Political Analysis of Social Media Data

Experiments

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

Today

- Experiments
- *No* video lectures or exercises today

3-day exam in three sections

1. Show that you understand the use of methods in existing research
    - You will be asked to explain a passage from an existing academic article about application of a method from the course
2. Show that you can explain in plain language a method from the course
3. Show that you can conduct analysis of social media data in R
    - You will be asked to load data; fit a model or technique to those data; graph it; and explain the results

Experiments in social media research

- Why are experiments useful?
- Types and examples of social media experiments
- Design your own experiment

Observational & experimental research

- Estimating causal effects is often very difficult
- Examples:
  - If I show you data that those who use social media are more polarized, is that strong evidence that social media causes polarization?
  - If I show you that people who are exposed to Russian bots on social media were less likely to support Hilary Clinton for president, does that mean that the Russian disinformation campaign caused a decrease in her support?
  - If I show you that people on social media who see messages about a protest were more likely to turn out for that protest, would you believe that the messages increased turnout?

An instructive comparison with “wellness” programs...

A yoga class at a New York wellness center. Randomized controlled trials can reverse the conclusions of observational studies. Chad Rhym for The New York Times

The empirical expectations of wellness programs:

- Increase well-being
- Increase productivity
- Decrease absenteeism
- Decrease medical spending

Observational designs suggest the hypotheses are correct. They:

- Increase well-being!
- Increase productivity!
- Decrease absenteeism!
- Decrease medical spending!

The problem? Those who partake in wellness programs are:

- Likely to be healthier to begin with
- Also less likely to require less medical spending

Can't we control for confounders

- Yes? And observational studies include controls
- So shouldn't estimates of the effect of wellness programs at least be close to the truth?

Large-scale randomized controlled trial

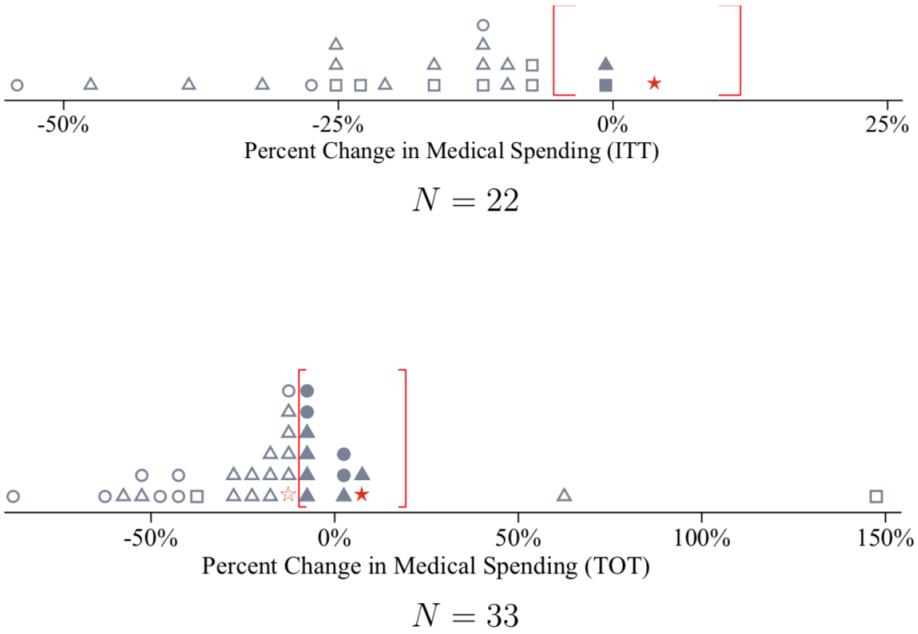
- Authors randomize subjects to a treatment or control group
- The benefit is that randomization breaks selection bias
  - Those assigned at random to treatment and control will be effectively equivalent on observed *and* unobserved characteristics
  - E.g. no differences (in expectation) on health, ideology, polarization, income, gender, age, political interest, etc.
  - Thus after running the experiment, any differences in the outcomes will not be due to selection bias

Results of the RCT on wellness programs?

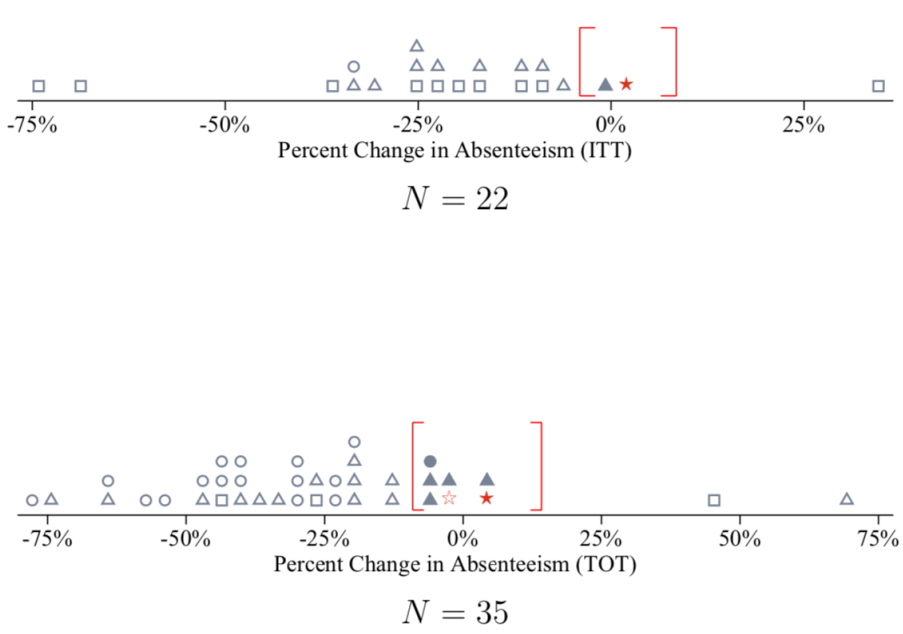
- Effect on well-being: **null effect**
- Effect on productivity: **null effect**
- Effect on medical spending: **null effect**
- Effect on absenteeism: **null effect**

The authors then analyzed their data as if it were an observational study (i.e. a regression with controls):

- Effect on well-being: **positive**
- Effect on productivity: **positive**
- Effect on medical spending: **negative**
- Effect on absenteeism: **negative**



(a) Medical Spending Estimates



(b) Absenteeism Estimates

IL Estimates: ★ RCT [ ] 95% Confidence Interval ☆ OLS
Prior Studies:
Not Ruled Out: ■ RCT ▲ Pre/Post ● Other
Ruled Out: □ RCT △ Pre/Post ○ Other

Figure 6: Comparison of Experimental Estimates to Prior Studies

"If we had published only these observational analyses, the headline result could have been that even after controlling for a battery of confounding variables, participation in a wellness program was associated with a significant reduction in health care spending, an improvement in exercise, and a lower chance of ceasing employment."

