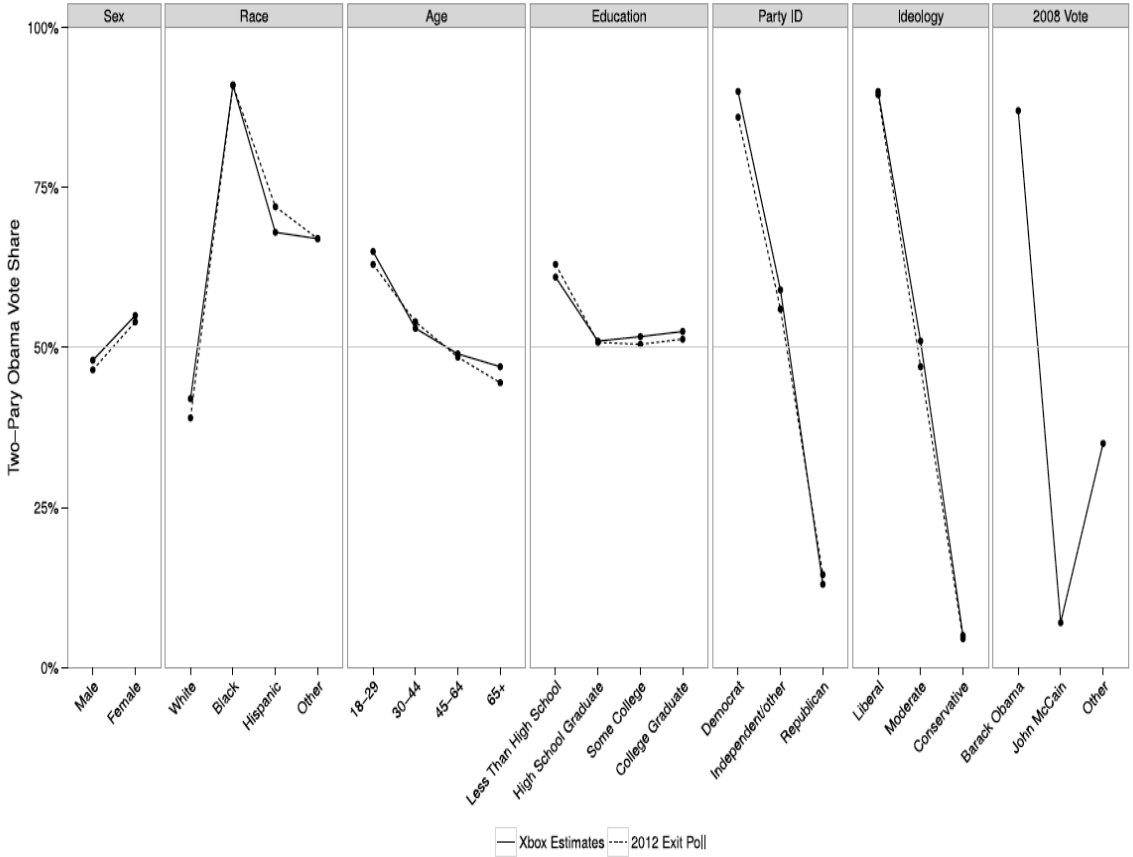
Online surveys

Demographics of Xbox users versus voters



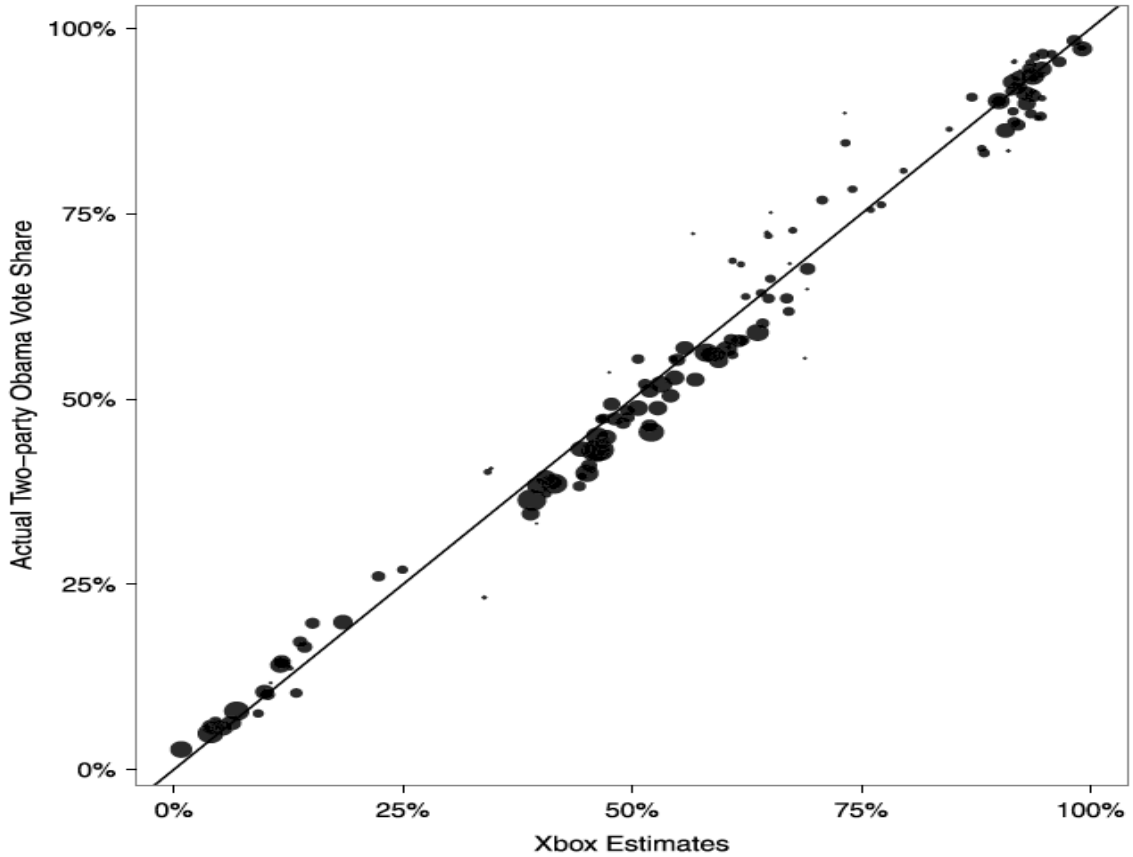
Online surveys

Population sub-group estimates



