Political Analysis of Social Media Data

Major Works

Instructor: Gregory Eady
Office: 18.2.10
Office hours: Fridays 13-15

Next week

1. Sign up for a presentation
2. I may email to ask you to sign up to the YouTube API

First, why presentations?

Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom

Louis Deslauriers[a,1], Logan S. McCarty[a,b], Kelly Miller[c], Kristina Callaghan[a], and Greg Kestin[a]

[a]Department of Physics, Harvard University, Cambridge, MA 02138; [b]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138; and [c]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved August 13, 2019 (received for review December 24, 2018)

We compared students' self-reported perception of learning with their actual learning under controlled conditions in large-enrollment introductory college physics courses taught using 1) active instruction (following best practices in the discipline) and 2) passive instruction (lectures by experienced and highly rated instructors). Both groups received identical class content and handouts, students were randomly assigned, and the instructor made no effort to persuade students of the benefit of either method. Students in active classrooms learned more (as would be expected based on prior research), but their perception of learning, while positive, was lower than that of their peers in passive environments. This suggests that attempts to evaluate instruction based on students' perceptions of learning could inadvertently promote inferior (passive) pedagogical methods. For instance, a superstar lecturer could create such a positive feeling of learning that students would choose those lectures over active learning. Most importantly, these results suggest that when students experience the increased cognitive effort associated with active learning, they initially take that effort to signify poorer learning. That disconnect may have a detrimental effect on students' motivation, engagement, and ability to self-regulate their own learning. Although students can, on their own, discover the increased value of being actively engaged during a semester-long course, their learning may be impaired during the initial part of the course. We discuss strategies that instructors can use, early in the semester, to improve students' response to being actively engaged in the classroom.

with the material. There is nothing known about how students naturally react to active learning without any promotion from the instructor. In addition, previous studies used different course materials for active versus passive instruction, potentially confounding the effect of pedagogy with that of course materials.

In this report, we identify an inherent student bias against active learning that can limit its effectiveness and may hinder the wide adoption of these methods. Compared with students in traditional lectures, students in active classes perceived that they learned less, while in reality they learned more. Students rated the quality of instruction in passive lectures more highly, and they expressed a preference to have "all of their physics classes taught this way," even though their scores on independent tests of learning were lower than those in actively taught classrooms. These findings are consistent with the observations that novices in a subject are poor judges of their own competence (27–29), and the cognitive fluency of lectures can be misleading (30, 31). Our findings also suggest that novice students may not accurately assess the changes in their own learning that follow from their experience in a class. These misperceptions must be understood and addressed in order for research-based active instructional strategies to be more effective and to become widespread.

Materials and Methods

Our study sought to measure students' perception of learning when active learning alone is toggled on and off. This contrasts with typical educational interventions that include active engagement as one component of many

Audience members

- A critical question about the paper
- A criticism of the paper
- Write out your question or comment so you can read it if called on

Reading empirically driven research helps build intuition for your own work

1. Read with research design in mind
2. Many questions are descriptive
3. Others must be tackled indirectly
4. Still others are examined experimentally
5. So think about what empirical limitations the authors note, and also why they run the robustness checks that they do
6. Read strategically...

Reading strategically

- Skim the abstract
- Skim the introduction
- Skim the theory/hypotheses
- Read the research design to a sufficient extent that you have a good understanding of the data and methods involved
- Read the results section in more depth
- Skim the conclusion

Some big questions in the study of social media

1. Why do governments censor social media?
2. What effects does censorship have on political behavior?
3. Does social media facilitate collective action? (e.g. protests)
4. Do governments benefit from using social media for political surveillance during civil conflict?
5. Does social media increase political radicalization?
6. Are social media algorithms biased?
7. Do women and racial/religious minorities face more toxicity on social media?

Some big questions in the study of social media

8. What interventions work to prevent prejudice and toxicity online?
9. Do people live in political echo chambers?
10. Is social media politically polarizing?
11. How extensive is fake-news sharing, and who shares it?
12. What are the effects of social media on knowledge, well-being, and voting behavior?
13. Is social media advertising beneficial or detrimental to democracy?
14. Does social media benefit populists & facilitate foreign interference in elections?

Why do governments censor social media?

King, Pan, and Roberts (2013, 2014, 2017)

Three papers:

1. Collect Chinese social media posts before they were censored (KPR 2013)
2. Run a large-scale field experiment by creating posts on Chinese social media sites to see what gets censored (KPR 2014)
3. Examine leaked 50c party members' social media posts to see how, when, and why the Chinese government fabricates content on social media (KPR 2017)

Results?

How Censorship in China Allows Government Criticism but Silences Collective Expression

GARY KING *Harvard University*
JENNIFER PAN *Harvard University*
MARGARET E. ROBERTS *Harvard University*

We offer the first large scale, multiple source analysis of the outcome of what may be the most extensive effort to selectively censor human expression ever implemented. To do this, we have devised a system to locate, download, and analyze the content of millions of social media posts originating from nearly 1,400 different social media services all over China before the Chinese government is able to find, evaluate, and censor (i.e., remove from the Internet) the subset they deem objectionable. Using modern computer-assisted text analytic methods that we adapt to and validate in the Chinese language, we compare the substantive content of posts censored to those not censored over time in each of 85 topic areas. Contrary to previous understandings, posts with negative, even vitriolic, criticism of the state, its leaders, and its policies are not more likely to be censored. Instead, we show that the censorship program is aimed at curtailing collective action by silencing comments that represent, reinforce, or spur social mobilization, regardless of content. Censorship is oriented toward attempting to forestall collective activities that are occurring now or may occur in the future—and, as such, seem to clearly expose government intent.

Figure 9. Content of All Censored Posts (Regardless of Topic Area)

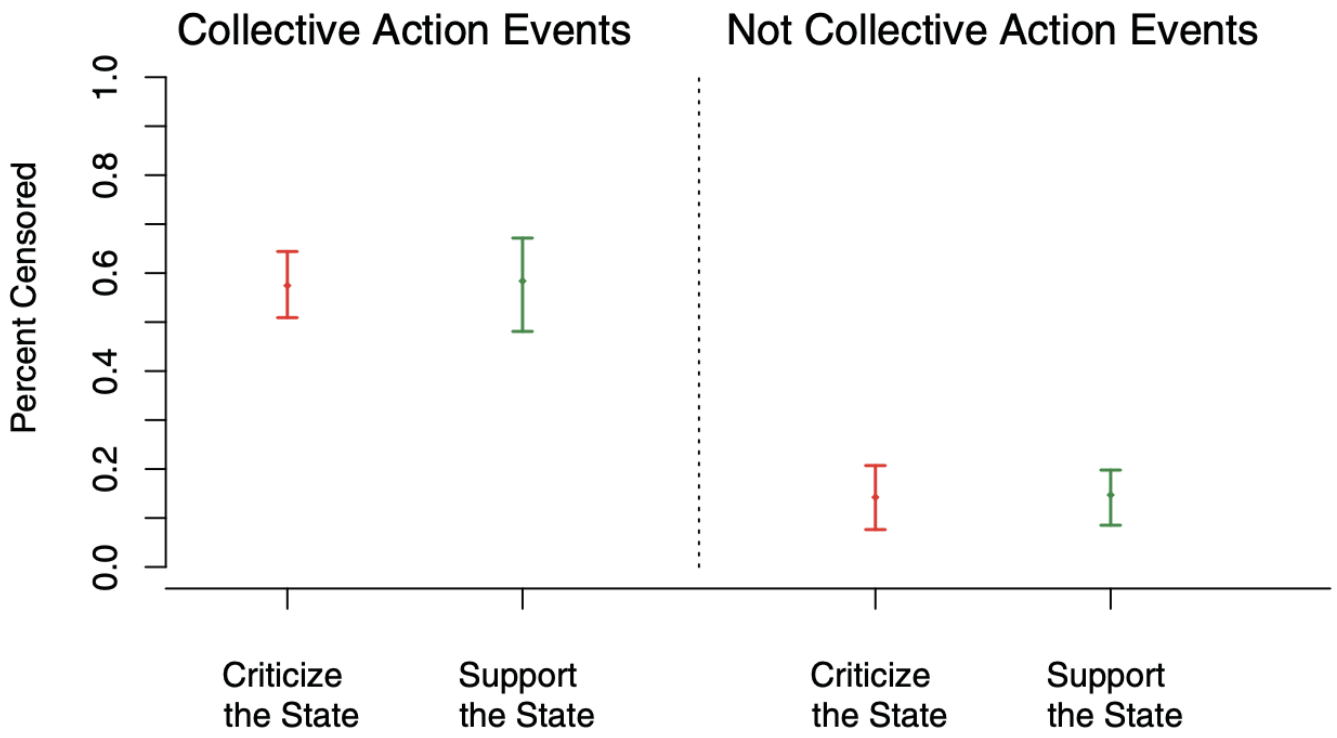
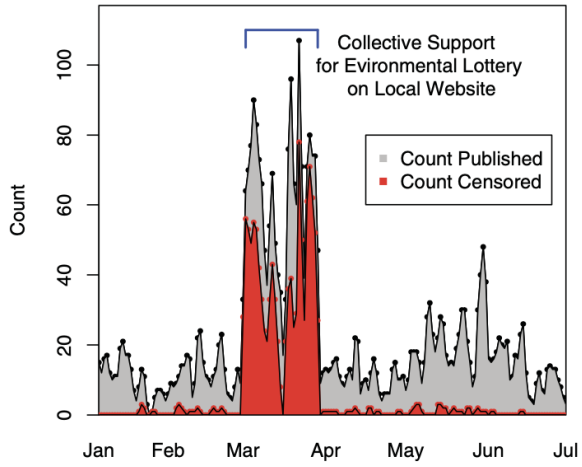
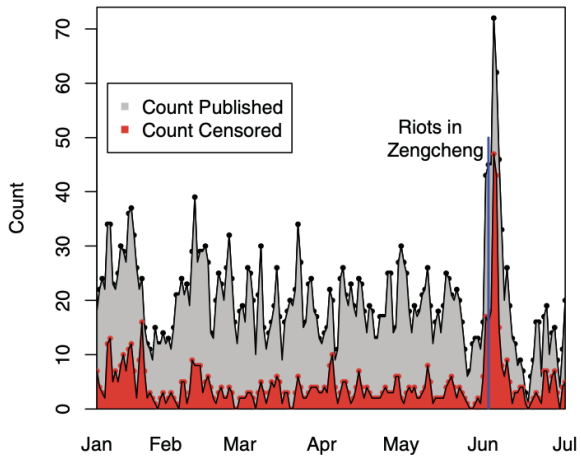


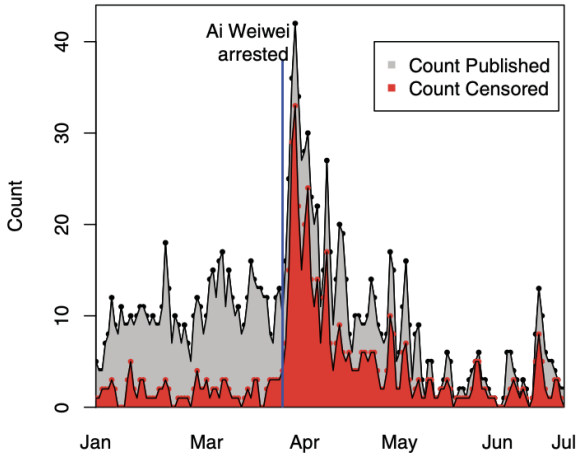
Figure 5. High Censorship During Collective Action Events (in 2011)



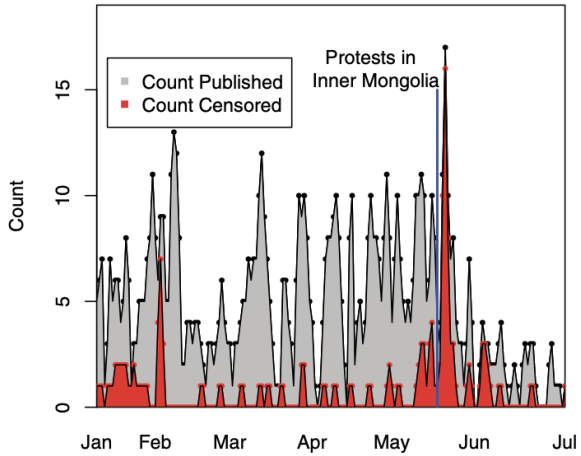
(a) Chen Fei's Environmental Lottery



(b) Riots in Zengcheng



(c) Dissident Ai Weiwei



(d) Inner Mongolia Protests

Reverse-engineering censorship in China: Randomized experimentation and participant observation

Gary King,[1*] Jennifer Pan,[1] Margaret E. Roberts[2]

Existing research on the extensive Chinese censorship organization uses observational methods with well-known limitations. We conducted the first large-scale experimental study of censorship by creating accounts on numerous social media sites, randomly submitting different texts, and observing from a worldwide network of computers which texts were censored and which were not. We also supplemented interviews with confidential sources by creating our own social media site, contracting with Chinese firms to install the same censoring technologies as existing sites, and—with their software, documentation, and even customer support—reverse-engineering how it all works. Our results offer rigorous support for the recent hypothesis that criticisms of the state, its leaders, and their policies are published, whereas posts about real-world events with collective action potential are censored.

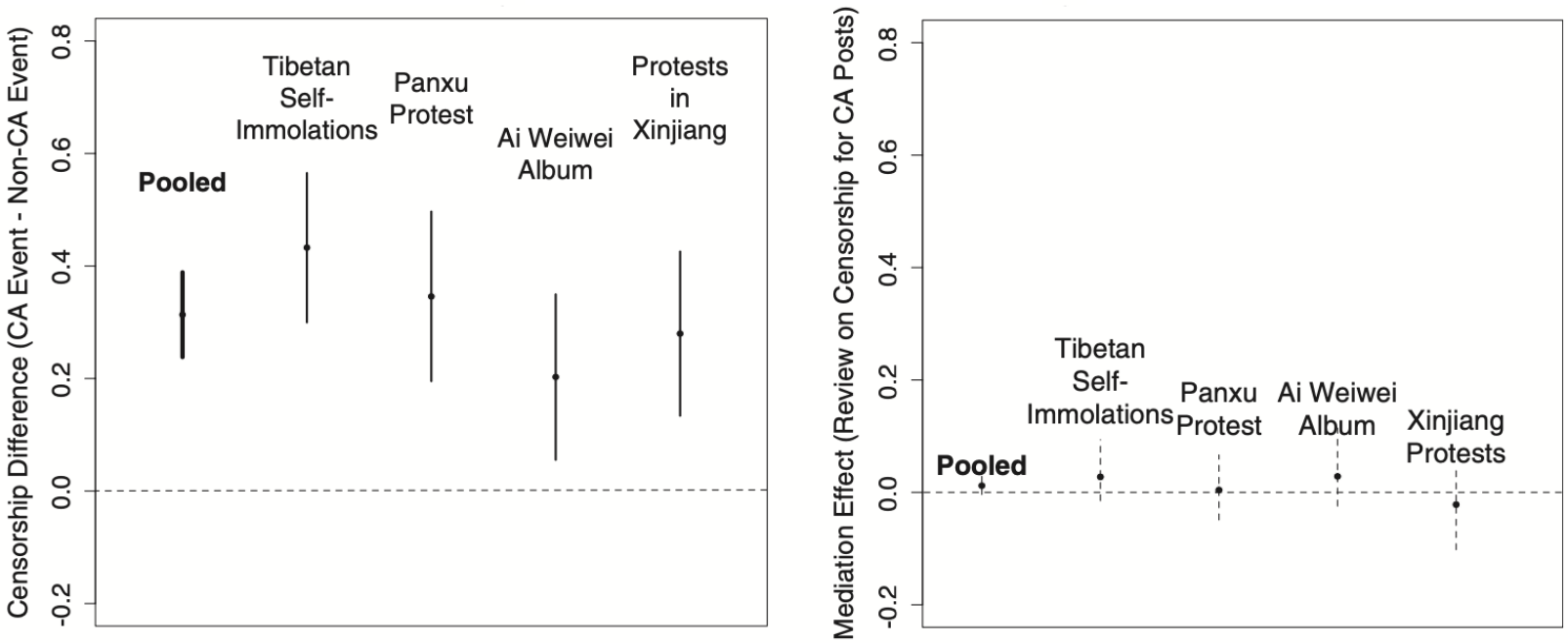


Fig. 2. The causal effect on censorship of posts with collective action potential (left panel) and the mediation effect of review (right panel). Collective action events are more highly censored than non–collective action events within the same time period. However, censorship of collective action events is not mediated through automated review.