Type of experiments

- Field experiments
  - High internal *and* external validity
- Survey experiments
  - High internal validity
- Lab experiments
  - High internal validity
- Natural or quasi-experiments
  - Massive literature on these
  - Used when cannot feasibly run an experiment
    - Difference-in-difference / event studies
    - Regression discontinuity designs
    - Instrumental variables

Social media experiments

- Increasing in number
- Many are lab- and survey-experiments
- Some of the best are field experiments
- Some you could conduct yourself, with relatively few resources
- Others are costly or require connections to, say, Facebook
- We'll walk through a few of the big ones

Answering some big questions

- What are the effects of social media on behavior, attitudes, and well-being?
  - Randomly assign social media
- What are the effects of censorship on attitudes and information consumption?
  - Randomly assign censorship

American Economic Review 2020, 110(3): 629–676
https://doi.org/10.1257/aer.20190658

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)

What are social media's effects on social well-being, political information, and polarization?

- Interpersonal connections are important drivers of happiness and well-being
- But also potential negative effects:
  - Echo chambers
  - Polarization
  - Fake news
  - Depression
- Certain type of person chooses to use social media, so many possible confounders

Experiment

- Offer users money to deactivate Facebook for a month
- Recruitment from Facebook ads
- Users surveyed before and after experiment
- Use "willingness-to-accept" mechanism to measure Facebook's value to each user:
  - "The computer has randomly generated an amount of money to offer you to deactivate your Facebook account for the next 4 weeks. Before we tell you what the offer is, we will ask you the smallest offer you would be willing to accept. If the offer the computer generated is above the amount you give, we will ask you to deactivate for 4 weeks and pay you the offered amount if you do. If the offer is below that amount, we will not ask you to deactivate."

Outcome variables:

1. Social interaction
2. News knowledge
3. Political engagement
4. Political polarization
5. Subjective well-being
6. Post-experiment Facebook use
7. Opinions about value of Facebook
8. Substitute time uses (what people do instead)
9. Substitute news sources (what people read instead)

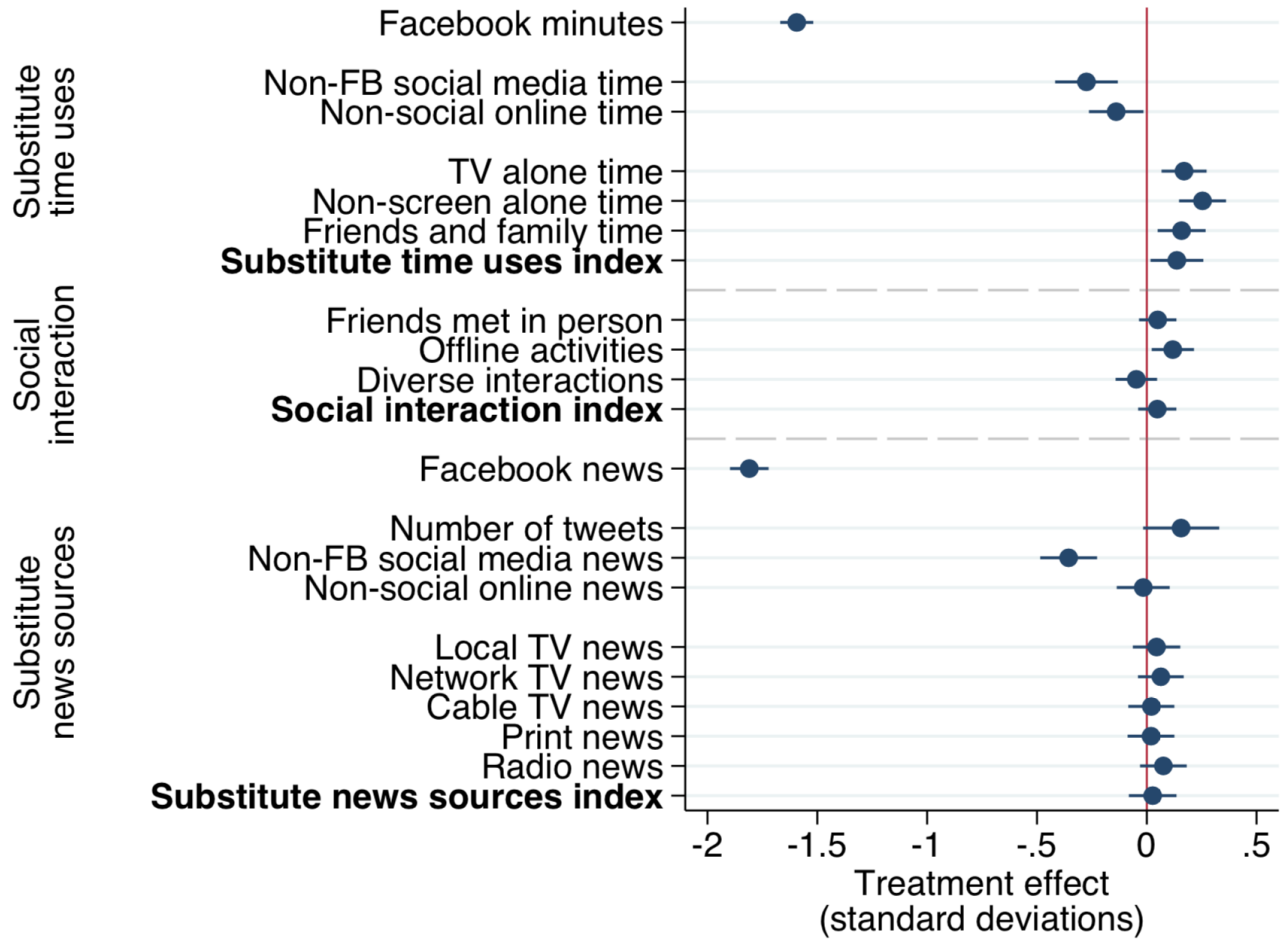
Sample unrepresentativeness:

Table 2: **Sample Demographics**

	(1) Impact evaluation sample	(2) Facebook users	(3) US population
Income under $50,000	0.40	0.41	0.42
College	0.51	0.33	0.29
Male	0.43	0.44	0.49
White	0.68	0.73	0.74
Age under 30	0.52	0.26	0.21
Republican	0.13		0.26
Democrat	0.42		0.20
Facebook minutes	74.52	45.00	

Results:

Figure 2: **Substitutes for Facebook**

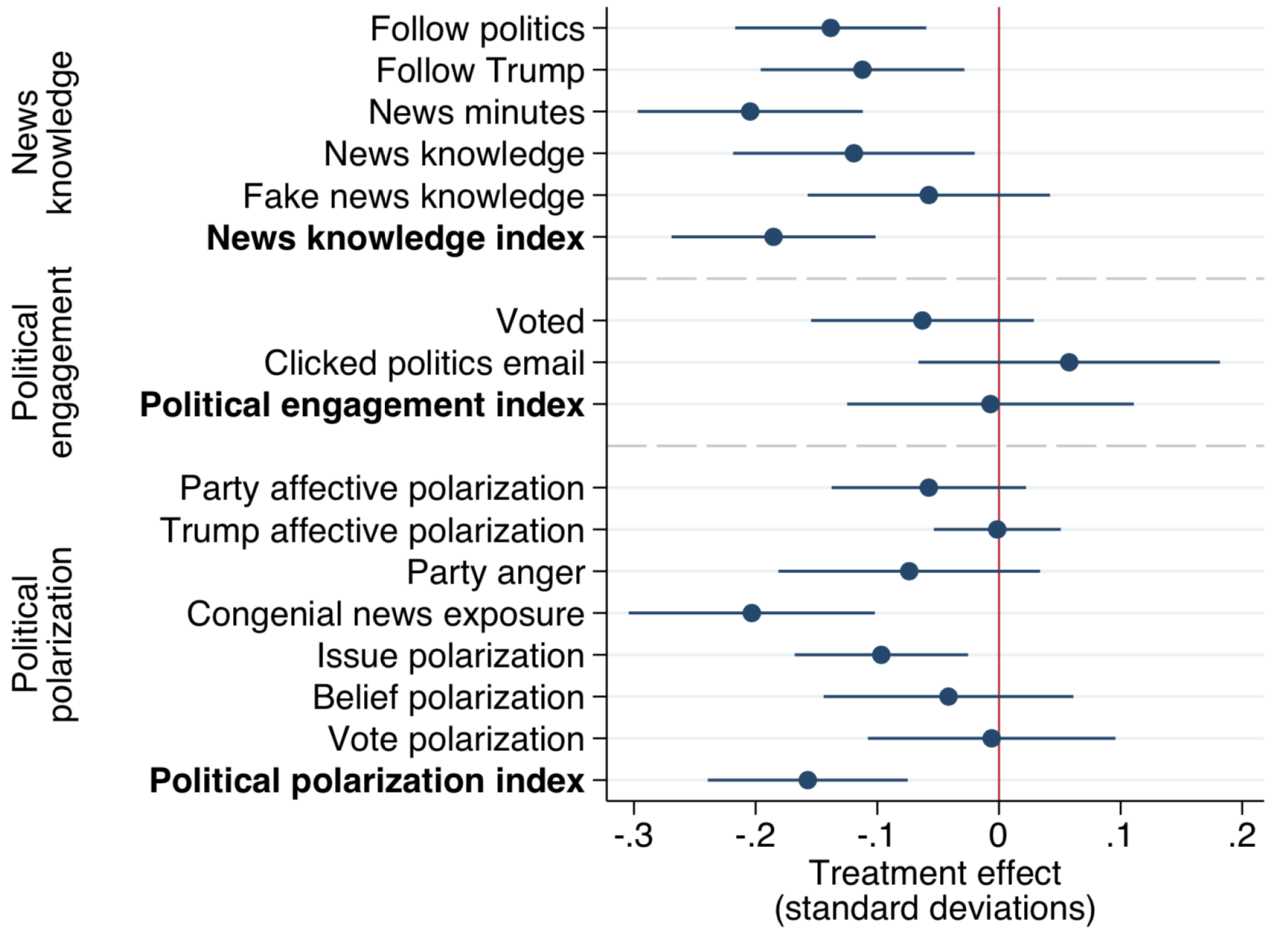


Highlights of these results:

- Decrease in use of *other* social media
- Read less news on other social media platforms
- Increase in solitary TV watching, other solitary activities, and time with friends and family
- No increases in going to the cinema, talking to friends, going to a party, going shopping, or time with one's children

Results:

Figure 3: **Effects on News and Political Outcomes**

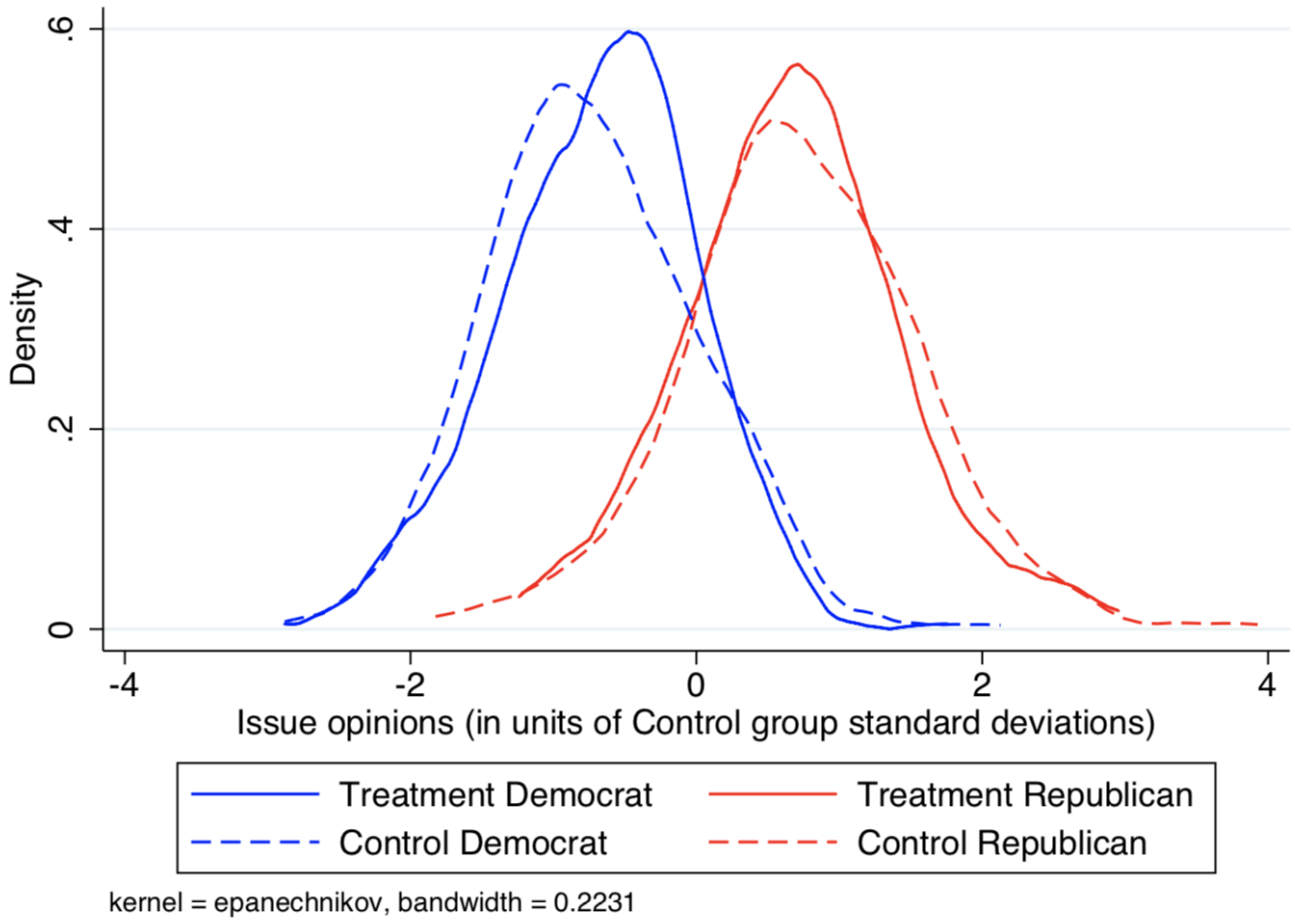


Highlights of these results:

- Lower attention to news and to political knowledge
- No effect on political engagement
- Lower political polarization...