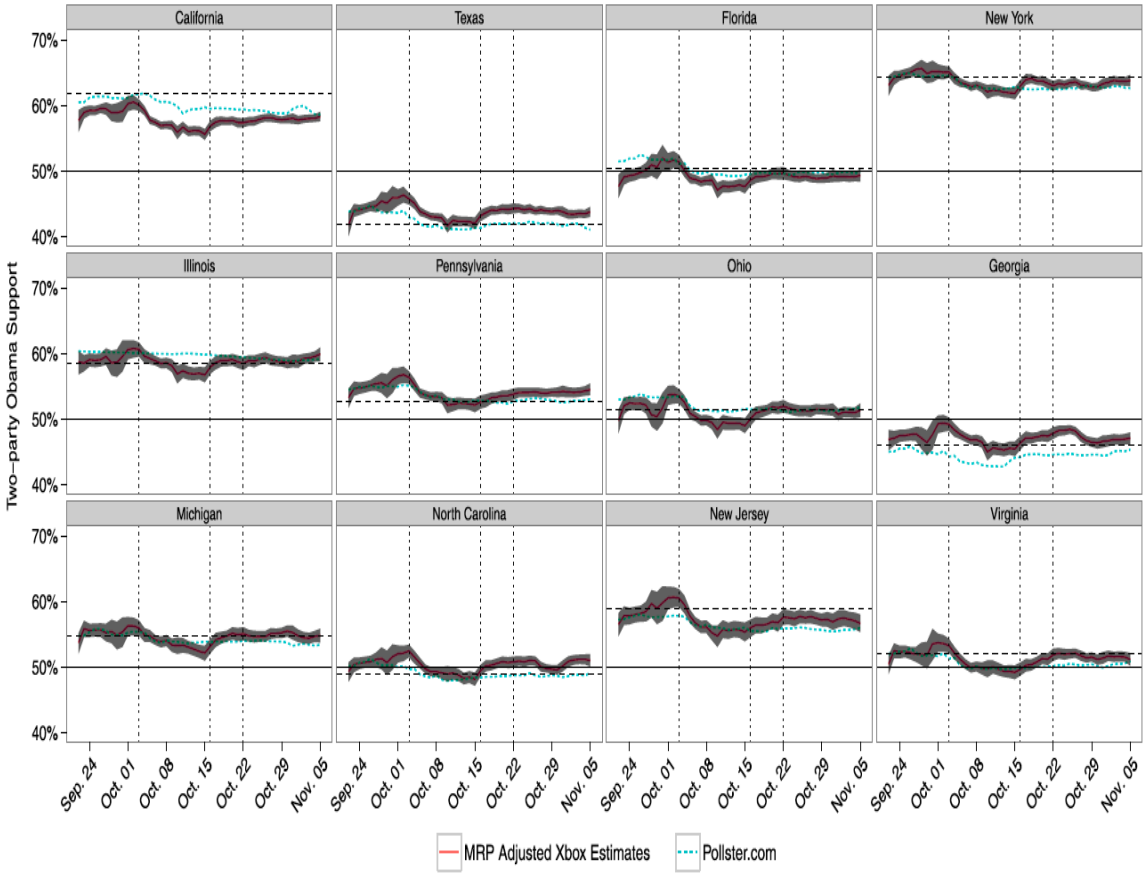
Online surveys

Errors



Online surveys

State-level estimates



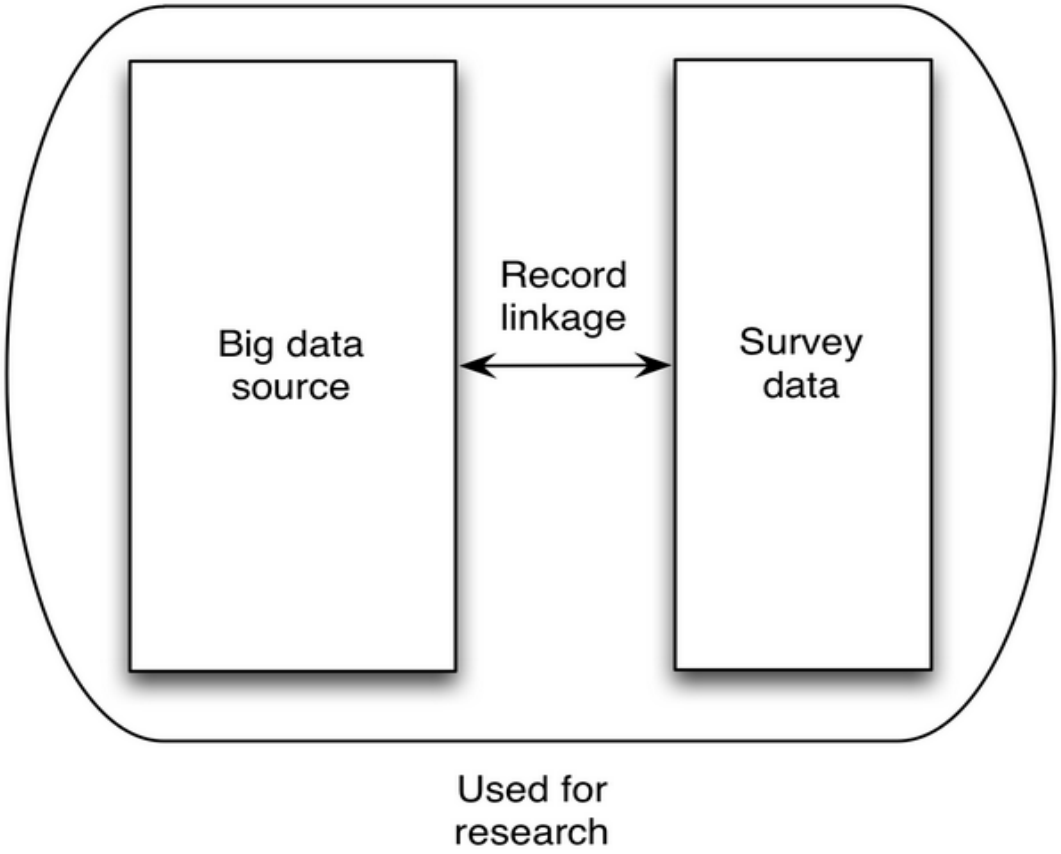
Online surveys

Working with non-probability samples

- Cheaper than fielding nationally-representative polls
- But more difficult to work with than conventional survey data
  - New statistical procedures and data sources non-probability sampling viable