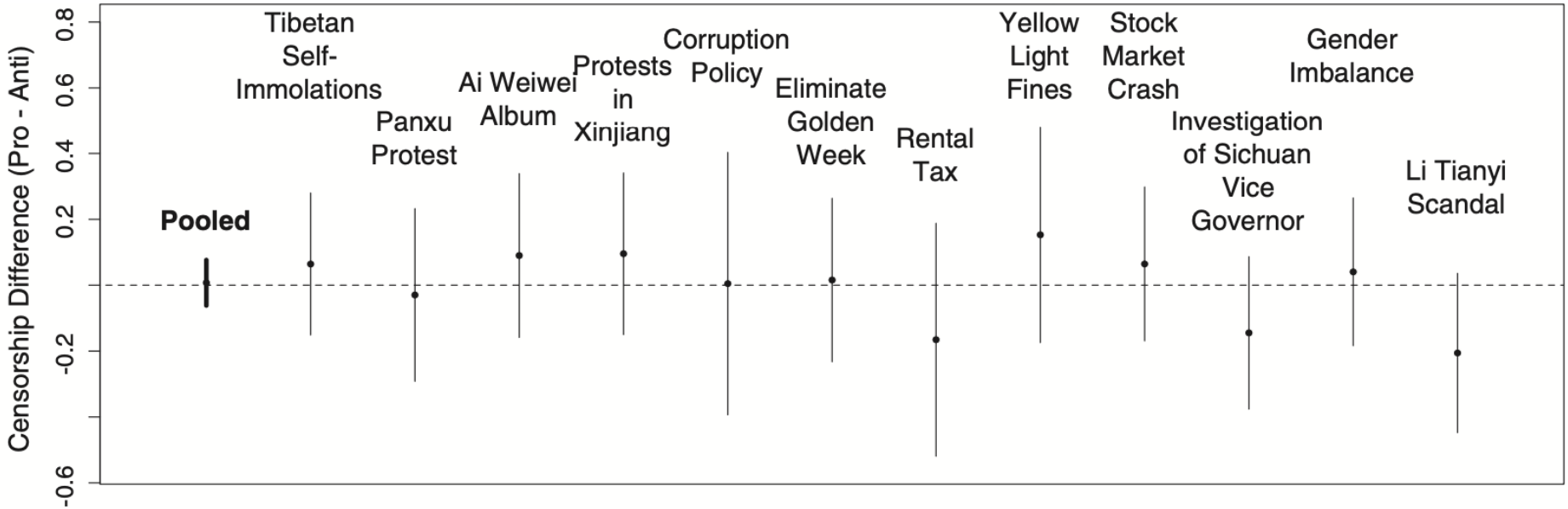


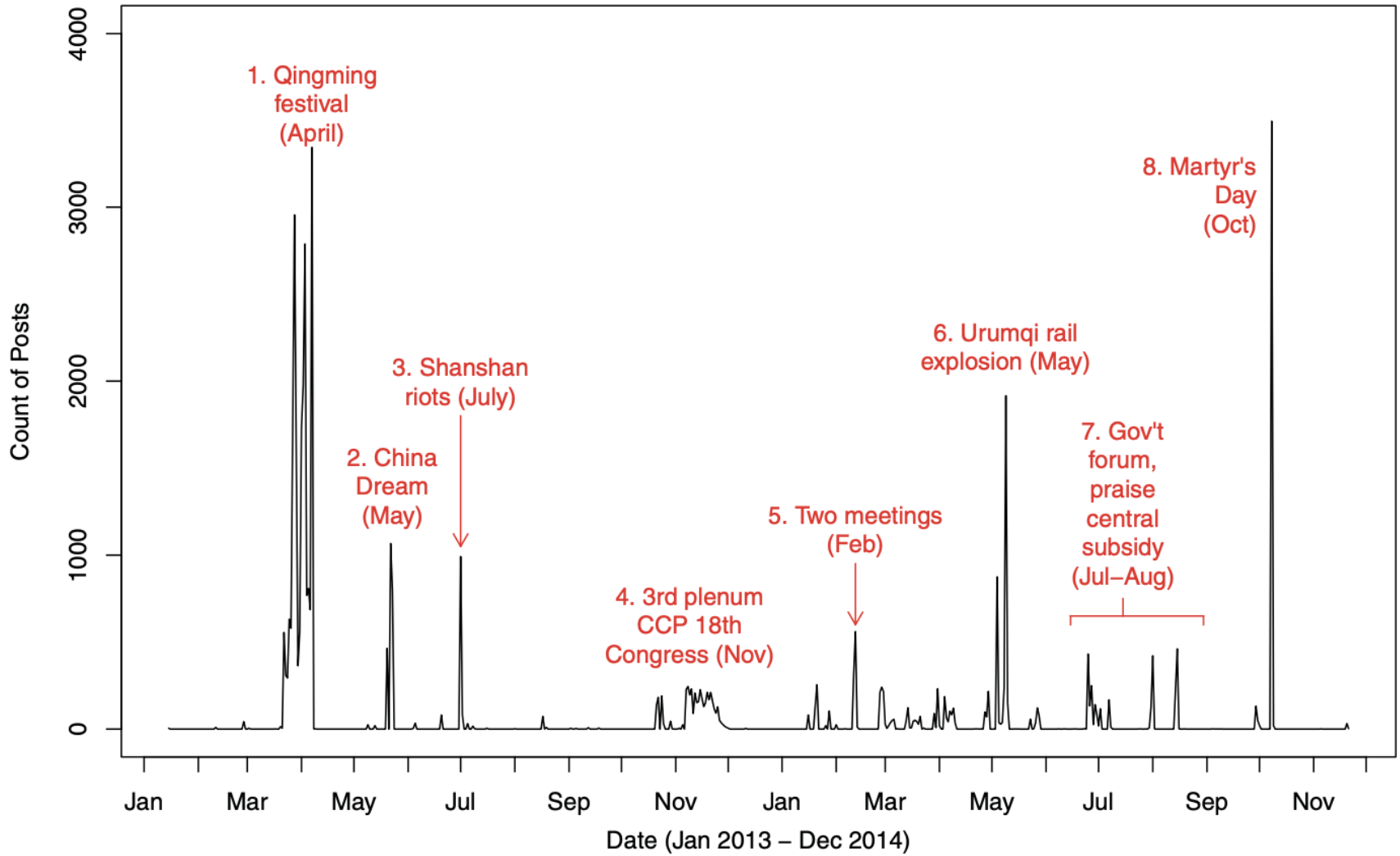
Fig. 3. The causal effect on censorship of posts for or against the government. Posts that support the government are not more or less likely to be censored than posts that oppose the government, within the same topic.

How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument

GARY KING *Harvard University*
JENNIFER PAN *Stanford University*
MARGARET E. ROBERTS *University of California, San Diego*

The Chinese government has long been suspected of hiring as many as 2 million people to surreptitiously insert huge numbers of pseudonymous and other deceptive writings into the stream of real social media posts, as if they were the genuine opinions of ordinary people. Many academics, and most journalists and activists, claim that these so-called 50c party posts vociferously argue for the government's side in political and policy debates. As we show, this is also true of most posts openly accused on social media of being 50c. Yet almost no systematic empirical evidence exists for this claim or, more importantly, for the Chinese regime's strategic objective in pursuing this activity. In the first large-scale empirical analysis of this operation, we show how to identify the secretive authors of these posts, the posts written by them, and their content. We estimate that the government fabricates and posts about 448 million social media comments a year. In contrast to prior claims, we show that the Chinese regime's strategy is to avoid arguing with skeptics of the party and the government, and to not even discuss controversial issues. We show that the goal of this massive secretive operation is instead to distract the public and change the subject, as most of these posts involve cheerleading for China, the revolutionary history of the Communist Party, or other symbols of the regime. We discuss how these results fit with what is known about the Chinese censorship program and suggest how they may change our broader theoretical understanding of "common knowledge" and information control in authoritarian regimes.

FIGURE 2. Time Series of 43,757 Known 50c Social Media Posts with Qualitative Summaries of the Content of Volume Bursts



What effects does censorship have on political behavior?

Hobbs & Roberts (2018) & Chen & Yang (2019)

1. Examine the ban of (non-political) Instagram in China
2. Experimentally provide users in China with access to a VPN

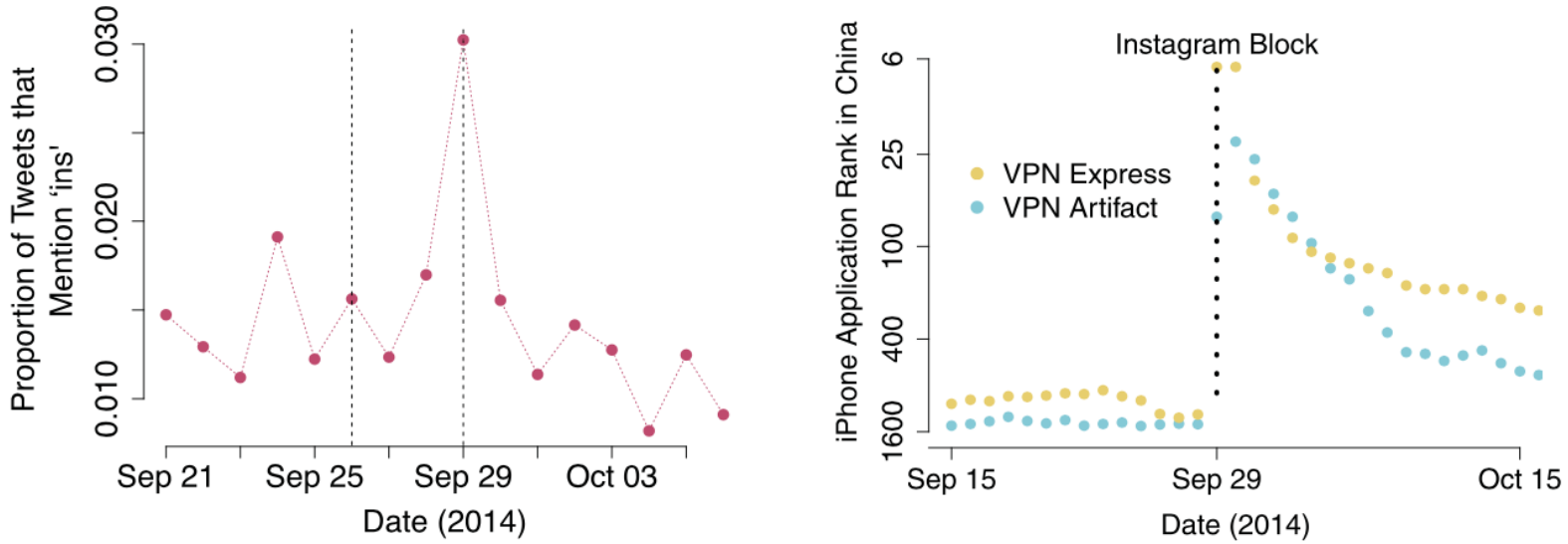
Results?

How Sudden Censorship Can Increase Access to Information

WILLIAM R. HOBBS *Northeastern University*
MARGARET E. ROBERTS *University of California, San Diego*

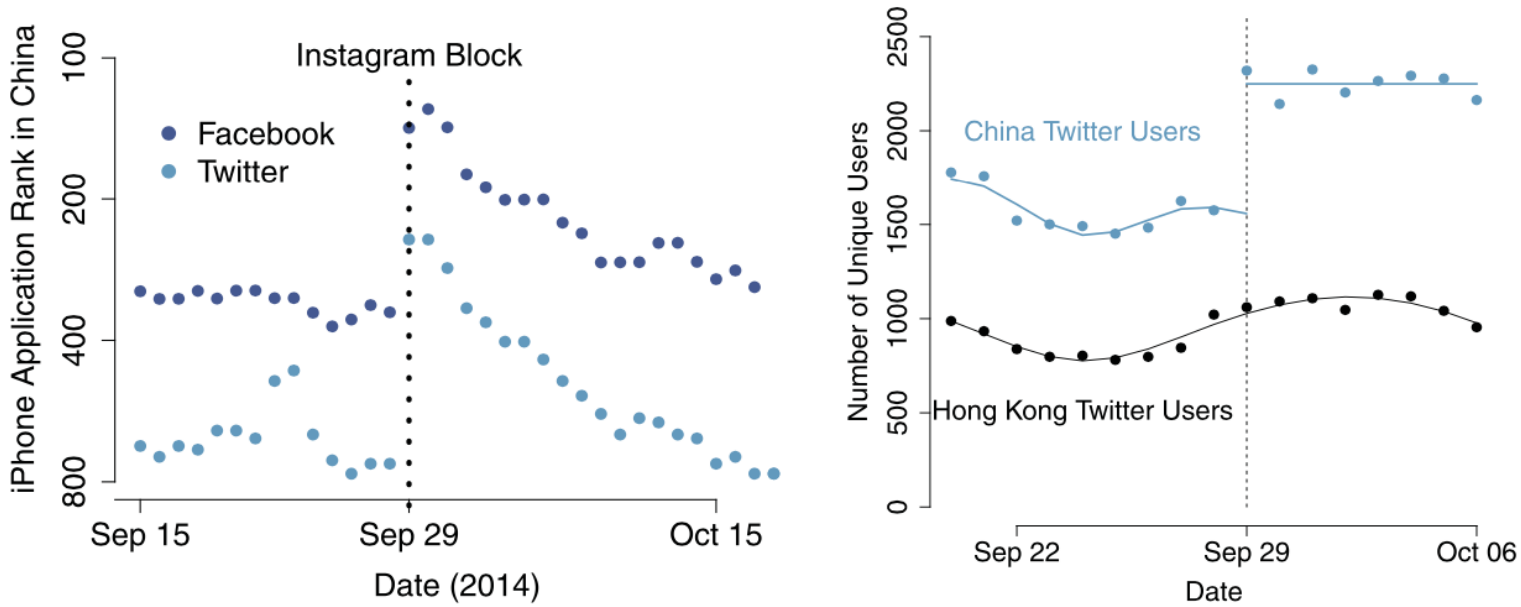
Conventional wisdom assumes that increased censorship will strictly decrease access to information. We delineate circumstances when increases in censorship expand access to information for a substantial subset of the population. When governments suddenly impose censorship on previously uncensored information, citizens accustomed to acquiring this information will be incentivized to learn methods of censorship evasion. These evasion tools provide continued access to the newly blocked information—and also extend users' ability to access information that has long been censored. We illustrate this phenomenon using millions of individual-level actions of social media users in China before and after the block of Instagram. We show that the block inspired millions of Chinese users to acquire virtual private networks, and that these users subsequently joined censored websites like Twitter and Facebook. Despite initially being apolitical, these new users began browsing blocked political pages on Wikipedia, following Chinese political activists on Twitter, and discussing highly politicized topics such as opposition protests in Hong Kong.

FIGURE 2. Left: Proportion of Tweets from China mentioning "ins" by day. Right: The Instagram block's effect on the rank of VPN applications on iPhones from mainland China, from AppAnnie.com.



In the left panel of this figure we show that 3% of tweets in China mentioned Instagram on the day of the Instagram block. The right panel of this figure shows that the download ranks of VPN Express and VPN Artifact increased from ranks lower than 1,000 to the top 10 most popular applications in China on the day of the Instagram block.

FIGURE 3. Left: The Instagram block's effect on the rank of Facebook and Twitter on iPhones from mainland China, from AppAnnie.com. Right: Comparison of tweets per day from Mainland China and Hong Kong before and after the Instagram block.



The left panel of this figure shows the change in download ranks for Facebook and Twitter before and after Instagram was blocked. The right panel of this figure shows that the Chinese Twitter users in our sample increased 30% the same day that we observe a spike in Instagram mentions and several days after the beginning of the Hong Kong protests. This increase only occurred in China and not in Hong Kong. The lines in this panel were fit using a smoothing spline.

The Impact of Media Censorship: 1984 or Brave New World?†

By Yuyu Chen and David Y. Yang*

Media censorship is a hallmark of authoritarian regimes. We conduct a field experiment in China to measure the effects of providing citizens with access to an uncensored internet. We track subjects' media consumption, beliefs regarding the media, economic beliefs, political attitudes, and behaviors over 18 months. We find four main results: (i) free access alone does not induce subjects to acquire politically sensitive information; (ii) temporary encouragement leads to a persistent increase in acquisition, indicating that demand is not permanently low; (iii) acquisition brings broad, substantial, and persistent changes to knowledge, beliefs, attitudes, and intended behaviors; and (iv) social transmission of information is statistically significant but small in magnitude. We calibrate a simple model to show that the combination of low demand for uncensored information and the moderate social transmission means China's censorship apparatus may remain robust to a large number of citizens receiving access to an uncensored internet. (*JEL* C93, D72, D83, L82, L86, L88, P36)

Does social media facilitate collective action?
(e.g. protests)

Clarke & Kocak (Forthcoming), & Larson et al. (2019)

1. Investigate how social media platforms used to mobilize and participate (Clarke & Kocak, Forthcoming)
2. Examine social media network structure and protest turnout (Larson et al. 2019)

Results?