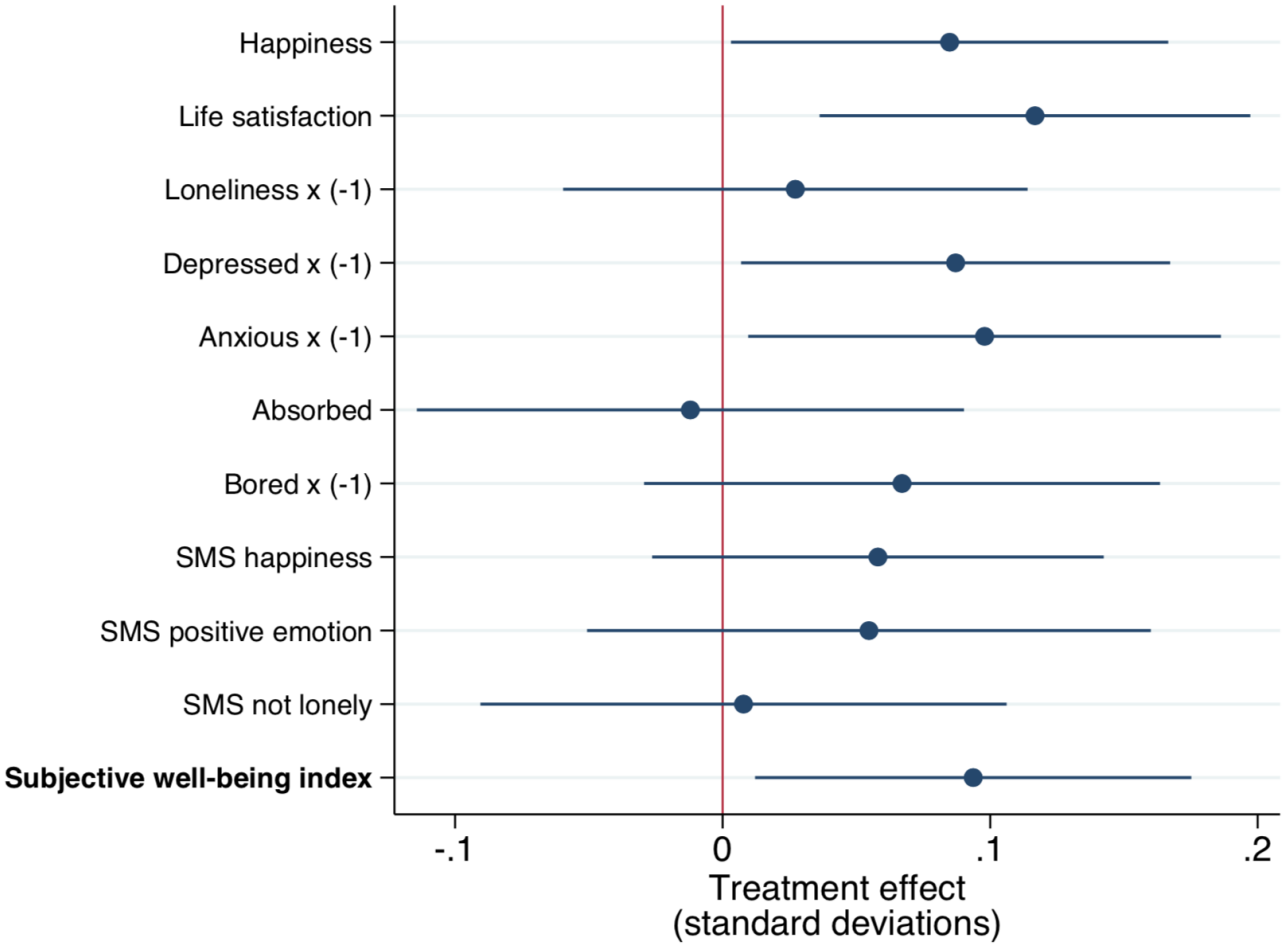
Lower political polarization:

Figure 4: **Issue Opinions by Party at Endline**



Results:

Figure 5: **Effects on Subjective Well-Being**



Highlights of these results:

- The increase in well-being is as high as 25-40% of a standard psychological intervention
- If they analyze the study as a standard observational study, the effects would be much higher (0.23 SD instead of the true 0.09 SD)
- These observational "results" are consistent with reverse causality, and highlights the benefit of an experiment, e.g. if people who are lonely or depressed spend more time on Facebook

Qualitative assessments by research subjects:

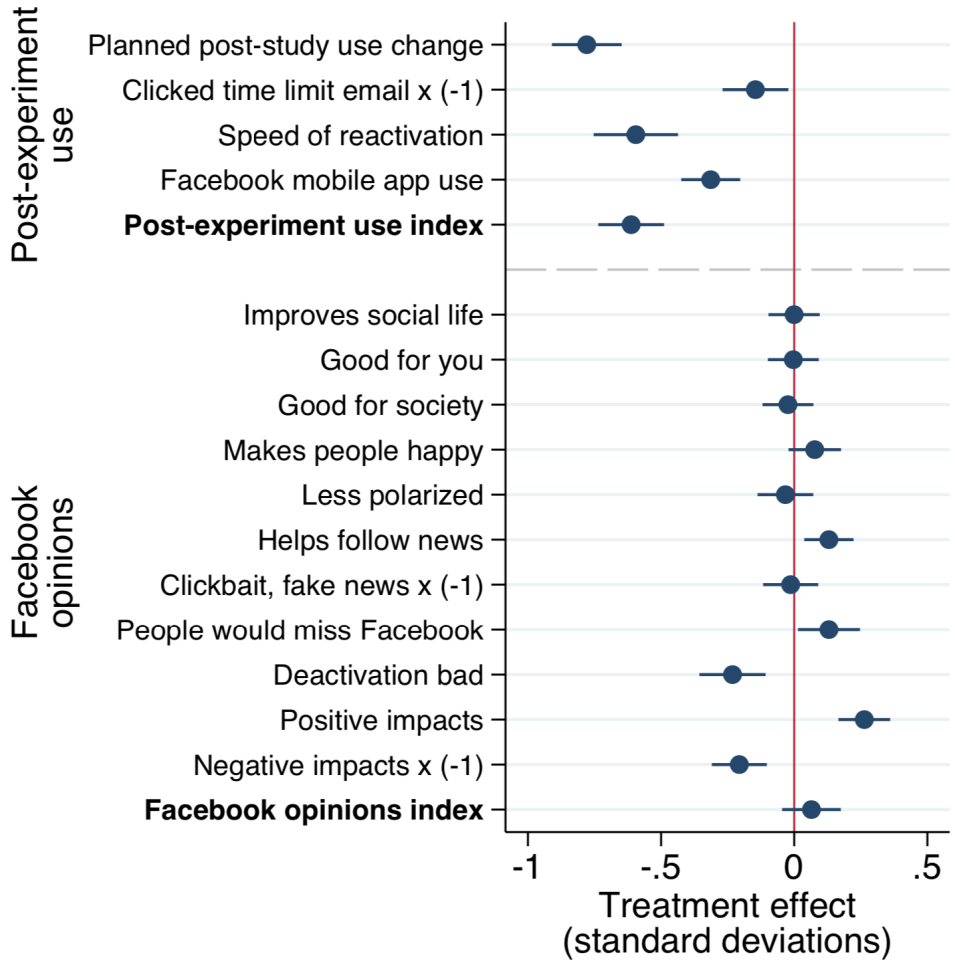
"I was way less stressed. I wasn't attached to my phone as much as I was before. And I found I didn't really care so much about things that were happening [online] because I was more focused on my own life... I felt more content. I think I was in a better mood generally. I thought I would miss seeing everyone's day-to-day activities... I really didn't miss it at all."