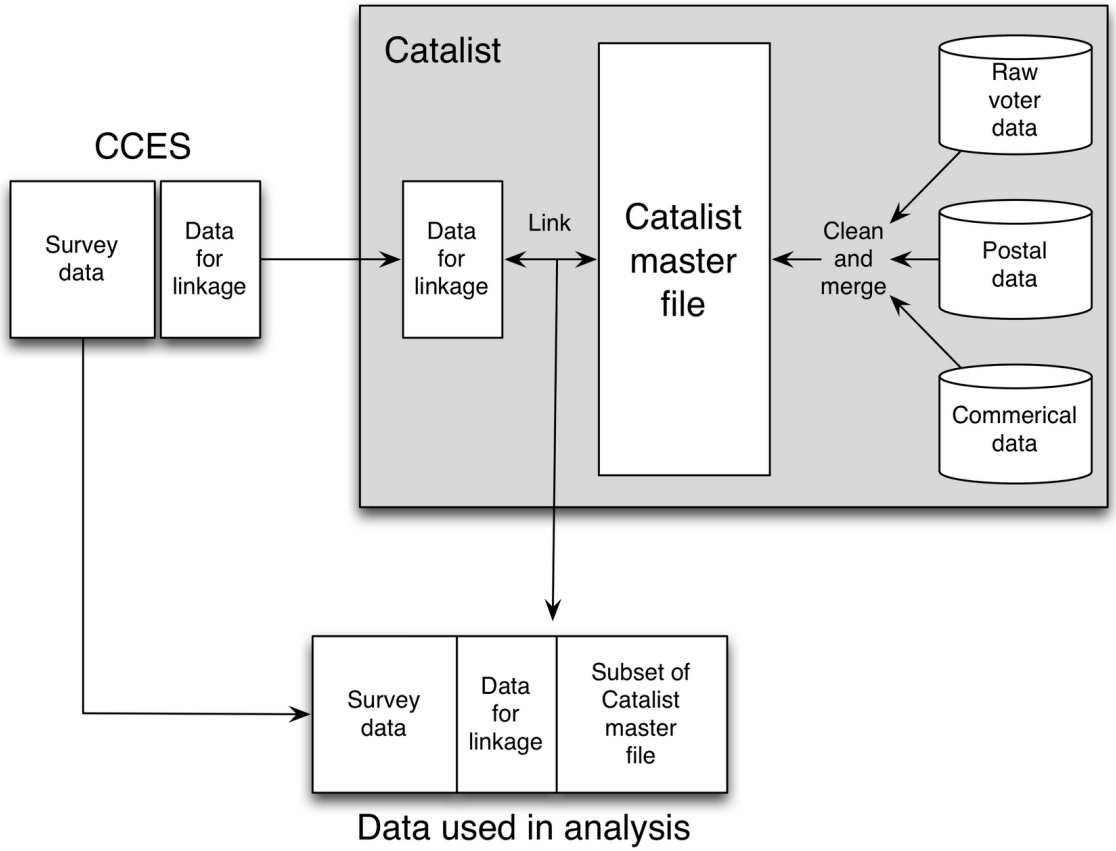
Online surveys

Record linkage / "enriched asking"