Launching Revolution: Social Media and the Egyptian Uprising's First Movers

Killian Clarke and Korhan Kocak*

Department of Politics, Princeton University
*Corresponding author. Email: kkocak@princeton.edu

(Received 7 November 2016; revised 14 December 2017; accepted 17 April 2018)

Abstract

Drawing on evidence from the 2011 Egyptian uprising, this article demonstrates how the use of two social media platforms – Facebook and Twitter – contributed to a discrete mobilizational outcome: the staging of a successful first protest in a revolutionary cascade, referred to here as 'first-mover mobilization'. Specifically, it argues that these two platforms facilitated the staging of a large, nationwide and seemingly leaderless protest on 25 January 2011, which signaled to hesitant but sympathetic Egyptians that a revolution might be in the making. It draws on qualitative and quantitative evidence, including interviews, social media data and surveys, to analyze three mechanisms that linked these platforms to the success of the January 25 protest: (1) protester recruitment, (2) protest planning and coordination, and (3) live updating about protest logistics. The article not only contributes to debates about the role of the Internet in the Arab Spring and other recent waves of mobilization, but also demonstrates how scholarship on the Internet in politics might move toward making more discrete, empirically grounded causal claims.

Coordination of updates and slogans (w/topic models)

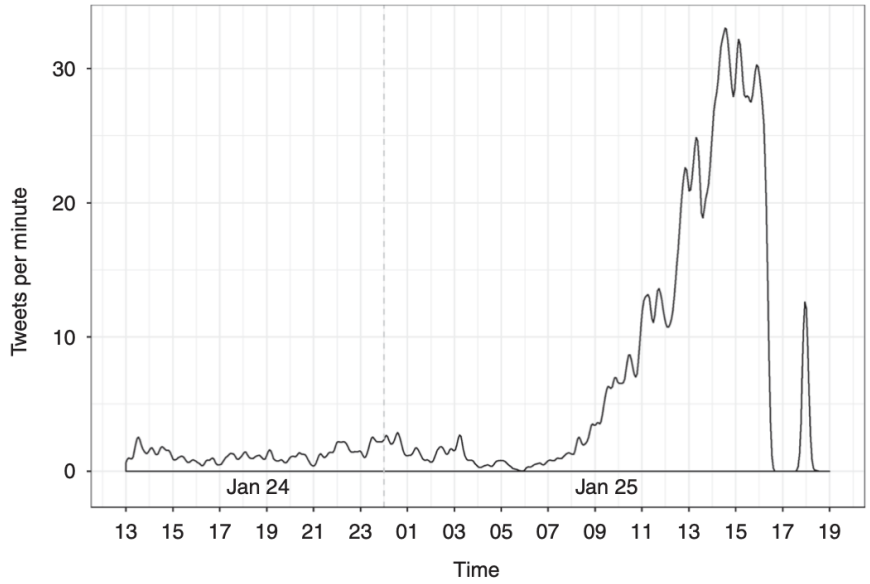
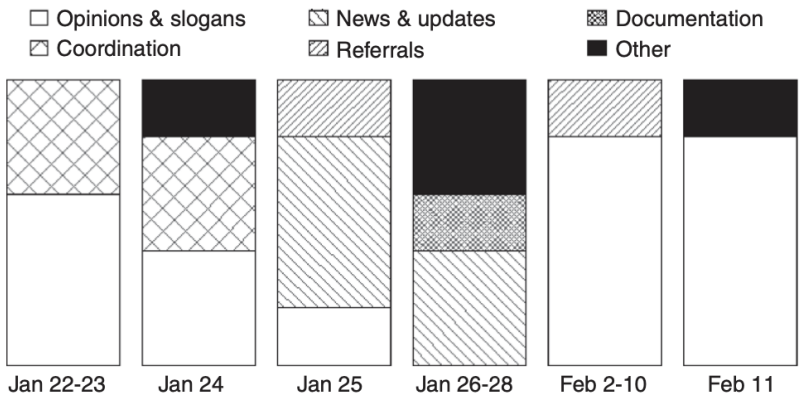


Figure 2. Twitter activity density, 24 January (1PM) to January 25 (7PM).



Social Networks and Protest Participation: Evidence from 130 Million Twitter Users

Jennifer M. Larson Vanderbilt University
Jonathan Nagler New York University
Jonathan Ronen Berlin Institute for Medical Systems Biology
Joshua A. Tucker New York University

Abstract: *Pinning down the role of social ties in the decision to protest has been notoriously elusive, largely due to data limitations. Social media and their global use by protesters offer an unprecedented opportunity to observe real-time social ties and online behavior, though often without an attendant measure of real-world behavior. We collect data on Twitter activity during the 2015* Charlie Hebdo *protest in Paris, which, unusually, record real-world protest attendance and network structure measured beyond egocentric networks. We devise a test of social theories of protest that hold that participation depends on exposure to others' intentions and network position determines exposure. Our findings are strongly consistent with these theories, showing that protesters are significantly more connected to one another via direct, indirect, triadic, and reciprocated ties than comparable nonprotesters. These results offer the first large-scale empirical support for the claim that social network structure has consequences for protest participation.*

Do governments benefit from social media surveillance during civil conflict?

Gohdes (2020)

- Examines whether areas in the Syrian civil war that are cut off from internet access experience more indiscriminate violence

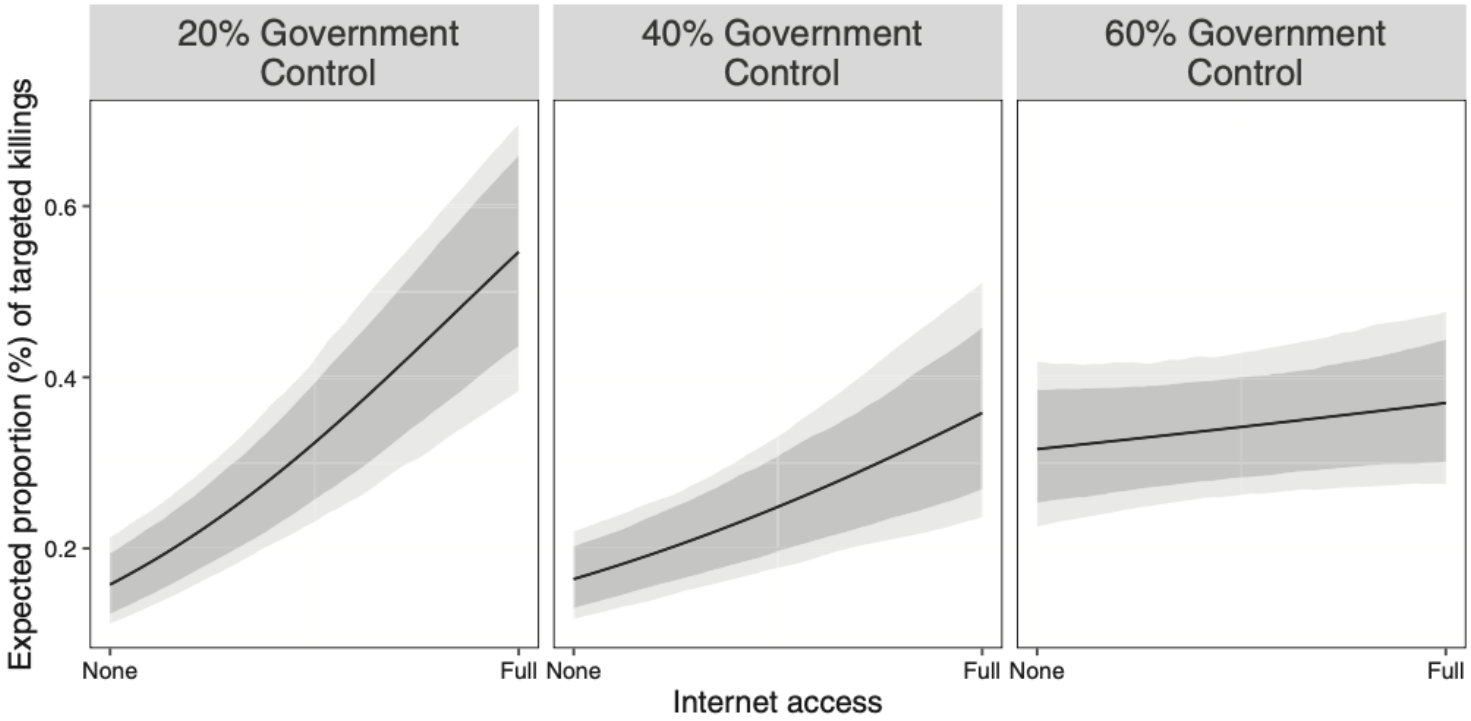
Results?

Repression Technology: Internet Accessibility and State Violence

Anita R. Gohdes Hertie School of Governance

Abstract: *This article offers a first subnational analysis of the relationship between states' dynamic control of Internet access and their use of violent repression. I argue that where governments provide Internet access, surveillance of digital information exchange can provide intelligence that enables the use of more targeted forms of repression, in particular in areas not fully controlled by the regime. Increasing restrictions on Internet accessibility can impede opposition organization, but they limit access to information on precise targets, resulting in an increase in untargeted repression. I present new data on killings in the Syrian conflict that distinguish between targeted and untargeted events, using supervised text classification. I find that higher levels of Internet accessibility are associated with increases in targeted repression, whereas areas with limited access experience more indiscriminate campaigns of violence. The results offer important implications on how governments incorporate the selective access to communication technology into their strategies of coercion.*

FIGURE 4 Expected Proportion (83% and 95% confidence intervals) of Targeted Killings, Given Internet Accessibility and Different Levels of Government Control



Does social media cause political radicalization?

Mitts (2019), Müller & Schwarz (2018)

1. Examines whether anti-Muslim hostility is tied to online radicalization (Mitts 2019)
2. Investigates whether anti-refugee sentiment on Facebook results in increases in hate crimes against refugees (Müller & Schwarz 2018)

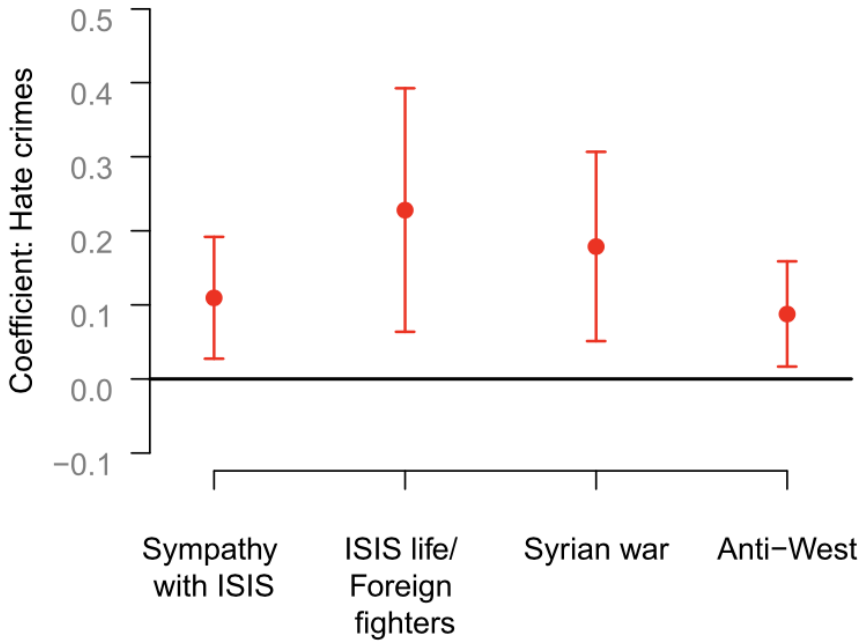
From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West

TAMAR MITTS *Columbia University*

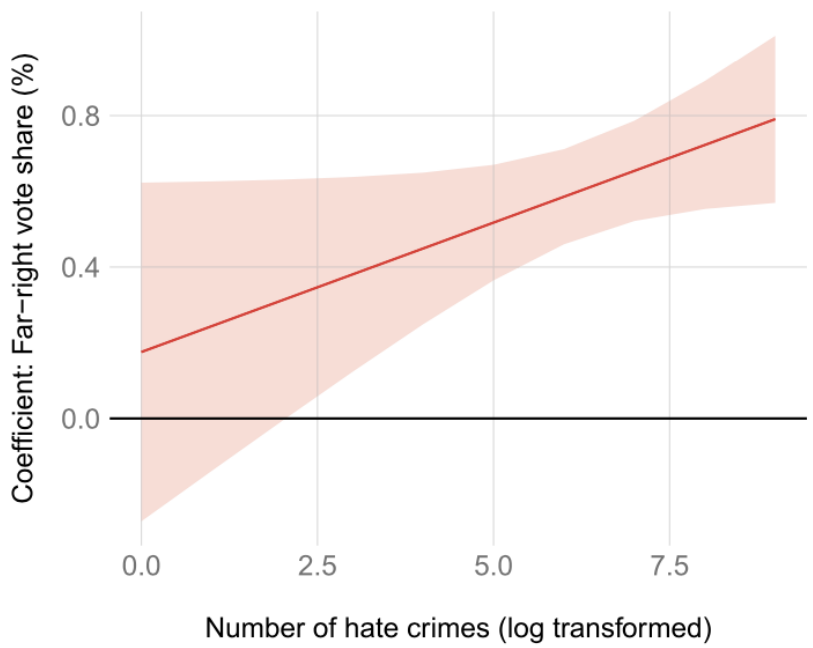
What explains online radicalization and support for ISIS in the West? Over the past few years, thousands of individuals have radicalized by consuming extremist content online, many of whom eventually traveled overseas to join the Islamic State. This study examines whether anti-Muslim hostility might drive pro-ISIS radicalization in Western Europe. Using new geo-referenced data on the online behavior of thousands of Islamic State sympathizers in France, the United Kingdom, Germany, and Belgium, I study whether the intensity of anti-Muslim hostility at the local level is linked to pro-ISIS radicalization on Twitter. The results show that local-level measures of anti-Muslim animosity correlate significantly and substantively with indicators of online radicalization, including posting tweets sympathizing with ISIS, describing life in ISIS-controlled territories, and discussing foreign fighters. High-frequency data surrounding events that stir support for ISIS—terrorist attacks, propaganda releases, and anti-Muslim protests—show the same pattern.

FIGURE 5. Far-Right Vote Share, Hate Crimes, and Support for ISIS in the U.K.

(A) Hate crimes and pro-ISIS and anti-West rhetoric



(B) The interaction between far-right vote share and hate crimes



Note: Panel (A) shows the correlation between the number of hate crimes and pro-ISIS tweeting in the U.K. Panel (B) presents the interaction between far-right vote share and hate crimes in a regression where the dependent variable combines all pro-ISIS tweets posted by users in the U.K.

FANNING THE FLAMES OF HATE: SOCIAL MEDIA AND HATE CRIME

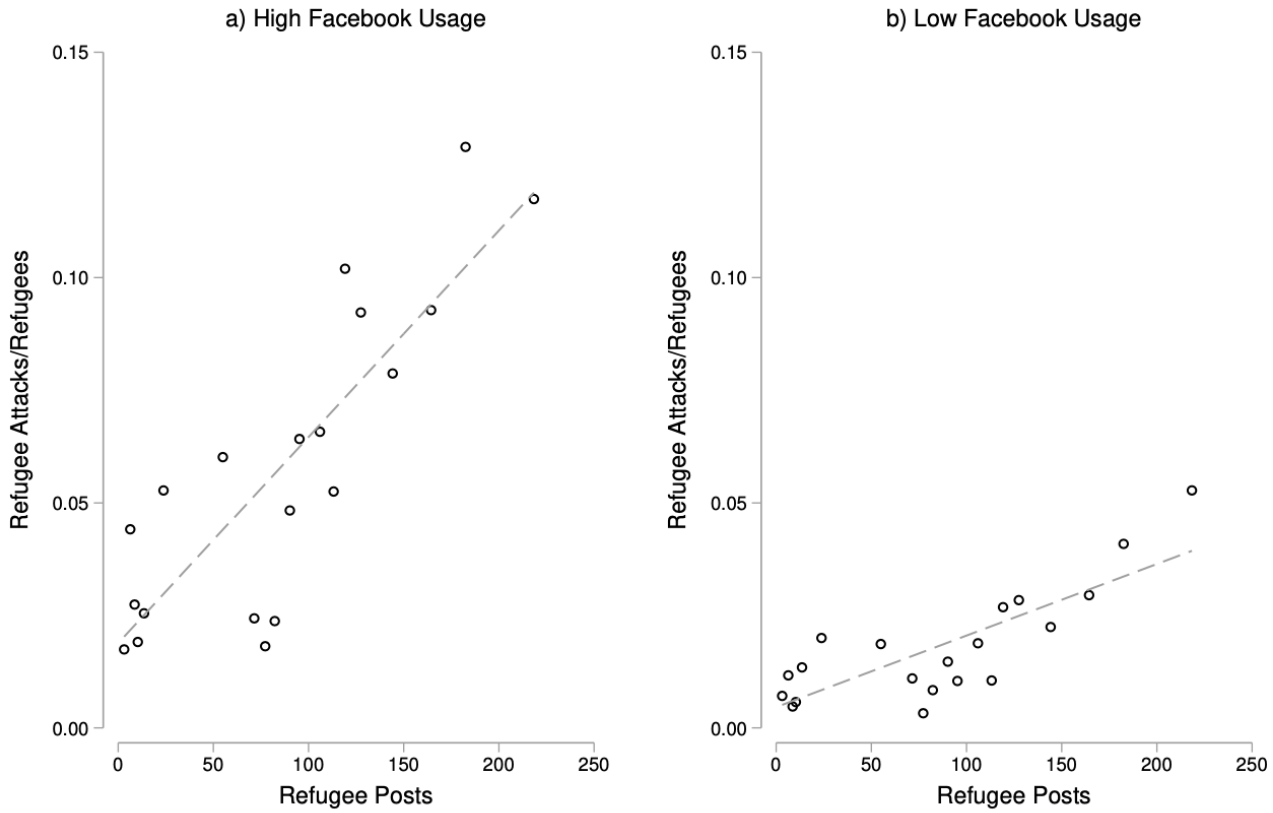
Karsten Müller
Princeton University, Department of Economics

Carlo Schwarz
Bocconi University, Department of Economics

Abstract

This paper investigates the link between social media and hate crime. We show that antirefugee sentiment on Facebook predicts crimes against refugees in otherwise similar municipalities with higher social media usage. To establish causality, we exploit exogenous variation in the timing of major Facebook and internet outages. Consistent with a role for "echo chambers," we find that right-wing social media posts contain narrower and more loaded content than news reports. Our results suggest that social media can act as a propagation mechanism for violent crimes by enabling the spread of extreme viewpoints. (JEL: D74, J15, Z10, D72, O35)

Figure 4: Exposure to Anti-Refugee Sentiment and Hate Crimes



Notes: This figure plots the average number of anti-refugee attacks against our measure of anti-refugee sentiment for municipalities below and above the median of $AfD\ Users/Pop$. Refugee attacks are binned by 20 quantiles of refugee posts. The number of anti-refugee attacks was residualized with respect to municipal population.