But also:

"I was shut off from those [online] conversations, or just from being an observer of what people are doing or thinking... I didn't like it at first at all, I felt very cut off from people that I like... I didn't like it because I spend a lot of time by myself anyway, I'm kind of an introvert, so I use Facebook in a social aspect in a very big way."

Results:

Figure 6: **Effects on Post-Experiment Facebook Use and Opinions**

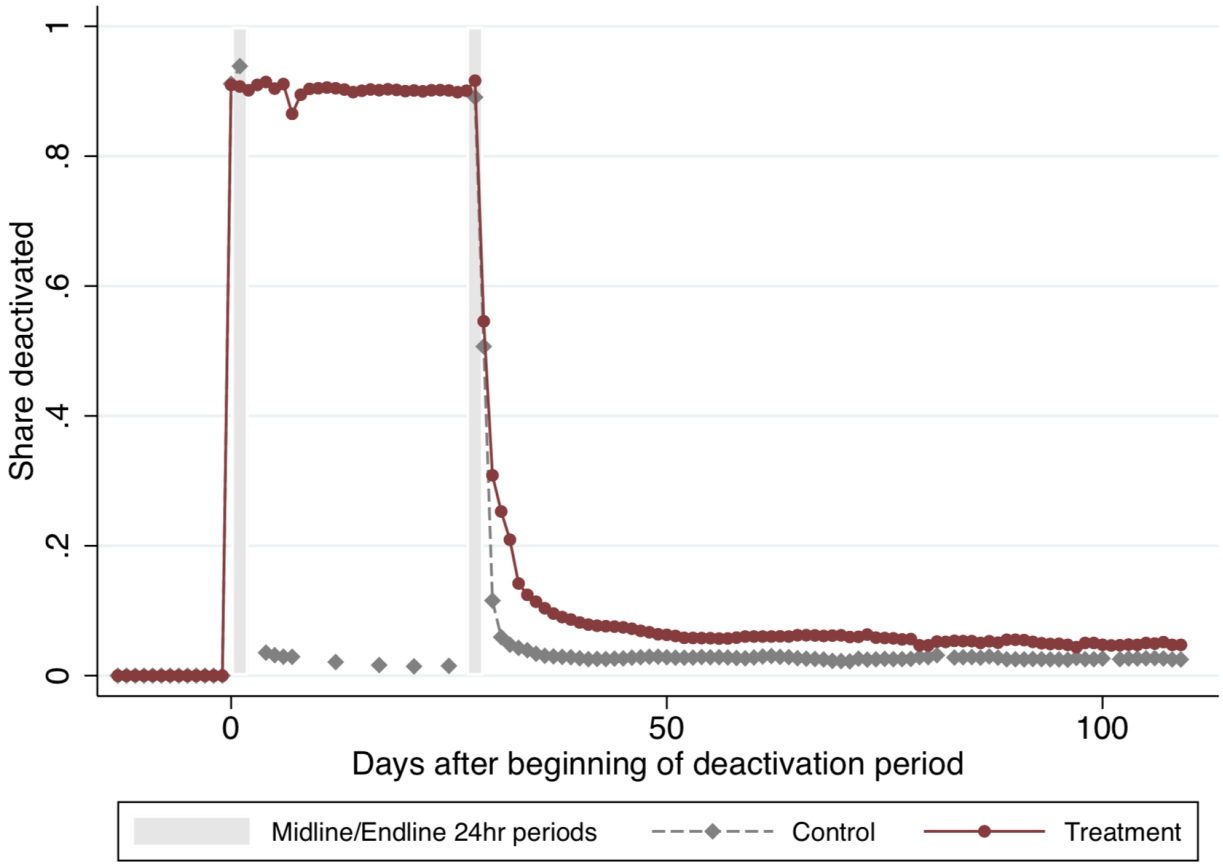


Highlights of these results:

- Massive reduction in demand for using Facebook
  - Consistent with a habit-forming model
- But de-activation also increased extent that people believe Facebook helps them follow the news better, and agree that people would miss using Facebook if they stopped

Results:

Figure 7: **Probability of Being Deactivated**



Do people misunderstand the value of social media?

- Might misunderstand its addictive quality or that it is making them unhappy
- Might misunderstand that social media is habit forming
- "Digital detox" may help consumers realize its value relative to other uses of time

Test by assessing the value of Facebook to the control and treatment group:

- Probe willingness to pay to stay off of Facebook at the end of the experiment
- People who were off of Facebook for a month asked for less money to stay off Facebook for another month
- Digital detox reduced the value of Facebook to users by 14%

Conclusions

- Facebook is an important sources of news and information
- Is a source of entertainment, facilitates charity and activist organizations, and provides a social lifeline for those who are isolated
- Discussion of downsides obscures basic fact that it fulfills deep and widespread needs
- Downsides are also real
- Four weeks without Facebook increased subjective well-being and post-experiment demand for its use
- People are less informed, but also less polarized

American Economic Review 2019, 109(6): 2294–2332
https://doi.org/10.1257/aer.20171765

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)

Two major empirical questions

1. Does access to an uncensored Internet cause people to acquire politically sensitive information?
2. Does politically sensitive information change citizens' beliefs, attitudes, and behaviors?

Why might the public not search out censored information?

1. Lack of interest in politics?
2. Fear of government reprisal?
3. Unaware or distrust of foreign news?

Empirical case: China

1. Great Firewall
2. Tools to bypass censorship: Virtual Private Networks (VPNs)
3. Yet relatively low use of VPNs. Why?

Experiment

- 1,800 Chinese students
- Treatment group given 18 months of a free VPN
- Some in treatment group also encourageed to access foreign news with monetary rewards and newsletters

Behavioral outcomes

1. Activation of VPN
2. Use of VPN
3. Time spent on foreign sites
4. Use of VPN after experiment is over

Belief-based outcomes

1. Valuation of uncensored internet
2. Trust in domestic & foreign news
3. Beliefs about censorship levels

Knowledge-based outcomes

1. Sensitive historial knowledge
2. China's GDP & stock market
3. Political attitudes
4. Past and future economic & political behaviors

Does VPN access increase acquisition of sensitive information?

- 55% of treated group actually activate the tool.
- 27% of those who activate it do not use it.