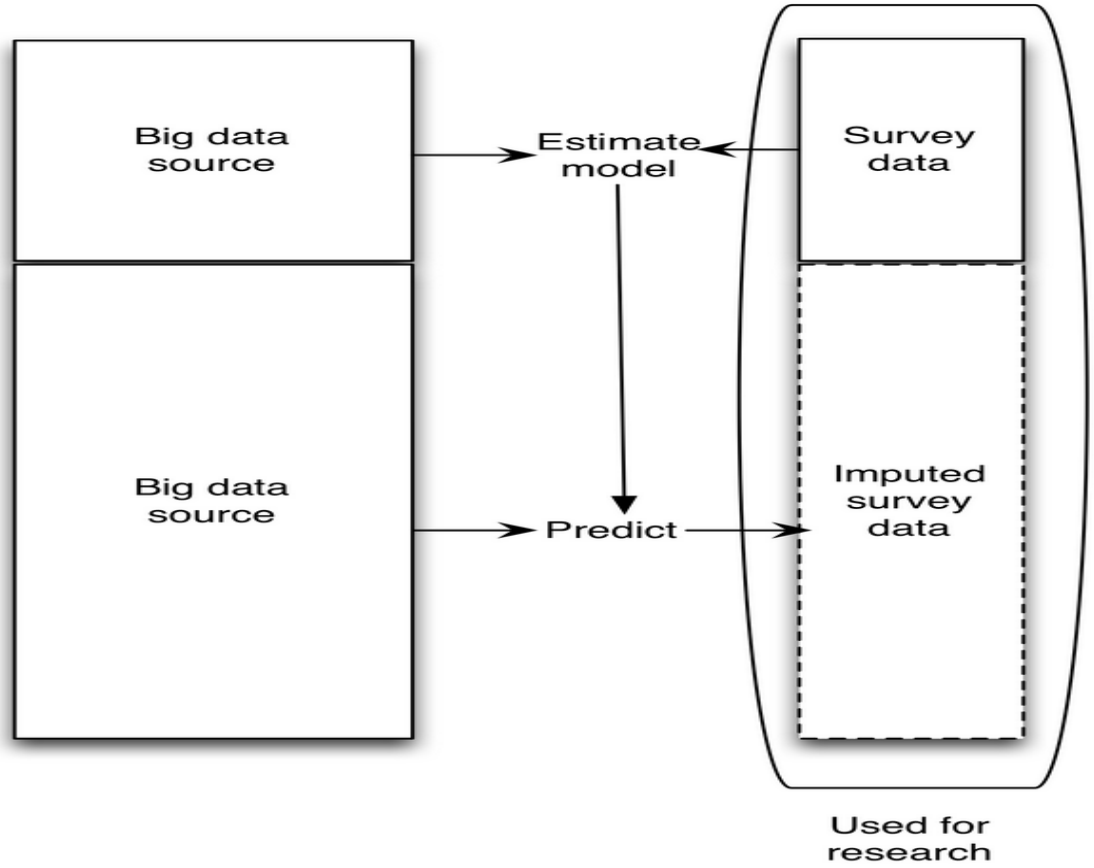
Online surveys

Enriched asking: voter behavior



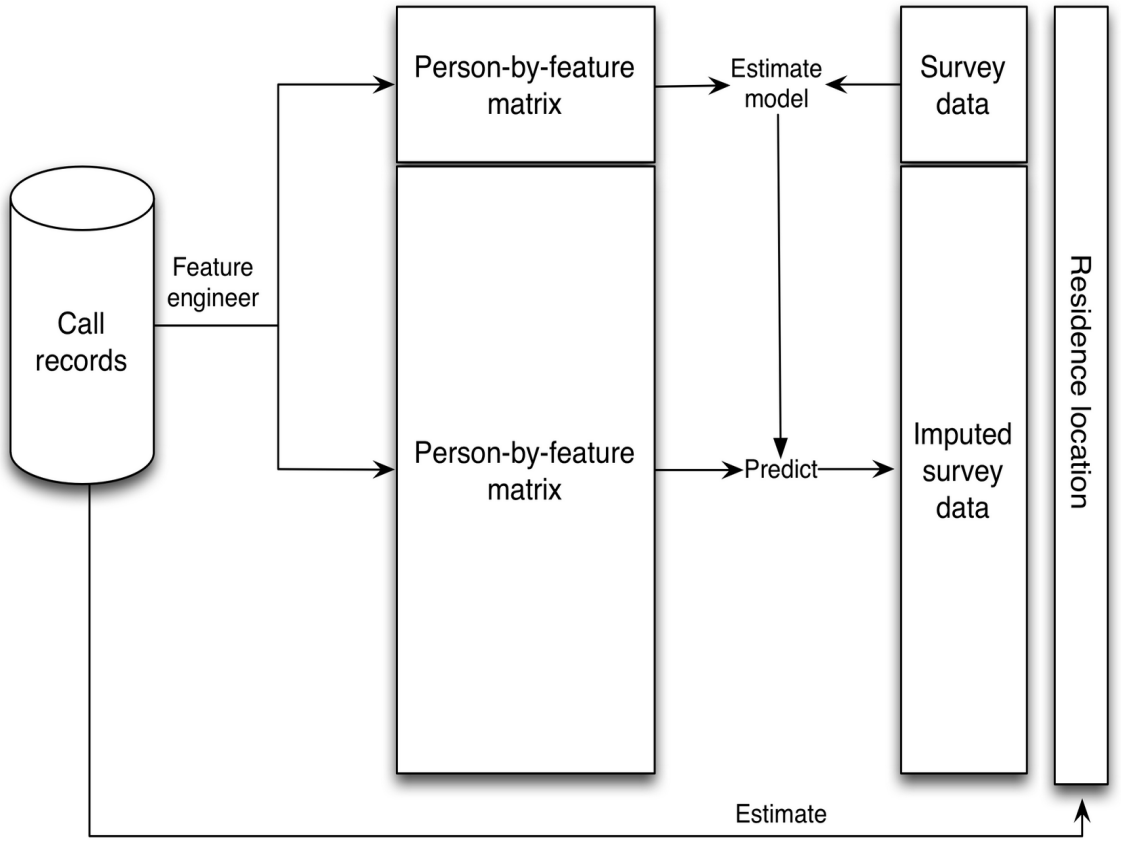
Online surveys

Big data imputation / "amplified asking"



Online surveys

Amplified asking: Mapping poverty in Rwanda



Final thoughts

- New technologies and data sources allow us to reinvent existing methods
  - Innovative work combines social scientific approaches, statistics, and programming in new ways
- Digital experiments and surveys open up many opportunities for social scientific research
  - These methods come with more challenges and require different skills to conventional methods
  - We must think more about ethics, related to informed consent, impacts on study participants, and implications of partnerships with other organizations

Next week

- Introduction to Natural Language Processing