Are social media algorithms biased?

Bakshy et al. (2015)

- Examine whether YouTube recommendations on one video lead to increasingly radical videos and recommendations (Haroon et al. 2023)
- Examine whether manipulating YouTube feeds causes attitude change (Liu et al. 2023)
- Investigate whether ideology of news exposure differs under Facebook's algorithm versus all potential content shared by friends (Bakshy et al. 2015)

PNAS RESEARCH ARTICLE | COMPUTER SCIENCES POLITICAL SCIENCES OPEN ACCESS

Auditing YouTube's recommendation system for ideologically congenial, extreme, and problematic recommendations

Muhammad Haroon[a,1], Magdalena Wojcieszak[b,1], Anshuman Chhabra[a], Xin Liu[a], Prasant Mohapatra[a], and Zubair Shafiq[a]

Edited by Douglas Massey, Princeton University, Princeton, NJ; received August 1, 2022; accepted September 21, 2023

Algorithms of social media platforms are often criticized for recommending ideologically congenial and radical content to their users. Despite these concerns, evidence on such filter bubbles and rabbit holes of radicalization is inconclusive. We conduct an audit of the platform using 100,000 sock puppets that allow us to systematically and at scale isolate the influence of the algorithm in recommendations. We test 1) whether recommended videos are congenial with regard to users' ideology, especially deeper in the watch trail and whether 2) recommendations deeper in the trail become progressively more extreme and come from problematic channels. We find that YouTube's algorithm recommends congenial content to its partisan users, although some moderate and cross-cutting exposure is possible and that congenial recommendations increase deeper in the trail for right-leaning users. We do not find meaningful increases in ideological extremity of recommendations deeper in the trail, yet we show that a growing proportion of recommendations comes from channels categorized as problematic (e.g., "IDW," "Alt-right," "Conspiracy," and "QAnon"), with this increase being most pronounced among the very-right users. Although the proportion of these problematic recommendations is low (max of 2.5%), they are still encountered by over 36.1% of users and up to 40% in the case of very-right users.

YouTube | congeniality | recommendation algorithms | extremity | ideology

Significance

YouTube's algorithm is often accused of putting users in filter bubbles and generating rabbit holes of radicalization. However, evidence on these issues is inconclusive. We conduct a systematic audit of the platform using 100,000 sock puppets that allow us to isolate the influence of the algorithm in recommendations to ideologically congenial and increasingly extreme and problematic videos. YouTube's algorithm recommends

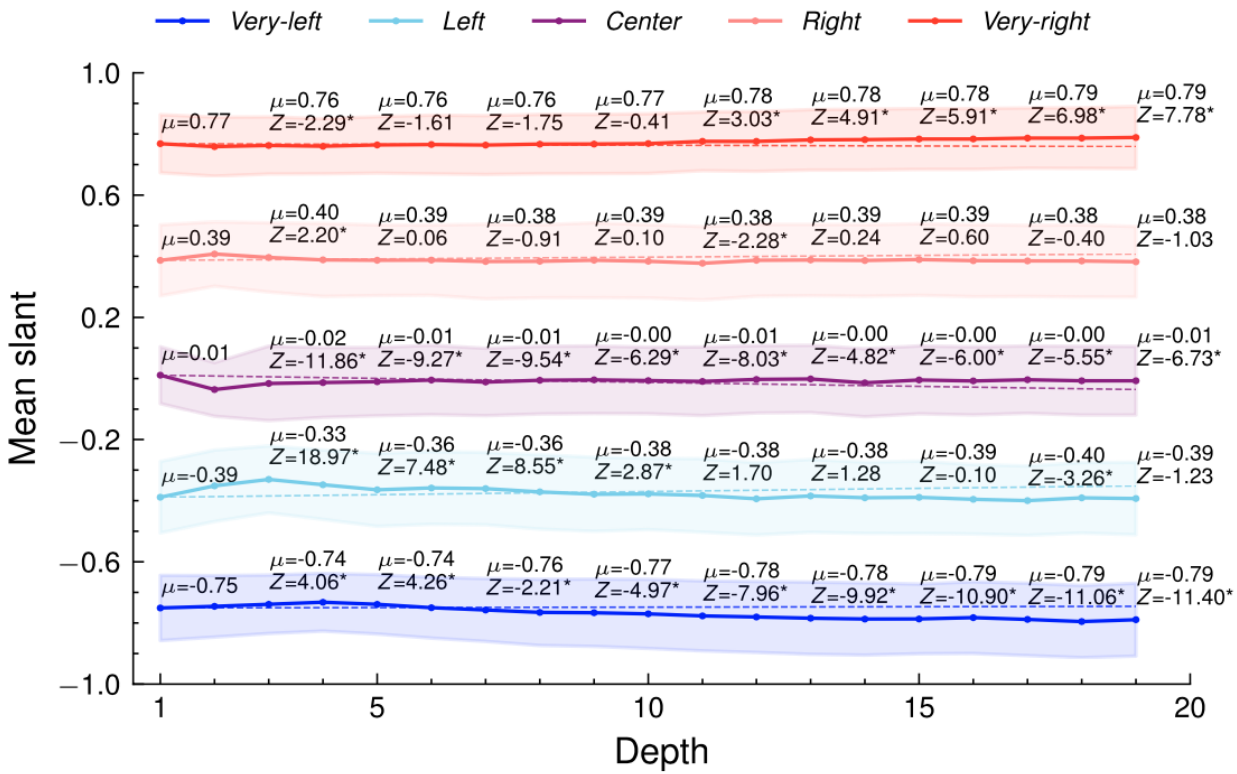


Fig. 5. The mean slant of the ideologically congenial videos watched by each ideology category in the up-next recommendations. For example, the *Top* red line corresponds to the mean slant of the very-right videos watched by the very-right sock puppet. The dotted and dashed lines correspond to 1-SD and mean slant at depth 1 of the trail, respectively. The mean and SD become progressively extreme for the very-left and very-right sock puppets.

Algorithmic recommendations have limited effects on polarization: A naturalistic experiment on YouTube*†

Naijia Liu[1], Matthew A. Baum[2], Adam J. Berinsky[3], Allison J.B. Chaney[4], Justin de Benedictis-Kessner[2], Andy Guess[5], Dean Knox[6], Christopher Lucas[7], Rachel Mariman[8], and Brandon M. Stewart[9]

[1]Department of Government, Harvard University
[2]John F. Kennedy School of Government, Harvard University
[3]Department of Political Science, Massachusetts Institute of Technology
[4]Fuqua School of Business, Duke University
[5]Department of Politics and School of Public and International Affairs, Princeton University
[6]Operations, Information, and Decisions Department, the Wharton School of the University of Pennsylvania
[7]Department of Political Science, Washington University in St. Louis
[8]Analytics at Wharton, the Wharton School of the University of Pennsylvania
[9]Department of Sociology and Office of Population Research, Princeton University

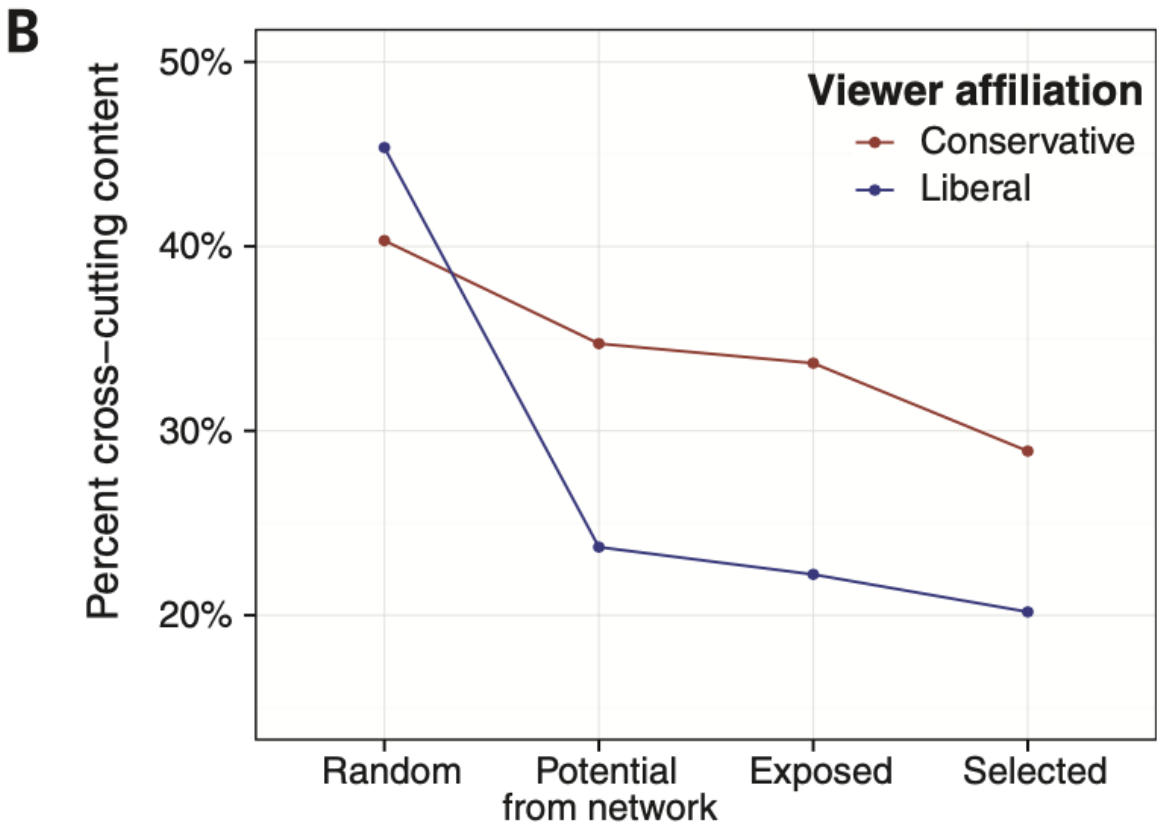
September 18, 2023

Exposure to ideologically diverse news and opinion on Facebook

Eytan Bakshy,[1*†] Solomon Messing,[1†] Lada A. Adamic[1,2]

Exposure to news, opinion, and civic information increasingly occurs through social media. How do these online networks influence exposure to perspectives that cut across ideological lines? Using deidentified data, we examined how 10.1 million U.S. Facebook users interact with socially shared news. We directly measured ideological homophily in friend networks and examined the extent to which heterogeneous friends could potentially expose individuals to cross-cutting content. We then quantified the extent to which individuals encounter comparatively more or less diverse content while interacting via Facebook's algorithmically ranked News Feed and further studied users' choices to click through to ideologically discordant content. Compared with algorithmic ranking, individuals' choices played a stronger role in limiting exposure to cross-cutting content.

Fig. 3. Cross-cutting content at each stage in the diffusion process. (**A**) Illustration of how algorithmic ranking and individual choice affect the proportion of ideologically cross-cutting content that individuals encounter. Gray circles illustrate the content present at each stage in the media exposure process. Red circles indicate conservatives, and blue circles indicate liberals. (**B**) Average ideological diversity of content (i) shared by random others (random), (ii) shared by friends (potential from network), (iii) actually appeared in users' News Feeds (exposed), and (iv) users clicked on (selected).



Do women politicians face more toxicity on social media?

Rheault et al. (2019)

- Surprisingly little systematic study
- Rheault et al. (2019) examine the proportion of incivility directed at US Senators and Canadian MPs on Twitter

Politicians in the line of fire: Incivility and the treatment of women on social media

Research and Politics
January- March 2019: 1–7

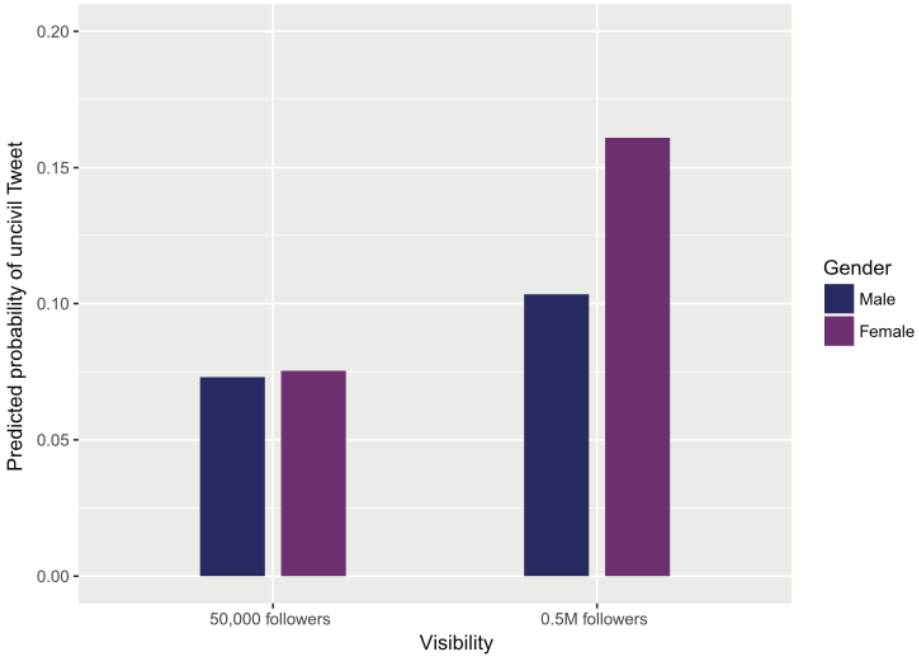
DOI: 10.1177/2053168018816228
journals.sagepub.com/home/rap

Ludovic Rheault, Erica Rayment and Andreea Musulan

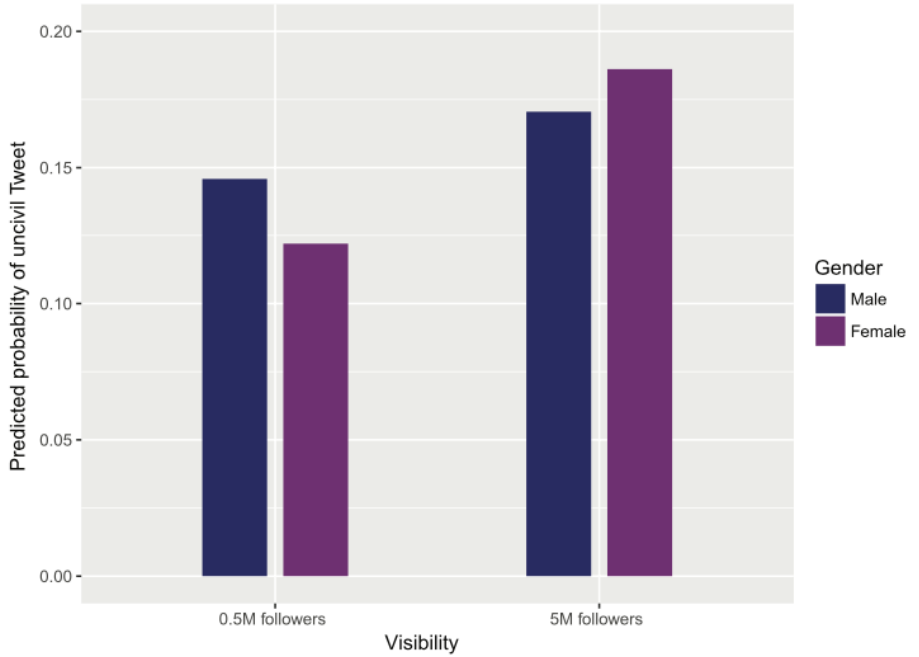
Abstract

A seemingly inescapable feature of the digital age is that people choosing to devote their lives to politics must now be ready to face a barrage of insults and disparaging comments targeted at them through social media. This article represents an effort to document this phenomenon systematically. We implement machine learning models to predict the incivility of about 2.2 m messages addressed to Canadian politicians and US Senators on Twitter. Specifically, we test whether women in politics are more heavily targeted by online incivility, as recent media reports suggested. Our estimates indicate that roughly 15% of public messages sent to Senators can be categorized as uncivil, whereas the proportion is about four points lower in Canada. We find evidence that women are more heavily targeted by uncivil messages than men, although only among highly visible politicians.