- Less than 5% in treatment group without the encouragement browse foreign news websites.
- Provision of access alone is thus not sufficient

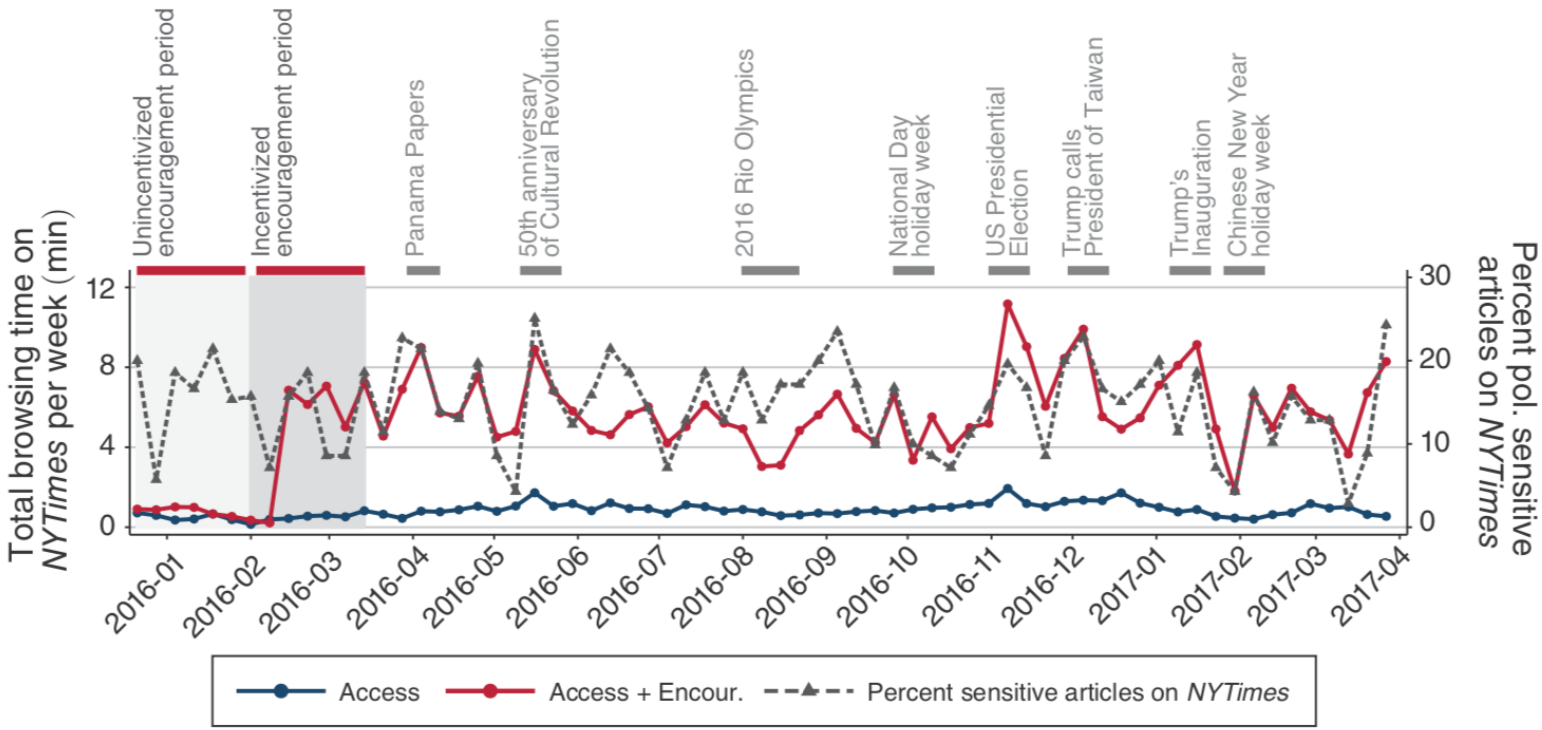


FIGURE 1

Does access increase acquisition of sensitive information?

- Those in encouragement group much more likely to use the VPN (14 percentage points)
- But don't visit foreign sites without first having monetary encouragement incentive
- Nevertheless, after the encouragement ends, students continue to visit foreign news sites

Do people continue to use a VPN?

- After experiment over, those in the treatment group are much more likely to subscribe to the VPN
- Are also willing to pay 70% more for VPN access after the experiment compared to the control group

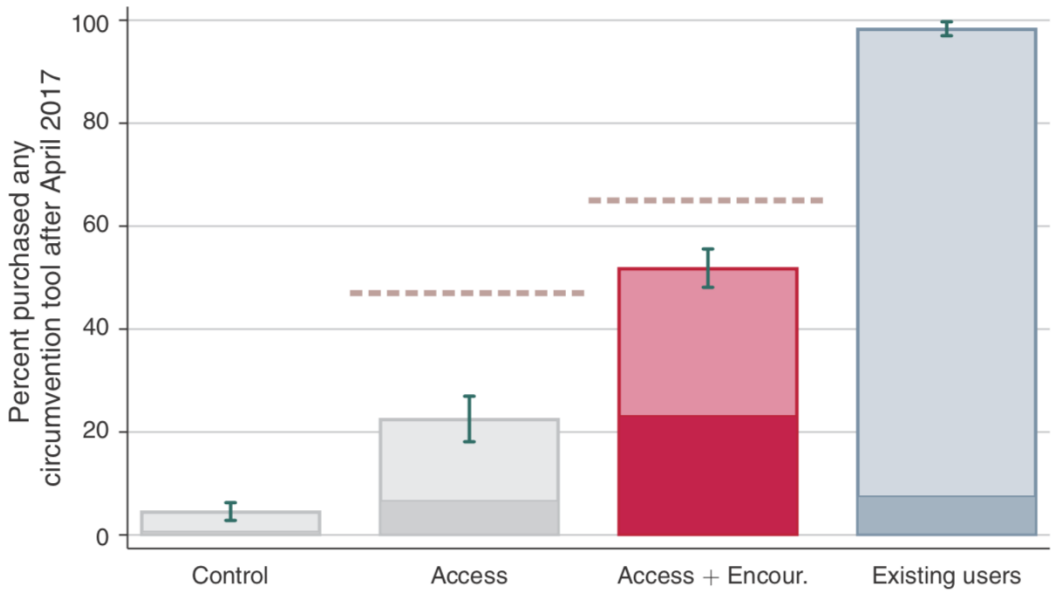


FIGURE 3

Is the increase in demand for a VPN due to the availability of politically sensitive information? No.

- *Not* because of foreign news access
- Non-encouragement group also likely to pay as much, and they did not visit foreign sites
- Thus because of other reasons (e.g. entertainment)

What explains lack of censorship avoidance?

- A lack of *demand* for foreign news information
- However, once aware of it, use it much more
- And it does affect beliefs:
  - See more value in censorship avoidance than control group
  - See domestic media as more censored
  - Less trust in domestic media

What is the effect of acquiring uncensored political information?

- More informed about current & historical political events
- More pessimistic about Chinese economy
- Lower trust in government
- Discuss politics more with friends
- More likely to want to study abroad
- But *not* more likely to engage in political action

Does information spillover to others?

- Relatively small knowledge spillover effects (roommates of those in the treatment)
  - Because those without knowledge wouldn't know what to ask?
  - Because those with new knowledge assume others already know?
  - Clustering among those with access?

A pessimistic conclusion?

- *Lack of demand* may be a large driver of lack of censorship avoidance
- "In fact, the Chinese government may not need to bear the extremely high costs of fully 'sealing' its Internet, as it can afford to leave some holes open."

Other large-scale field experiments

- Bond et al. (2012)
  - Show Facebook users a social message about friends who voted
- King, Pan, and Roberts (2014)
  - Write posts on 100 Chinese social media sites to test what is censored
- Kramer et al. (2014)
  - Remove negative or positive sentiment posts from users' Facebook feeds
- Bail et al. (2018)
  - Pay people to follow a bot that retweets posts from other-partisans

A growing list of experiments using interactions with users on social media

- Coppock et al. (2016), Eckles et al. (2016), Munger (2017a), Bohren et al. (2019), Gallego et al. (2019), Siegel & Badaan (2020), Yang et al. (2020), Pennycook et al. (2021), Mosleh et al. (2021a), Mosleh et al. (2021b), Foos et al. (2021), Munger (Forthcoming)...
- Siegel & Badaan (2020) as an example...

American Political Science Review (2020) 114, 3, 837–855
doi:10.1017/S0003055420000283

#No2Sectarianism: Experimental Approaches to Reducing Sectarian Hate Speech Online

ALEXANDRA A. SIEGEL *University of Colorado at Boulder*
VIVIENNE BADAAN *New York University*

We use an experiment across the Arab Twittersphere and a nationally representative survey experiment in Lebanon to evaluate what types of counter-speech interventions are most effective in reducing sectarian hate speech online. We explore whether and to what extent messages priming common national identity or common religious identity, with and without elite endorsements, decrease the use of hostile anti-outgroup language. We find that elite-endorsed messages that prime common religious identity are the most consistently effective in reducing the spread of sectarian hate speech. Our results provide suggestive evidence that religious elites may play an important role as social referents—alerting individuals to social norms of acceptable behavior. By randomly assigning counter-speech treatments to actual producers of online hate speech and experimentally evaluating the effectiveness of these messages on a representative sample of citizens that might be incidentally exposed to such language, this work offers insights for researchers and policymakers on avenues for combating harmful rhetoric on and offline.

What prevents religious sectarian hate speech on social media?

- Sectarian hate speech is harmful to inter-group relations & to politics more broadly
- What strategies work best to prevent it?
  - Appeals to broader religious identity?
  - Appeals to broader national identity?
  - Counter-speech by elites (elite endorsements)?
  - Counter-speech by religious leaders (religious leader endorsements)?

Experimental setup

- Create a Twitter bot to respond to Arab Twitter users who regularly tweet hostile sectarian language
- Use streaming API to collect tweets from those who regularly tweet sectarian slurs (at least 5) over a six month period
- Subset:
  - Exclude users whose profiles are less than two months old
  - Exclude users with $>$ 10,000 followers
  - Exclude those who appear to be very young or bots

Experimental setup

- Create a Twitter bot called “Mohammed Ahmed”
- To make him seem like a normal user, regularly tweet news about soccer and Quranic verses

“Mohammed Ahmed”

Figure 1: Gulf Twitter Sock Puppet for all Treatments

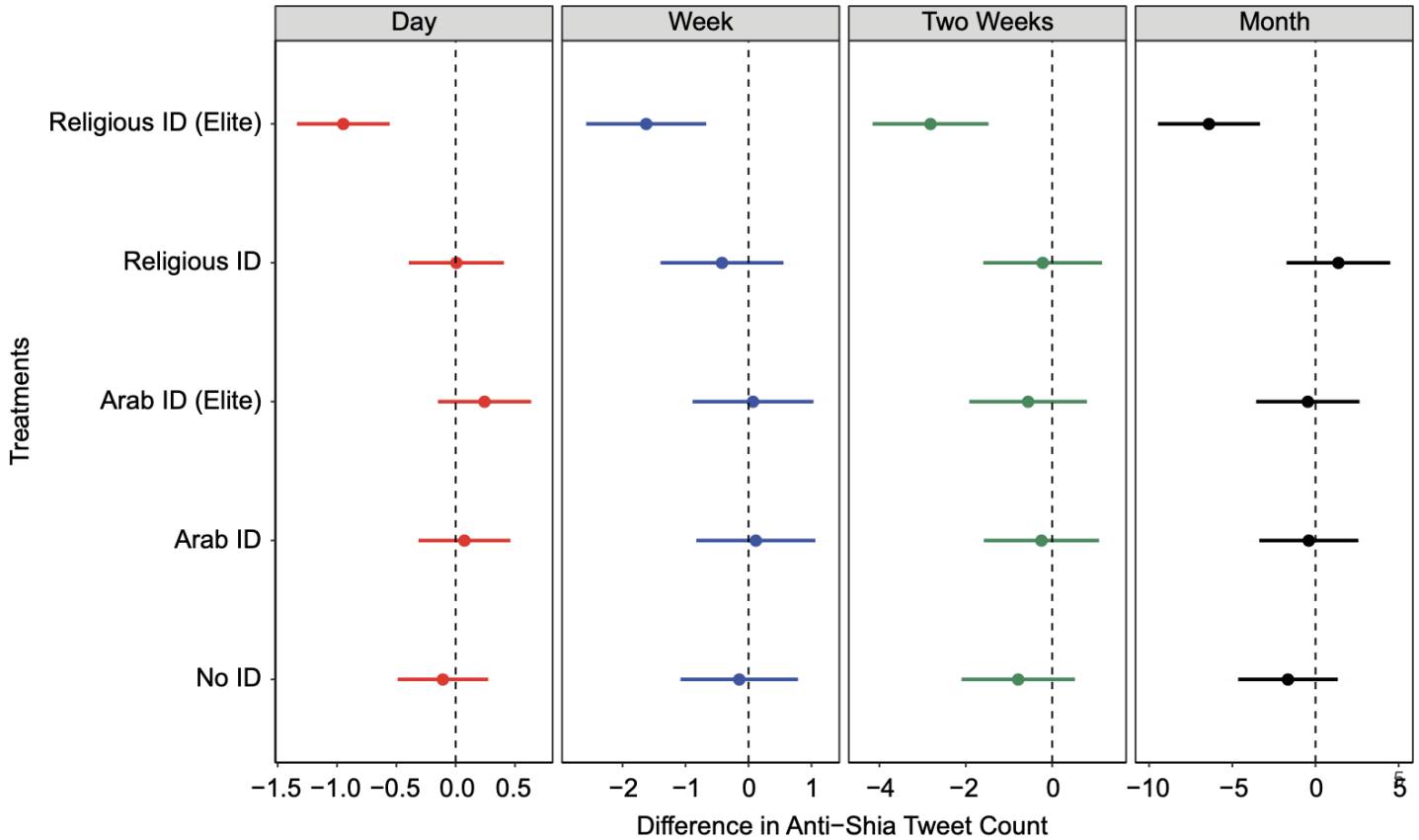
6 experimental treatments:

1. Control (no message)
2. No prime ("That language sows (sectarian) discord/strife.")
3. Common national identity prime ("That language sows (sectarian) discord/strife. We are all Arab.")
4. Common religious identity ("[...] We are all Muslim")
5. Elite common national identity
6. Elite common religious identity

Post-treatment data collection:

1. Use API to collect all subsequent tweets from users in the experiment
2. Measure the count of the sectarian hate speech tweets
3. Test whether some treatment messages worked better than others

FIGURE 2. Effect of Treatment on Volume of Anti-Shia Tweets



Note: This coefficient plot shows the results of four ordinary least squares models, where the outcome variable is the difference in the number of anti-Shia tweets produced by subjects (Twitter users) in our experiment one day, one week, two weeks, and one month after treatment. Treatment periods are nonoverlapping. The error bars show 90% and 95% confidence intervals. The full output is displayed in Table A3 in the Online Appendix.