(a) Canada



(b) USA



What interventions prevent prejudice and toxicity online?

Munger (2017), Chandrasekharan et al. (2017)

1. Experimentally tests whether social sanctioning of racial abuse is effective (Munger 2017)
2. Investigate whether banning racist subreddits drive racists away (Chandrasekharan et al. 2017)

Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment

Kevin Munger[1]

Published online: 11 November 2016

Abstract I conduct an experiment which examines the impact of group norm promotion and social sanctioning on racist online harassment. Racist online harassment de-mobilizes the minorities it targets, and the open, unopposed expression of racism in a public forum can legitimize racist viewpoints and prime ethnocentrism. I employ an intervention designed to reduce the use of anti-black racist slurs by white men on Twitter. I collect a sample of Twitter users who have harassed other users and use accounts I control ("bots") to sanction the harassers. By varying the identity of the bots between in-group (white man) and out-group (black man) and by varying the number of Twitter followers each bot has, I find that subjects who were sanctioned by a high-follower white male significantly reduced their use of a racist slur. This paper extends findings from lab experiments to a naturalistic setting using an objective, behavioral outcome measure and a continuous 2-month data collection period. This represents an advance in the study of prejudiced behavior.

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

You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech

ESHWAR CHANDRASEKHARAN, Georgia Institute of Technology
UMASHANTHI PAVALANATHAN, Georgia Institute of Technology
ANIRUDH SRINIVASAN, Georgia Institute of Technology
ADAM GLYNN, Emory University
JACOB EISENSTEIN, Georgia Institute of Technology
ERIC GILBERT, University of Michigan

In 2015, Reddit closed several subreddits—foremost among them r/fatpeoplehate and r/CoonTown—due to violations of Reddit's anti-harassment policy. However, the effectiveness of banning as a moderation approach remains unclear: banning might diminish hateful behavior, or it may relocate such behavior to different parts of the site. We study the ban of r/fatpeoplehate and r/CoonTown in terms of its effect on both participating users and affected subreddits. Working from over 100M Reddit posts and comments, we generate hate speech lexicons to examine variations in hate speech usage via causal inference methods. We find that the *ban worked for Reddit.* More accounts than expected discontinued using the site; those that stayed drastically decreased their hate speech usage—by at least 80%. Though many subreddits saw an influx of r/fatpeoplehate and r/CoonTown "migrants," those subreddits saw no significant changes in hate speech usage. In other words, other subreddits did not inherit the problem. We conclude by reflecting on the apparent success of the ban, discussing implications for online moderation, Reddit and internet communities more broadly.

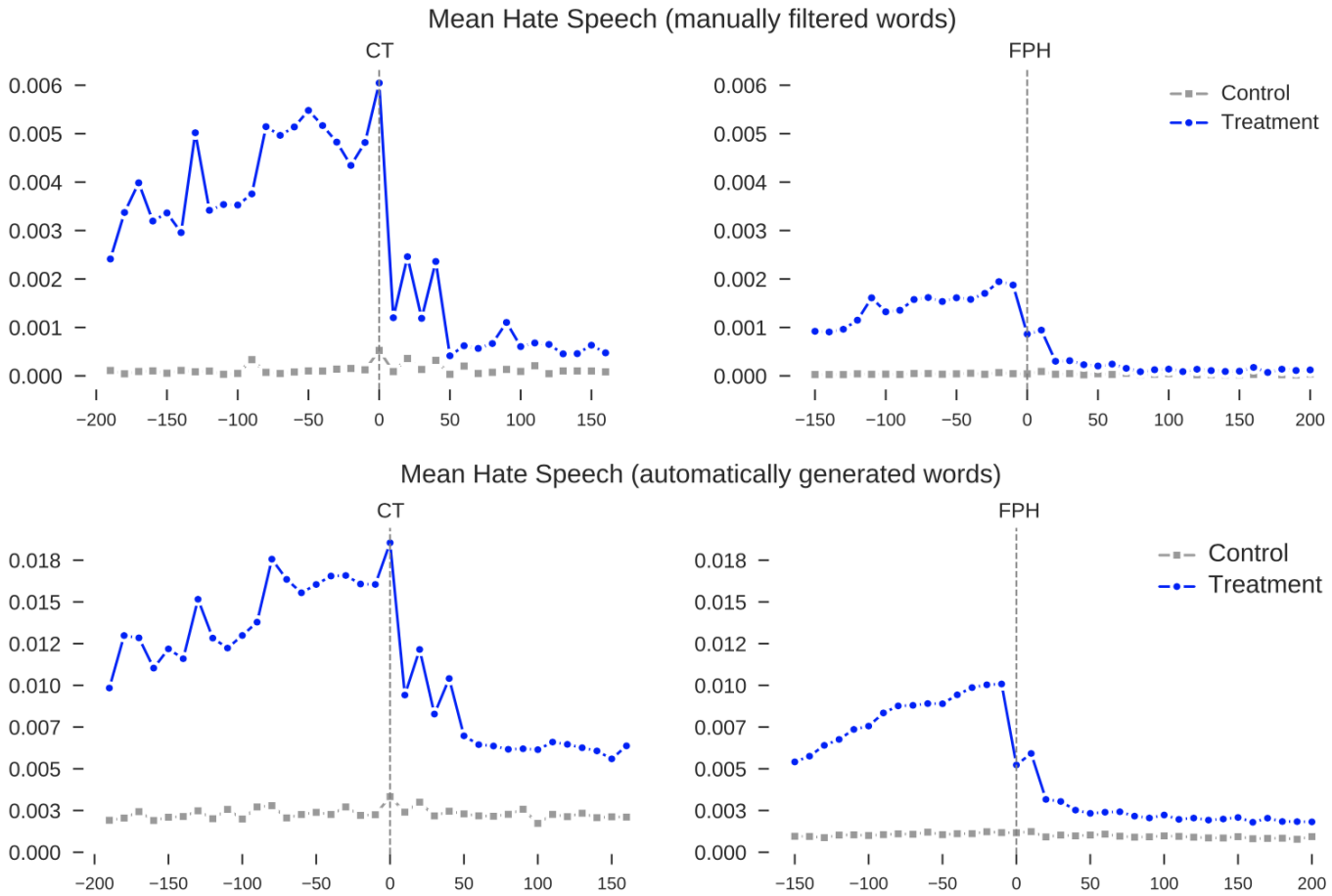


Fig. 3. Variations in users' hate speech usage on Reddit. We compute mean hate speech usage on Reddit by treatment and control users across all of 2015, using time-windows of 10 days, before and after the respective bans. We calculate hate speech usage as the sum of individual frequencies of each term in the hate lexicon and normalize it per post.

Do people live in political echo chambers?

Flaxman et al. (2016) & Messing & Westwood (2014)

1. Examine the ideology of news media linked by social media, search, and direct browsing (Flaxman et al. 2016)
2. Test experimentally whether news media organizations or social media endorsements drive exposure (Messing & Westwood 2014)

FILTER BUBBLES, ECHO CHAMBERS, AND ONLINE NEWS CONSUMPTION

SETH FLAXMAN*
SHARAD GOEL
JUSTIN M. RAO

Abstract Online publishing, social networks, and web search have dramatically lowered the costs of producing, distributing, and discovering news articles. Some scholars argue that such technological changes increase exposure to diverse perspectives, while others worry that they increase ideological segregation. We address the issue by examining web-browsing histories for 50,000 US-located users who regularly read online news. We find that social networks and search engines are associated with an increase in the mean ideological distance between individuals. However, somewhat counterintuitively, these same channels also are associated with an increase in an individual's exposure to material from his or her less preferred side of the political spectrum. Finally, the vast majority of online news consumption is accounted for by individuals simply visiting the home pages of their favorite, typically mainstream, news outlets, tempering the consequences—both positive and negative—of recent technological changes. We thus uncover evidence for both sides of the debate, while also finding that the magnitude of the effects is relatively modest.

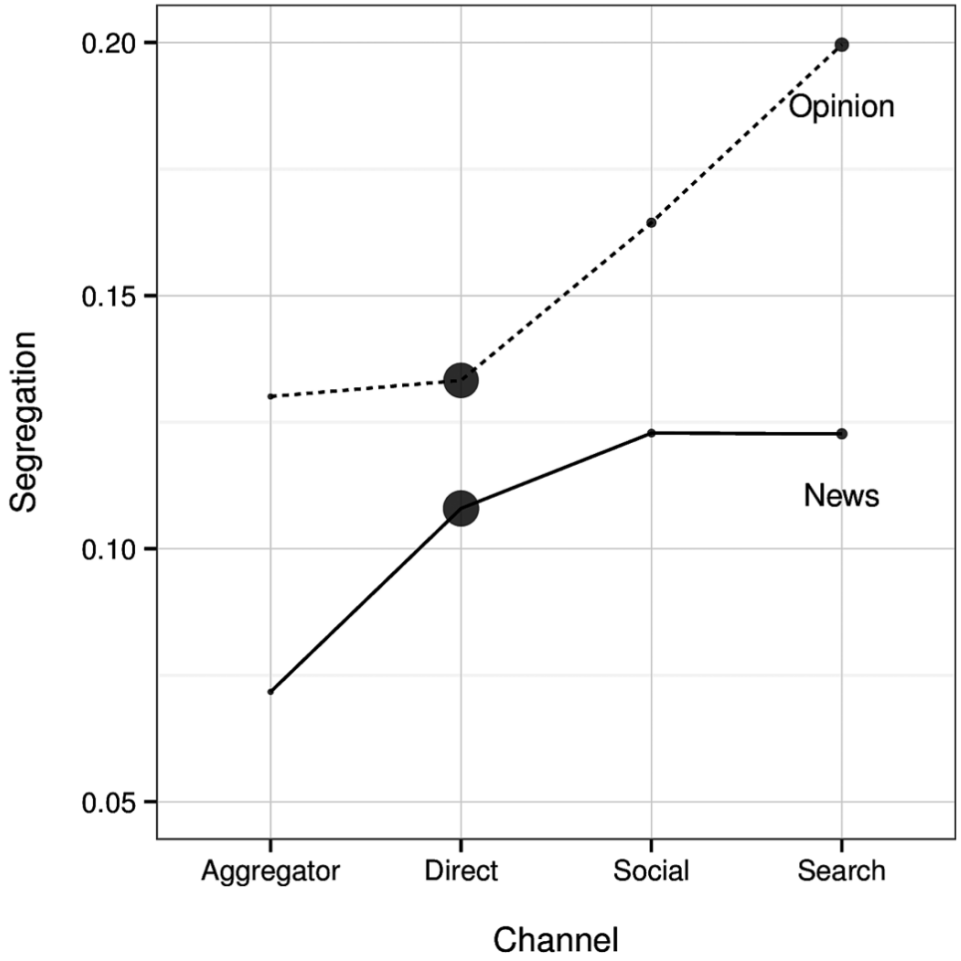


Figure 3. Estimates of Ideological Segregation across Consumption Channels. Point sizes indicate traffic fraction, normalized separately within the news and opinion lines.

Communication Research
2014, Vol. 41(8) 1042–1063

DOI: 10.1177/0093650212466406
crx.sagepub.com
SAGE

Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online

Solomon Messing[1] and Sean J. Westwood[1]

Abstract

Much of the literature on polarization and selective exposure presumes that the internet exacerbates the fragmentation of the media and the citizenry. Yet this ignores how the widespread use of social media changes news consumption. Social media provide readers a choice of stories from different sources that come recommended from politically heterogeneous individuals, in a context that emphasizes social value over partisan affiliation. Building on existing models of news selectivity to emphasize information utility, we hypothesize that social media's distinctive feature, social endorsements, trigger several decision heuristics that suggest utility. In two experiments, we demonstrate that stronger social endorsements increase the probability that people select content and that their presence reduces partisan selective exposure to levels indistinguishable from chance.

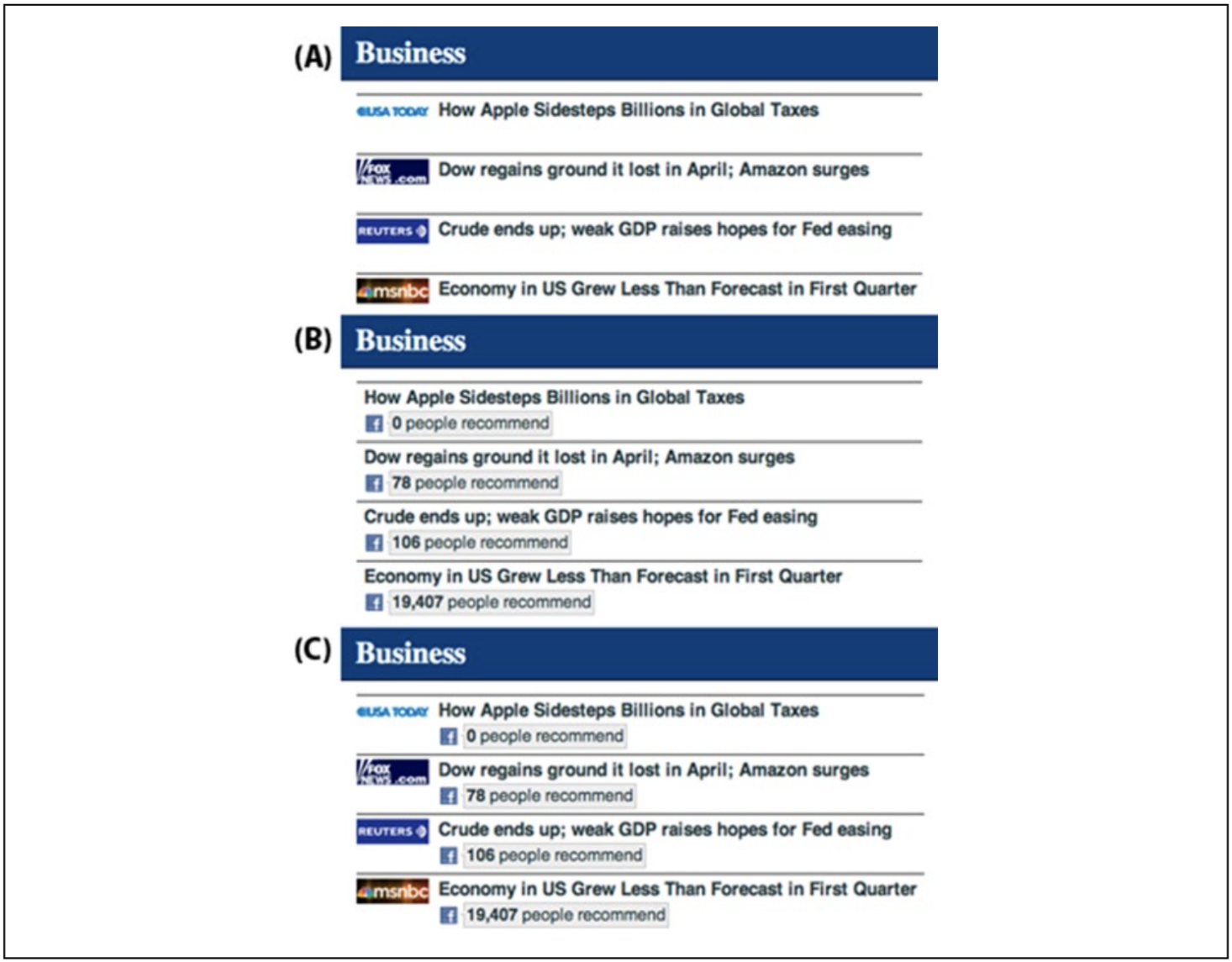


Figure 1. Design for Study 1.

Is social media politically polarizing?

Boxell et al. (2017) & Bail et al. (2018)

1. Examine increases in polarization by demographics groups, and their Internet and social media use (Boxell et al. 2017)
2. Test experimentally whether manipulating the ideological content of people's Twitter feeds increases or decreases polarization (Bail et al. 2018)

Greater Internet use is not associated with faster growth in political polarization among US demographic groups

Levi Boxell[a,1], Matthew Gentzkow[a,b,c], and Jesse M. Shapiro[c,d]

[a]Department of Economics, Stanford University, Stanford, CA 94305; [b]Stanford Institute for Economic Policy Research, Stanford University, Stanford, CA 94305; [c]National Bureau of Economic Research, Cambridge, MA 02138; and [d]Economics Department, Brown University, Providence, RI 02912

Edited by Larry M. Bartels, Vanderbilt University, Nashville, TN, and approved August 16, 2017 (received for review April 20, 2017)

We combine eight previously proposed measures to construct an index of political polarization among US adults. We find that polarization has increased the most among the demographic groups least likely to use the Internet and social media. Our overall index and all but one of the individual measures show greater increases for those older than 65 than for those aged 18–39. A linear model estimated at the age-group level implies that the Internet explains a small share of the recent growth in polarization.

politics | polarization | Internet | social media

work that analyzes demographic differences to evaluate the role of the Internet in the 2016 presidential election outcome (17).