Things to check in these experiments

- Is there a larger effect among users with fewer followers?
  - Users with more followers less likely to see the treatment
- Differences by users' network: if follow more users who post less anti-Shia content
  - Users with fewer such friends may be more easily sanctioned

Would these interventions generalize?

- Sample population are those who posted anti-Shia tweets
- But would this intervention work on the general population?
- To test, authors run a survey experiment among representative sample of 500 people in Lebanon...

Tweet-based survey experiment

- Treatment group respondents read one of four counter-speech primes
- Then were presented with tweet containing anti-Shia content
- Outcomes:
  - How favorable feeling toward the author of the tweet
  - How favorable feeling toward the content of the tweet
  - Willingness to share the tweet

Example prime and tweet

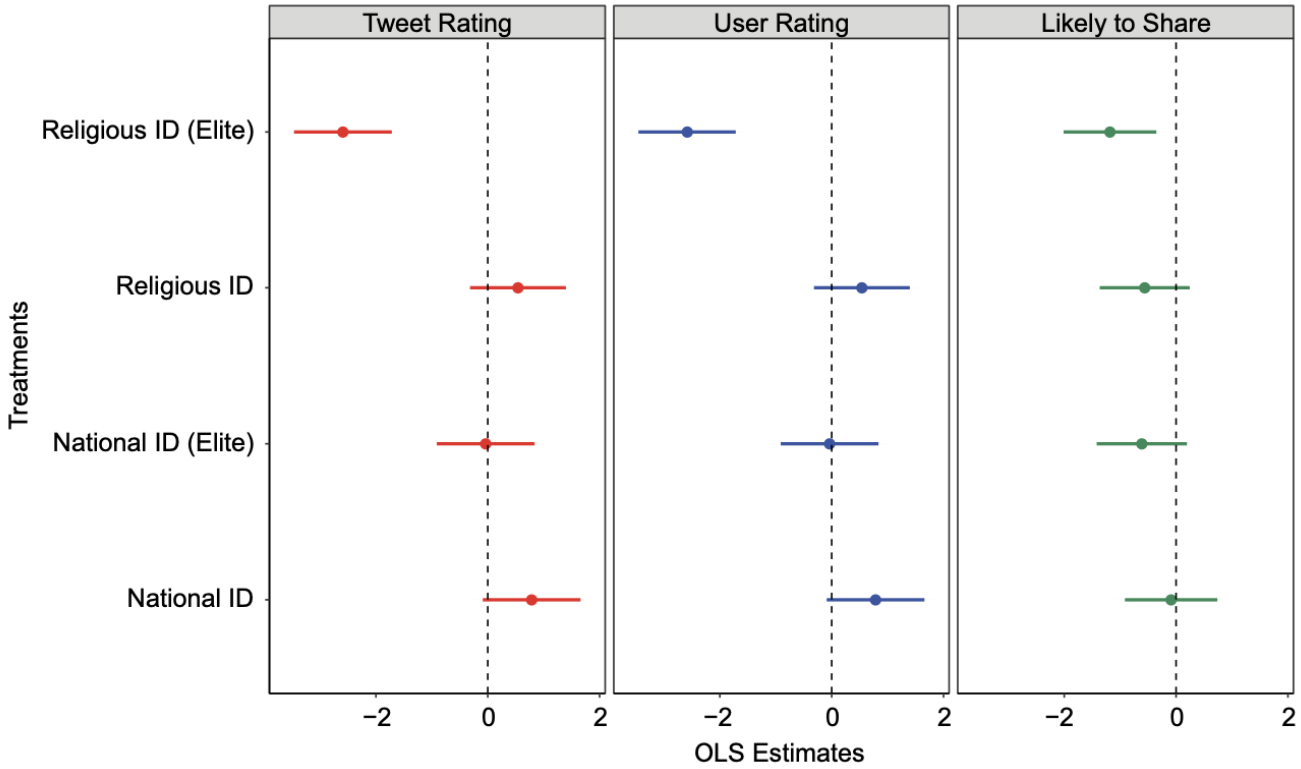
- **Elite Common-Religious-Identity Prime:** Over the past few years, there has been a rise in sectarian tensions in Lebanon and across the Middle East. *But many prominent Christian, Sunni, and Shia religious leaders have issued religious decrees calling for people to come together and stop inciting sectarian hatreds. They agree that we all believe in one God and we all should be equal. We all live on one land. We share the same history and the same future; we share the same culture, the same food, and language. Most importantly, we share a common belief in God.*Such sectarian issues have been widely discussed on Twitter, a popular social networking site on which users can post messages of 140 characters or less and share messages with their friends. You will now be presented with several messages from Twitter users on this topic and will answer a few questions about each message.

Sectarian Tweet Example (Translated Text):

- #HezbollahDestroysLebanon. Most people are fully aware that *Hezb al-Lat* [derogatory sectarian term for Hezbollah] and its *rawafidh* [anti-Shia slur] followers are a brutal subversive arm of Iran.

Results

FIGURE 6. Effect of Primes on all Tweet Ratings



Note: This coefficient plot shows the results of three ordinary least squares models, where each outcome variable is an index created by subtracting subjects' average ratings of counter-sectarian tweets from their ratings of sectarian tweets. The first column displays the effect of treatment on ratings of the tweets themselves, the second displays the effect on ratings of the users who produced the tweets, and the third column displays respondents' likelihood of sharing the tweets. Negative or lower values of this index signify lower ratings of sectarian content and/or higher ratings of counter-sectarian content. The error bars show 90% and 95% confidence intervals. This plot displays results without covariates. The full regression output from the models with and without covariates is displayed in Table A22 in the Online Appendix.

In sum

- Field experiments can demonstrate the effect of a treatment in the real world
- On a platform like Twitter, they are relatively easy to implement
- Can supplement field experiments with survey experiments to test plausibility of generalizing to other populations, or to test for mechanisms (e.g. why does the religious elite treatment work?)

Other examples

- Pennycook et al. (2021)
  - Direct message to users who share fake news to ask them to rate the accuracy of a single non-political headline → Less subsequent fake news sharing
- Mosleh et al. (2021a)
  - Two partisan bots follow Democratic and Republican users → Users more likely to follow co-partisans
- Coppock et al. (2016)
  - Advocacy group sends direct messages or tweets public messages to group members → Direct messages increase petition signing
- Mosleh et al. (2021b)
  - Being corrected for sharing fake news → *increases* partisanship slant and incivility

Types of social media experiments

- "Offline" interventions to examine online & online behavior
  - e.g. Allcott et al. (2020), Chen and Yang (2019)
- Online interventions to examine offline & online attitudes
  - e.g. Bail et al. (2018)
- Interventions delivered through interactions with users
  - e.g. Siegel and Badaar (2020), Coppock et al. (2016)
- Sending political advertisements (e.g. on Facebook)
  - e.g. Ryan and Brader (2017)

Develop your own hypothesis & experiment

- Partisanship, issue position, Race, gender, religion
- Effectiveness of messages (campaigns, corrections, interventions)
- Effectiveness of ads on Facebook

Who is the sample and how would you select?

- Followers of a user or campaign?
- Users who post certain keywords or @mention certain users?
- Users who follow back a bot you create?

What is the outcome and how is it measured?

- Number or content of social media posts?
- Whether a bot is followed back?
- Whether a user responds?
- Whether a user clicks on an ad?

Are there any ethical problems involved?