We divide respondents according to demographics that predict Internet and social media use. The main predictor we focus on is age. We show, using data from the ANES and the Pew Research Center, that Internet and social media use rates are far higher among the young than the old, with rates of social media use in 2016 of 0.88, 0.65, and 0.30, respectively, among those aged 18–39, 40–64, and 65 and older (65+).

A normalized index of our eight polarization measures

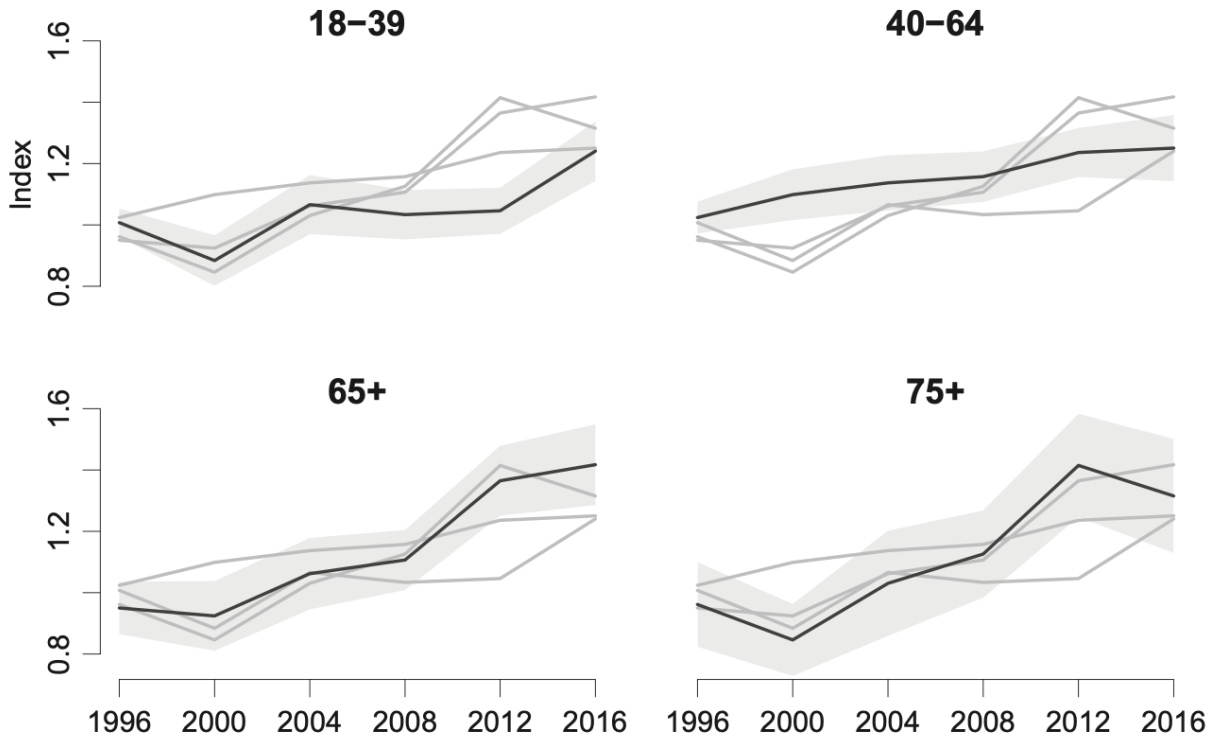


Fig. 3. Trends in polarization by age group. Each plot shows the polarization index for each of four age groups. Each plot highlights the series for one age group in bold. Shaded regions represent 95% pointwise CIs for the bold series constructed from a nonparametric bootstrap with 100 replicates. See main text for definitions and *SI Appendix*, section 3 for details on the bootstrap procedure.

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

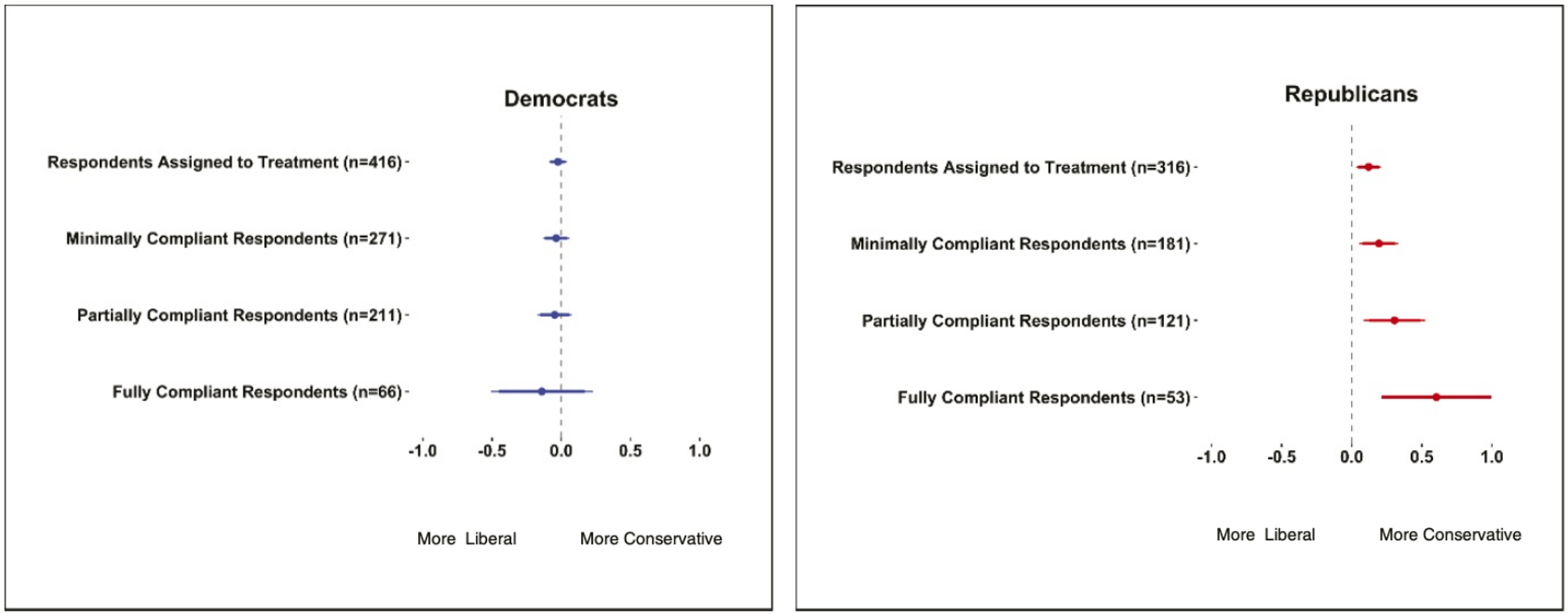


Fig. 3. Effect of following Twitter bots that retweet messages by elected officials, organizations, and opinion leaders with opposing political ideologies for 1 mo, on a seven-point liberal/conservative scale where larger values indicate more conservative opinions about social policy issues, for experiments with Democrats (n = 697) and Republicans (n = 542). Models predict posttreatment liberal/conservative scale score and control for pretreatment score on this scale as well as 12 other covariates described in *SI Appendix*. Circles describe unstandardized point estimates, and bars describe 90% and 95% confidence intervals. "Respondents Assigned to Treatment" describes the ITT effect for Democrats (ITT = —0.02, t = —0.76, p = 0.45, n = 416) and Republicans (ITT = 0.12, t = 2.68, p = 0.008, n = 316). "Minimally-Compliant Respondents" describes the CACE for respondents who followed one of the study's bots for Democrats (CACE = —0.04, t = —0.75, p = 0.45, n of compliant respondents = 271) and Republicans (CACE = 0.19, t = 2.73, $p <$ 0.007, n of compliant respondents = 181). "Partially-Compliant Respondents" describes the CACE for respondents who correctly answered at least one question, but not all questions, about the content of a bot's tweets during weekly surveys throughout the study period for Democrats (CACE = —0.05, t = —0.75, p = 0.45, n of compliant respondents = 211) and Republicans (CACE = 0.31, t = 2.73, $p <$.007, n of compliant respondents = 121). "Fully-Compliant Respondents" describes the CACE for respondents who answered all questions about the content of the bot's tweets correctly for Democrats (CACE = —0.14, t = —0.75, p = 0.46, n of compliant respondents = 66) and Republicans (CACE = 0.60, t = 2.53, $p <$ 0.01, n of compliant respondents = 53). Although treated Democrats exhibited slightly more liberal attitudes posttreatment that increase in size with level of compliance, none of these effects were statistically significant. In contrast, treated Republicans exhibited substantially more conservative views posttreatment that increase in size with level of compliance, and these effects are highly significant.

How widespread is fake news, and who shares it?

Grinberg et al. (2019) & Guess et al. (2019)

1. Examines who shares and is exposed to fake news on Twitter (Grinberg et al. 2019)
2. Examines who shares fake news on Facebook (Guess et al. 2019)

Fake news on Twitter during the 2016 U.S. presidential election

**Nir Grinberg[1,2]*, Kenneth Joseph[3]*, Lisa Friedland[1]*,
Briony Swire-Thompson[1,2], David Lazer[1,2]†**

The spread of fake news on social media became a public concern in the United States after the 2016 presidential election. We examined exposure to and sharing of fake news by registered voters on Twitter and found that engagement with fake news sources was extremely concentrated. Only 1% of individuals accounted for 80% of fake news source exposures, and 0.1% accounted for nearly 80% of fake news sources shared. Individuals most likely to engage with fake news sources were conservative leaning, older, and highly engaged with political news. A cluster of fake news sources shared overlapping audiences on the extreme right, but for people across the political spectrum, most political news exposure still came from mainstream media outlets.

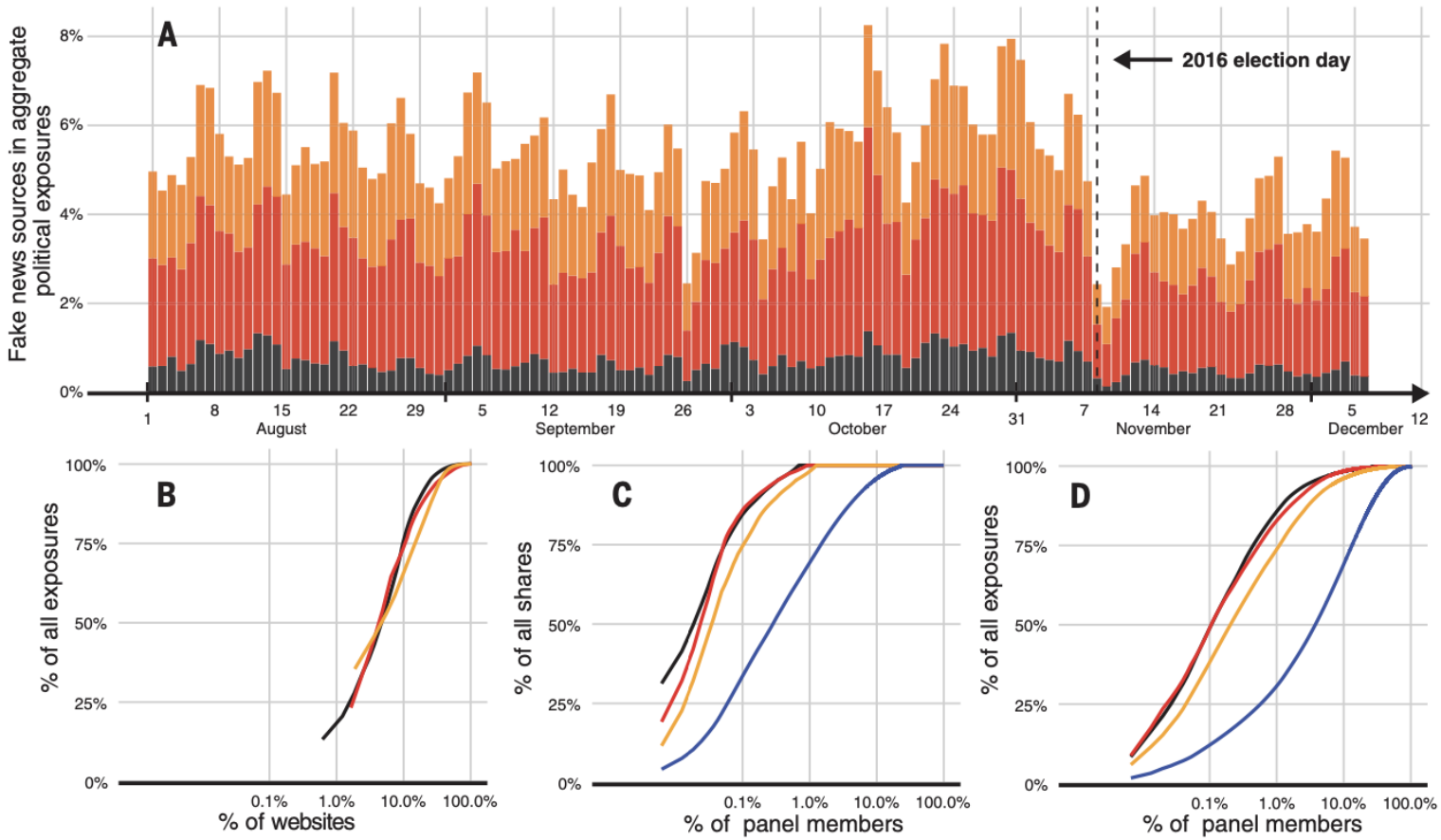


Fig. 1. Prevalence over time and concentration of fake news sources. (**A**) Daily percentage of exposures to black, red, and orange fake news sources, relative to all exposures to political URLs. Exposures were summed across all panel members. (**B** to **D**) Empirical cumulative distribution functions showing distribution of exposures among websites (B), distribution of shares by panel members (C), and distribution of exposures among panel members (D). The *x* axis represents percentage of websites or panel members responsible for a given percentage (*y* axis) of all exposures or shares. Black, red, and orange lines represent fake news sources; blue line denotes all other sources. This distribution was not comparable for (B) because of the much larger number of sources in its tail and the fundamentally different selection process involved.

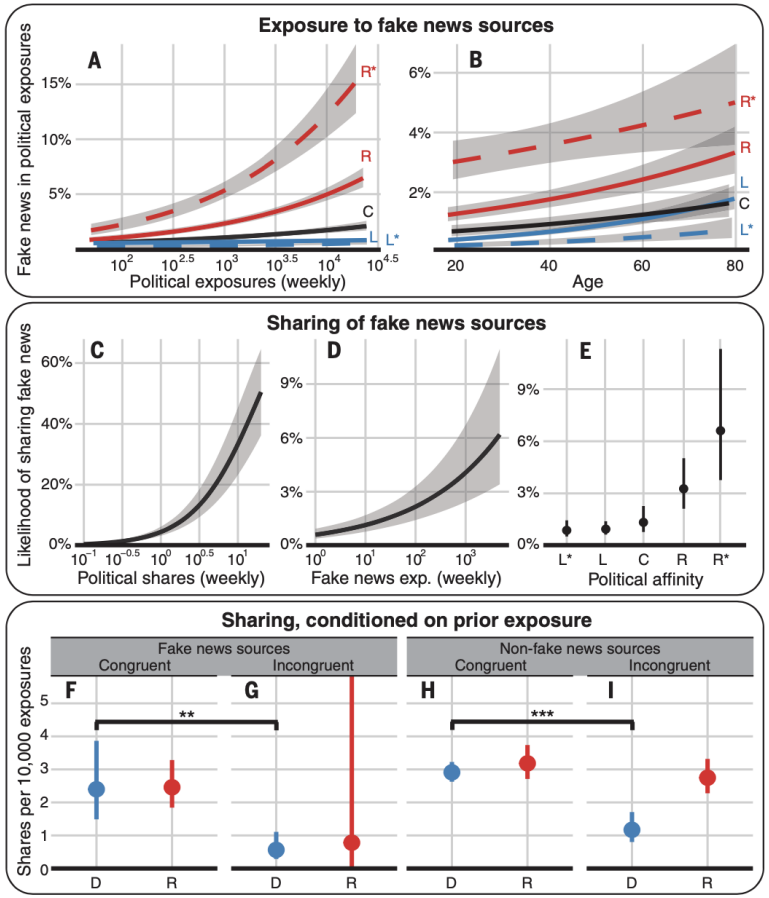


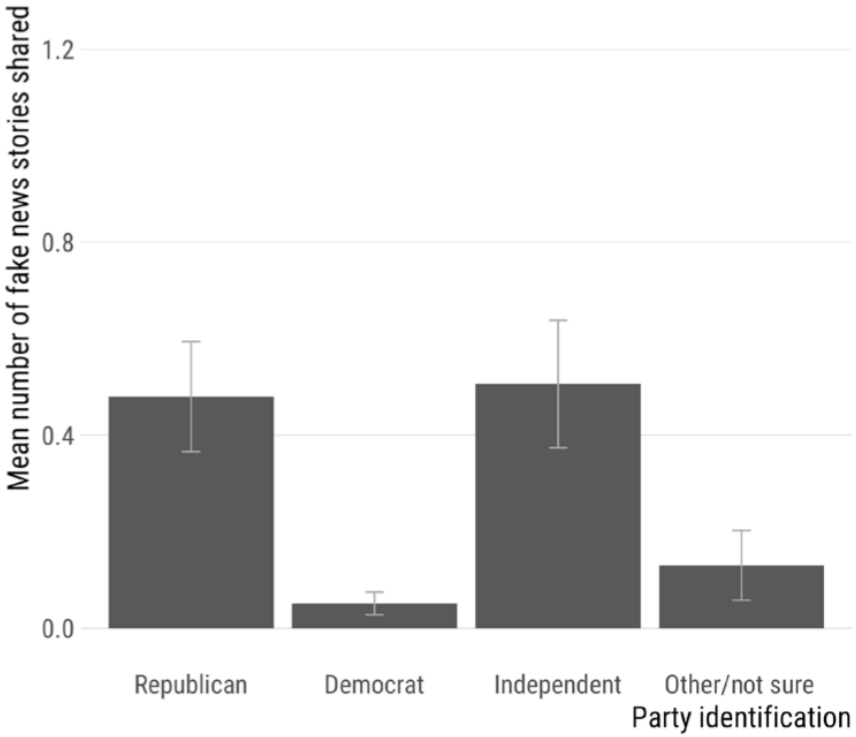
Fig. 4. Key individual characteristics associated with exposure to and sharing of fake news sources. The proportion of an individual's political exposures coming from fake news sources as a function of (**A**) number of political exposures, excluding fake news sources, and (**B**) age. Estimates are based on binomial regression models fitted separately to each political affinity subgroup. Blue, liberal; black, center; red, conservative. (**C** to **E**) An individual's likelihood of sharing one or more URLs from fake news sources as a function of (C) number of shares of political URLs, (D) number of exposures to fake news sources, and (E) political affinity. (**F** to **I**) Likelihood of a liberal (D) or conservative (R) individual sharing a political URL to which they have been exposed, depending on the political congruency and veracity of the source: (F) congruent and fake, (G) incongruent and fake, (H) congruent and nonfake, and (I) incongruent and nonfake. Brackets indicate significantly different pairs: $^{**}P < 0.01$, $^{***}P < 0.001$. All estimates and 95% CIs [gray shaded regions in (A) to (D); line segments in (E) to (I)] are based on regression models specified in SM S.11 to S.13, with the remaining model variables held constant to their median or most common level.

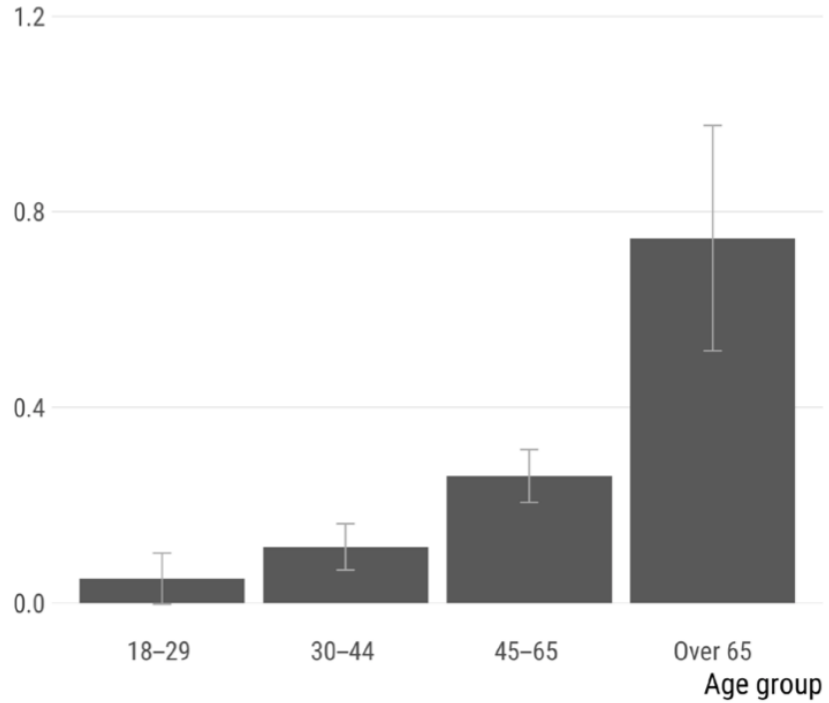
SOCIAL SCIENCES

Less than you think: Prevalence and predictors of fake news dissemination on Facebook

Andrew Guess[1]*, Jonathan Nagler[2], Joshua Tucker[2]

So-called "fake news" has renewed concerns about the prevalence and effects of misinformation in political campaigns. Given the potential for widespread dissemination of this material, we examine the individual-level characteristics associated with sharing false articles during the 2016 U.S. presidential campaign. To do so, we uniquely link an original survey with respondents' sharing activity as recorded in Facebook profile data. First and foremost, we find that sharing this content was a relatively rare activity. Conservatives were more likely to share articles from fake news domains, which in 2016 were largely pro-Trump in orientation, than liberals or moderates. We also find a strong age effect, which persists after controlling for partisanship and ideology: On average, users over 65 shared nearly seven times as many articles from fake news domains as the youngest age group.


Mean number of fake news stories shared
1.2
0.8
0.4
0.0
Republican
Democrat
Independent
Other/not sure
Party identification


1.2
0.8
0.4
0.0
18–29
30–44
45–65
Over 65
Age group

What are the effects of social media on knowledge, well-being, and voting behavior?

Bond et al. (2012), Kramer et al. (2014), Allcott et al. (Forthcoming)

1. Experimentally test the effect of an "I voted" button on Facebook on voter turnout (Bond et al. 2012)
2. Experimentally remove negative and positive content from Facebook feeds to test effects on user emotions (Kramer et al. 2014)
3. Experimentally prevent people from using Facebook to test its effects on a wide range of political and non-political outcomes (Allcott et al. Forthcoming)

A 61-million-person experiment in social influence and political mobilization

Robert M. Bond[1], Christopher J. Fariss[1], Jason J. Jones[2], Adam D. I. Kramer[3], Cameron Marlow[3], Jaime E. Settle[1] & James H. Fowler[1,4]

Human behaviour is thought to spread through face-to-face social networks, but it is difficult to identify social influence effects in observational studies[9–13], and it is unknown whether online social networks operate in the same way[14–19]. Here we report results from a randomized controlled trial of political mobilization messages delivered to 61 million Facebook users during the 2010 US congressional elections. The results show that the messages directly influenced political self-expression, information seeking and real-world voting behaviour of millions of people. Furthermore, the messages not only influenced the users who received them but also the users' friends, and friends of friends. The effect of social transmission on real-world voting was greater than the direct effect of the messages themselves, and nearly all the transmission occurred between 'close friends' who were more likely to have a face-to-face relationship. These results suggest that strong ties are instrumental for spreading both online and real-world behaviour in human social networks.

with all users of at least 18 years of age in the United States who accessed the Facebook website on 2 November 2010, the day of the US congressional elections. Users were randomly assigned to a 'social message' group, an 'informational message' group or a control group. The social message group ($n = 60{,}055{,}176$) was shown a statement at the top of their 'News Feed'. This message encouraged the user to vote, provided a link to find local polling places, showed a clickable button reading 'I Voted', showed a counter indicating how many other Facebook users had previously reported voting, and displayed up to six small randomly selected 'profile pictures' of the user's Facebook friends who had already clicked the I Voted button (Fig. 1). The informational message group ($n = 611{,}044$) was shown the message, poll information, counter and button, but they were not shown any faces of friends. The control group ($n = 613{,}096$) did not receive any message at the top of their News Feed.

The design of the experiment allowed us to assess the impact that the treatments had on three user actions; clicking the I Voted button,

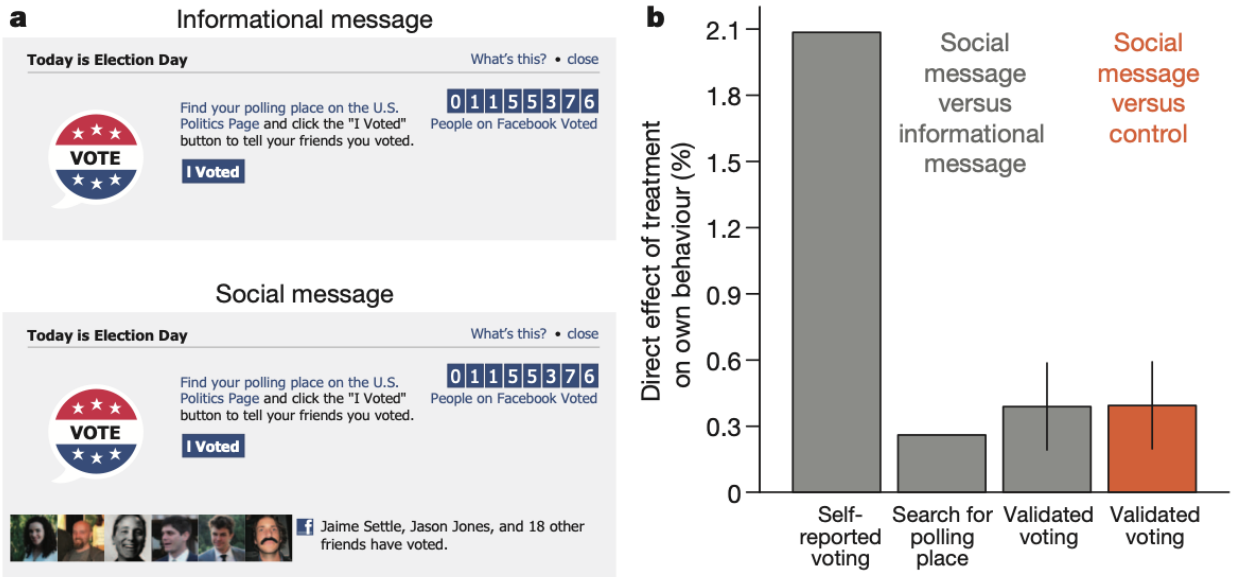


Figure 1 | The experiment and direct effects. **a**, **b**, Examples of the informational message and social message Facebook treatments (**a**) and their direct effect on voting behaviour (**b**). Vertical lines indicate s.e.m. (they are too small to be seen for the first two bars).

"To put these results in context, it is important to note that turnout has been steadily increasing in recent US midterm elections, from 36.3% of the voting age population in 2002 to 37.2% in 2006, and to 37.8% in 2010. **Our results suggest that the Facebook social message increased turnout directly by about 60,000 voters and indirectly through social contagion by another 280,000 voters, for a total of 340,000 additional votes.** That represents about 0.14% of the voting age population of about 236 million in 2010."

Experimental evidence of massive-scale emotional contagion through social networks

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

[a]Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; [b]Center for Tobacco Control Research and Education, University of California, San Francisco, CA 94143; and Departments of [c]Communication and [d]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.

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 =

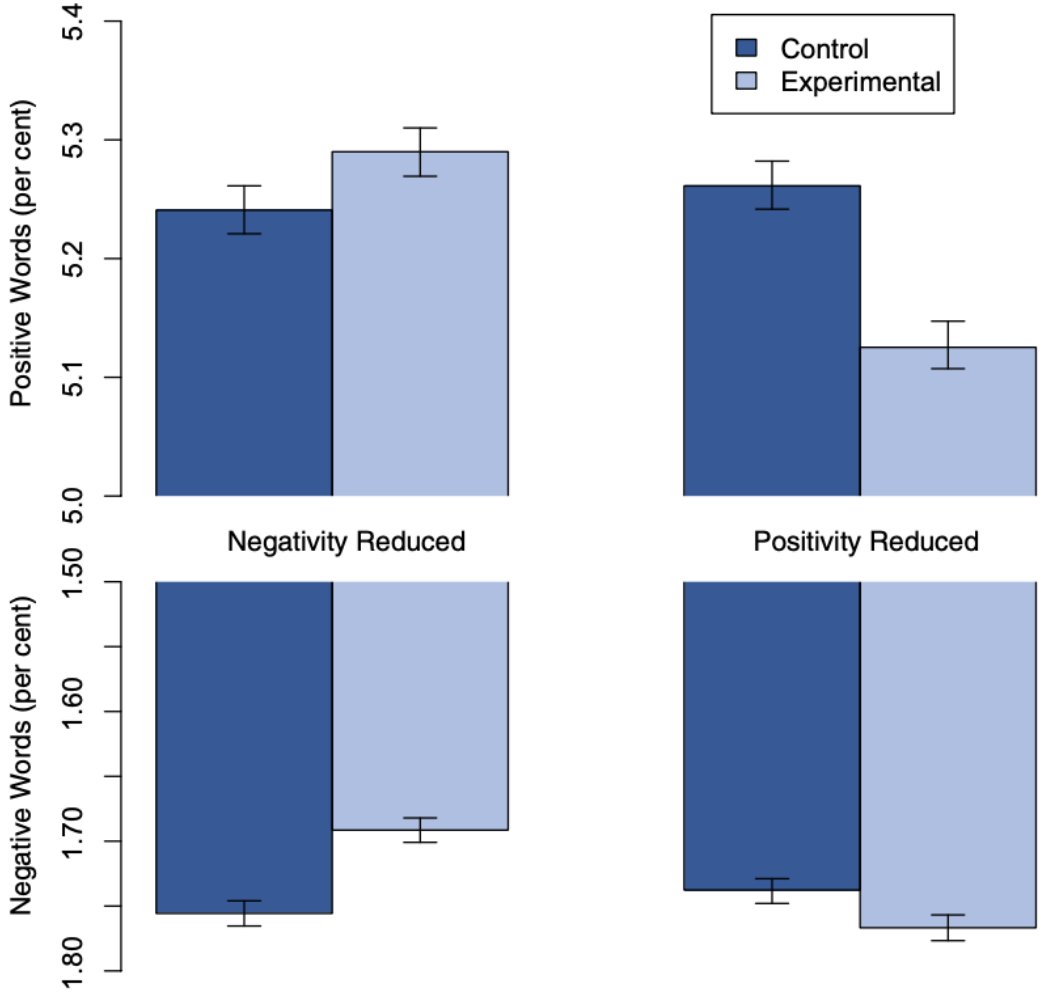


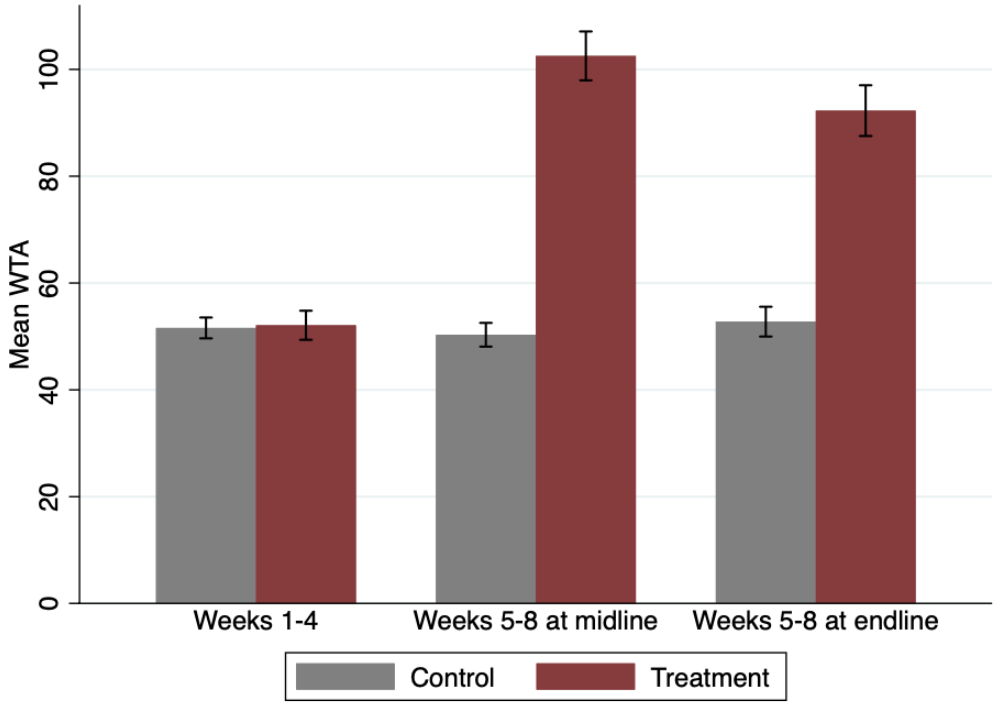
Fig. 1. Mean number of positive (*Upper*) and negative (*Lower*) emotion words (percent) generated people, by condition. Bars represent standard errors.

The Welfare Effects of Social Media†

By Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow*

The rise of social media has provoked both optimism about potential societal benefits and concern about harms such as addiction, depression, and political polarization. In a randomized experiment, we find that deactivating Facebook for the four weeks before the 2018 US midterm election (i) reduced online activity, while increasing offline activities such as watching TV alone and socializing with family and friends; (ii) reduced both factual news knowledge and political polarization; (iii) increased subjective well-being; and (iv) caused a large persistent reduction in post-experiment Facebook use. Deactivation reduced post-experiment valuations of Facebook, suggesting that traditional metrics may overstate consumer surplus.
(*JEL* D12, D72, D90, I31, L82, L86, Z13)

Figure 12: **Average Valuation of Facebook in Treatment and Control**



Notes: This figure presents the mean willingness-to-accept (WTA) to deactivate Facebook in Treatment and Control, for the impact evaluation sample: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered $p = \$102$ or $p = \$0$ to do so. The first pair of bars is the mean WTA for deactivation in weeks 1-4, the four weeks after the midline survey. The second pair of bars is mean WTA for deactivation in weeks 5-8, the four weeks after the endline survey, as elicited in the midline survey. The third pair of bars is mean WTA for deactivation in weeks 5-8, as elicited in the endline survey.

Is social media advertising beneficial or detrimental to democracy?

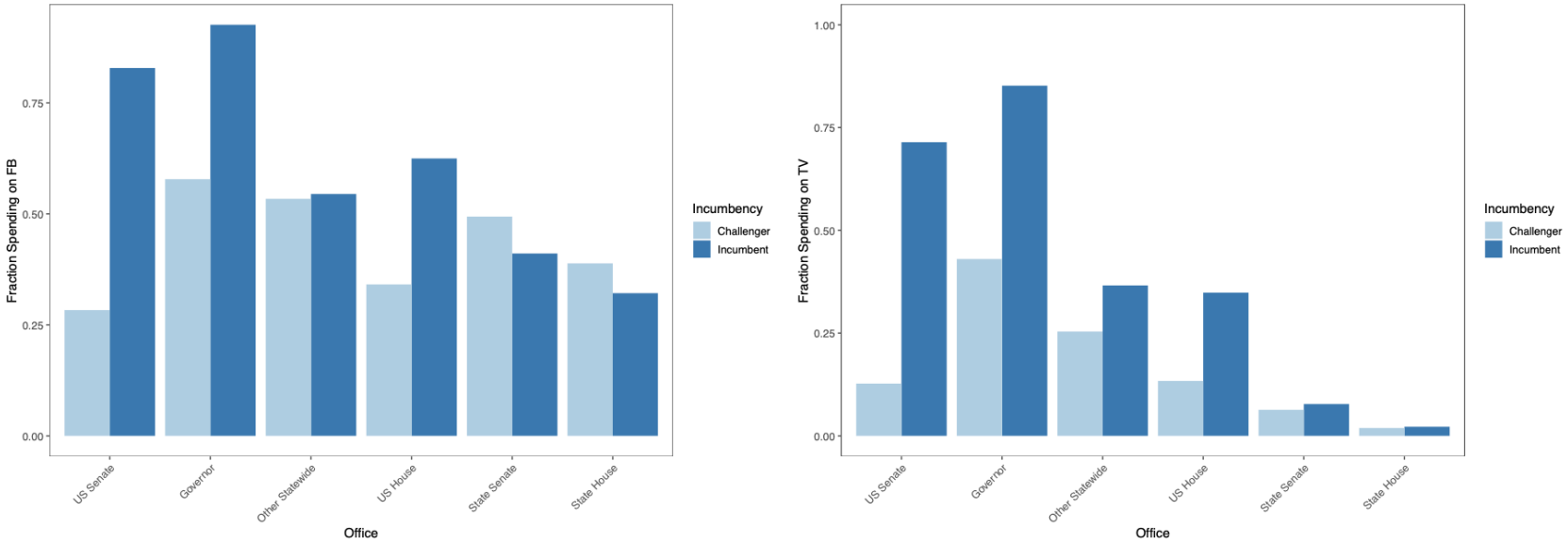
Fowler et al. (2019)

- Examine whether Facebook advertising broadens the set of candidates who politically advertise, and the tone of advertising online and offline

Political Advertising Online and Offline

ERIKA FRANKLIN FOWLER *Wesleyan University*
MICHAEL M. FRANZ *Bowdoin College*
GREGORY J. MARTIN *Stanford University*
ZACHARY PESKOWITZ *Emory University*
TRAVIS N. RIDOUT *Washington State University*

Despite the rapid growth of online political advertising, the vast majority of scholarship on political advertising relies exclusively on evidence from candidates' television advertisements. The relatively low cost of creating and deploying online advertisements and the ability to target online advertisements more precisely may broaden the set of candidates who advertise and allow candidates to craft messages to more narrow audiences than on television. Drawing on data from the newly released Facebook Ad Library API and television data from the Wesleyan Media Project, we find that a much broader set of candidates advertises on Facebook than television, particularly in down-ballot races. We then examine within-candidate variation in the strategic use and content of advertising on television relative to Facebook for all federal, statewide, and state legislative candidates in the 2018 election. Among candidates who use both advertising media, Facebook advertising occurs earlier in the campaign, is less negative, less issue focused, and more partisan than television advertising.



(a) Facebook

(b) TV

Figure 2: Fraction of candidates with positive spending on each medium, by office and incumbency status.

Does social media benefit populists & facilitate foreign interference in elections?

Guriev et al. (2021) & Bail et al. (2019)

1. Examines the effect of internet access on the growth of mistrust in government and support for populist parties (Guriev et al. 2021)
2. Investigates the effect of the the Russian Internet Research Agency on the political attitudes and behavior of US citizens (Bail et al. 2019)

3G internet and confidence in government

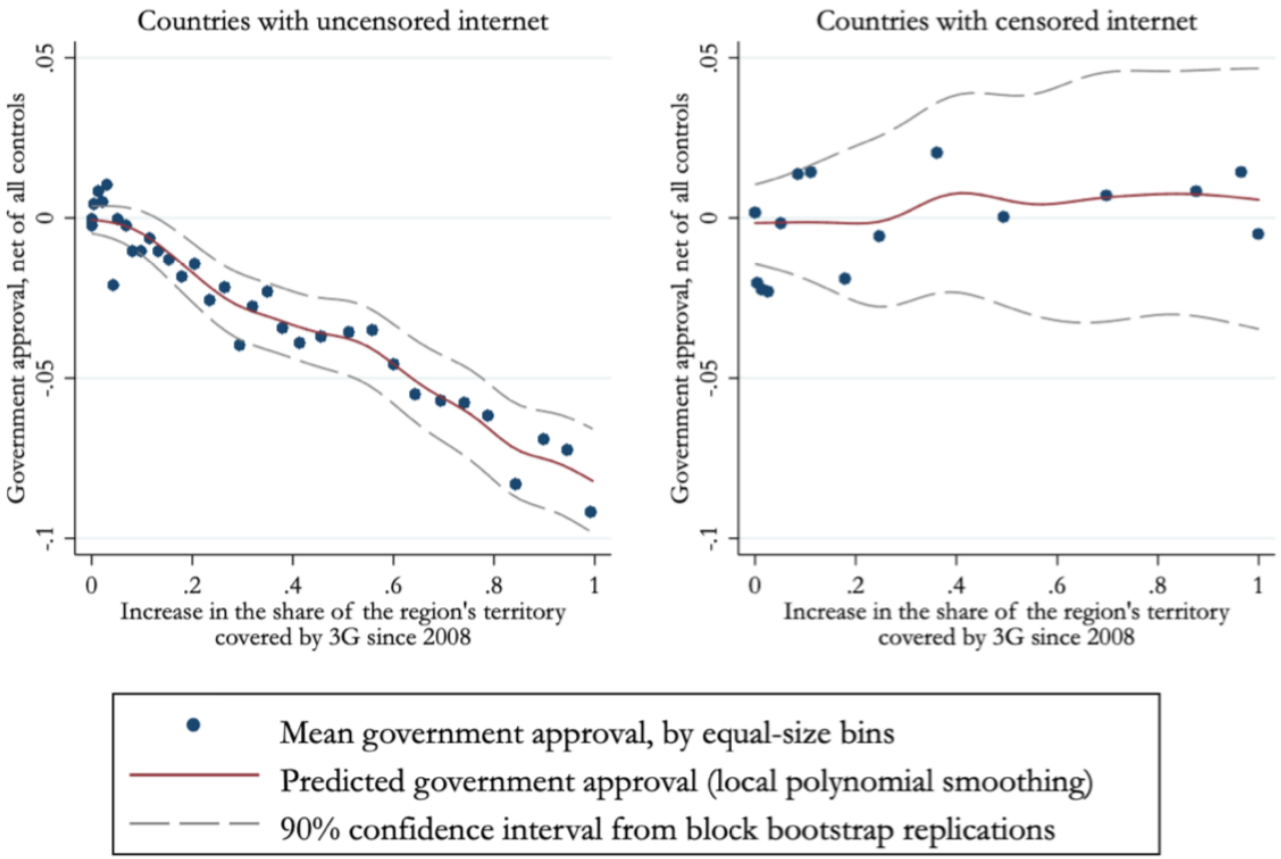
Sergei Guriev, Nikita Melnikov and Ekaterina Zhuravskaya

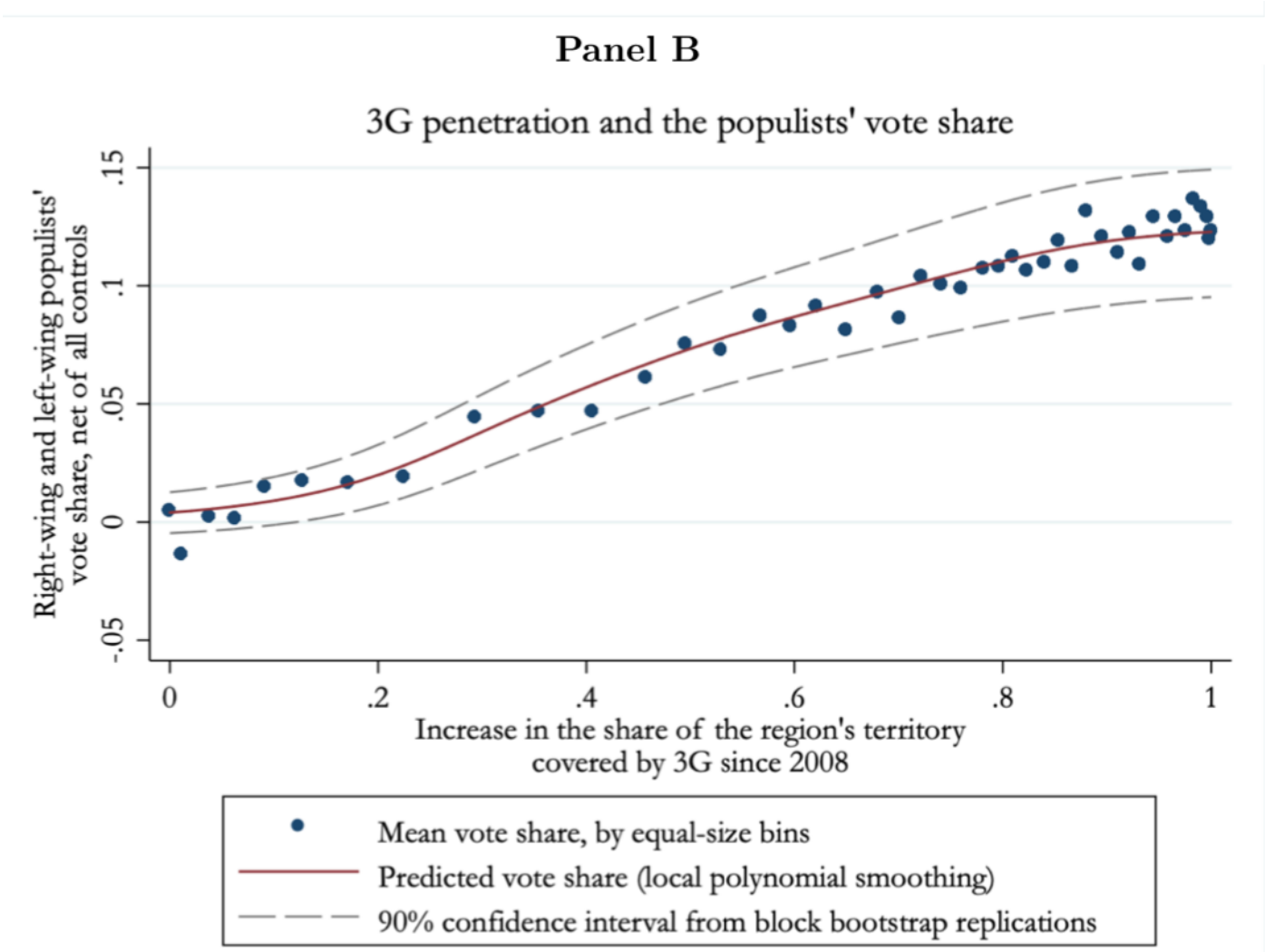
Abstract

How does the internet affect government approval? Using surveys of 840,537 individuals from 2,232 subnational regions in 116 countries in 2008-2017 from the Gallup World Poll and the global expansion of 3G networks, we show that an increase in internet access, on average, reduces government approval and increases the perception of corruption in government. This effect is present only when internet is not censored and is stronger when traditional media is. Actual incidence of corruption translates into higher corruption perception only in places covered by 3G. In Europe, the expansion of mobile internet increased vote shares of anti-establishment populist parties.

Panel A

3G penetration and government approval across the globe




Panel B
3G penetration and the populists' vote share
Right-wing and left-wing populists' vote share, net of all controls
-.05
0
.05
.1
.15
0
.2
.4
.6
.8
1
Increase in the share of the region's territory covered by 3G since 2008
Mean vote share, by equal-size bins
Predicted vote share (local polynomial smoothing)
90% confidence interval from block bootstrap replications

Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017

Christopher A. Bail[a,b,c,1], Brian Guay[a,d,2], Emily Maloney[a,b,2], Aidan Combs[a,b], D. Sunshine Hillygus[a,c,d], Friedolin Merhout[a,e], Deen Freelon[f], and Alexander Volfovsky[a,g]

[a]Polarization Lab, Duke University, Durham, NC 27708; [b]Department of Sociology, Duke University, Durham, NC 27708; [c]Sanford School of Public Policy, Duke University, Durham, NC 27708; [d]Department of Political Science, Duke University, Durham, NC 27708; [e]Department of Sociology, University of Copenhagen, 1353 Copenhagen, Denmark; [f]School of Media and Journalism, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [g]Department of Statistical Science, Duke University, Durham, NC 27708

Edited by Arild Underdal, University of Oslo, Oslo, Norway, and approved October 22, 2019 (received for review April 15, 2019)

There is widespread concern that Russia and other countries have launched social-media campaigns designed to increase political divisions in the United States. Though a growing number of studies analyze the strategy of such campaigns, it is not yet known how these efforts shaped the political attitudes and behaviors of Americans. We study this question using longitudinal data that describe the attitudes and online behaviors of 1,239 Republican and Democratic Twitter users from late 2017 merged with nonpublic data about the Russian Internet Research Agency (IRA) from Twitter. Using Bayesian regression tree models, we find no evidence that interaction with IRA accounts substantially impacted 6 distinctive measures of political attitudes and behaviors over a 1-mo period. We also find that interaction with IRA accounts were most common among respondents with strong ideological homophily within their Twitter network, high interest in politics, and high frequency of Twitter usage. Together, these findings suggest that Russian trolls might have failed to sow discord because they mostly interacted with those who were already highly polarized. We conclude by discussing several important limitations of our study—especially our inability to determine whether IRA accounts influenced the 2016 presidential election—as well as its implications for future research on social media influence campaigns, political polarization, and computational social science.

ature examines the breadth and depth of IRA activity on social media to gain insight into Russian social-influence strategies (6, 7, 9). Yet, to our knowledge, no studies have examined whether these efforts actually impacted the attitudes and behaviors of the American public (10, 11). Our study is an initial attempt to fill this research gap.

Popular wisdom indicates that Russia's social-media campaign exerted profound influence on the political attitudes and behaviors of the American public. This is perhaps because of the sheer scale and apparent sophistication of this campaign. In 2016 alone, the IRA produced more than 57,000 Twitter posts, 2,400 Facebook posts, and 2,600 Instagram posts—and the numbers increased significantly in 2017 (6). There is also anecdotal evidence that IRA accounts succeeded in inspiring American activists to attend rallies (12). The scope of this effort prompted *The New York Times* to describe the Russian campaign as "the Pearl Harbor of the social media age: a singular act of aggression that ushered in an era of extended conflict" (13).

Significance

While numerous studies analyze the strategy of online influence campaigns, their impact on the public remains an open question. We investigate this question combining longitudi-

Next

- Assignment to presentation groups
- 20 minutes to discuss what article you want to present