

Class 20: Fixing social media

Matthew J. Salganik

Sociology 204: Social Networks
Princeton University



Feedback on the feedback

- ▶ thank you for your feedback on the quizzes

Feedback on the feedback

- ▶ thank you for your feedback on the quizzes
- ▶ thank you for your suggestions for the final class topic

Social media:

- ▶ Lecture 17: Social media and individuals
- ▶ Lecture 18: Social media and society
- ▶ [Lecture 19: Social ads in social media](#)
- ▶ Lecture 20: Fixing social media
- ▶ Lecture 21: Facebook Files

Social media:

- ▶ Lecture 17: Social media and individuals
- ▶ Lecture 18: Social media and society
- ▶ Lecture 19: Social ads in social media
- ▶ [Lecture 20: Fixing social media](#)
- ▶ Lecture 21: Facebook Files

1. Frank, R.H. (2021). The economic case for regulating social media. *New York Times*.
2. Pennycook, G. and Rand, D. (2020). The right way to fight fake news. *New York Times*.
3. Bail, C.A. et al. (2018). Exposure to opposing views on social media can increase political polarization. *PNAS*.
4. Kaiser, B., Mayer, J. and Mattias, J.M. (2021). Warnings that work: Combating misinformation without deplatforming. *Lawfare*.
5. Kaiser, B. et al. (2021). Adapting security warnings to counter online disinformation. *30th USENIX Security Symposium (USENIX Security 21)*.

Community Minute

Fixing social media

“Fixing” social media

Changing social media



Tristan Harris, Congressional testimony, January 8, 2020: “Americans at Risk: Manipulation and Deception in the Digital Age”

<https://www.youtube.com/watch?v=LUNErhONqCY>

Two broad categories: Structural changes vs tweaks

ECONOMIC VIEW

The Economic Case for Regulating Social Media

The core business model of platforms like Facebook and Twitter poses a threat to society and requires retooling, an economist says.

Force social media companies to switch to a subscription model.

ECONOMIC VIEW

The Economic Case for Regulating Social Media

The core business model of platforms like Facebook and Twitter poses a threat to society and requires retooling, an economist says.

Force social media companies to switch to a subscription model. What kinds of changes would happen? How can we know? Who would be harmed?

Possible structural changes

- ▶ require subscriptions for social media

Possible structural changes

- ▶ require subscriptions for social media
- ▶ creating online privacy regulations (end surveillance capitalism)

Possible structural changes

- ▶ require subscriptions for social media
- ▶ creating online privacy regulations (end surveillance capitalism)
- ▶ breaking up monopolies

Possible structural changes

- ▶ require subscriptions for social media
- ▶ creating online privacy regulations (end surveillance capitalism)
- ▶ breaking up monopolies
- ▶ changing Section 230 of the Communications Decency Act ("No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider")

All the Ways Congress Wants to Change Section 230

Republicans and Democrats alike want to change Section 230 of the Communications Decency Act. Here's a comprehensive list of the proposed legislation so far.

BY KIRAN JEEVANJEE, BRIAN LIM, IRENE LY, MATT PERAULT, JENNA RUDDOCK, TIM SCHMELING, NIHARIKA VATTIKONDA, AND JOYCE ZHOU

MARCH 23, 2021 • 5:45 AM



[https://slate.com/technology/2021/03/
section-230-reform-legislative-tracker.html](https://slate.com/technology/2021/03/section-230-reform-legislative-tracker.html)

- ▶ We don't know what impact any of these would have.

- ▶ We don't know what impact any of these would have.
- ▶ This need require inaction but it should encourage humility.

- ▶ We don't know what impact any of these would have.
- ▶ This need require inaction but it should encourage humility.
- ▶ Some argue that risks of action exceed the benefits

Opinion

The Right Way to Fight Fake News

Social media platforms need to make sure their anti-misinformation strategies are empirically grounded.

By Gordon Pennycook and David Rand

Dr. Pennycook and Dr. Rand are cognitive psychologists.

March 24, 2020

But just because a strategy sounds reasonable doesn't mean it works. Although the platforms are making some progress in their fight against misinformation, recent research by us and other scholars suggests that many of their tactics may be ineffective — and can even make matters worse, leading to confusion, not clarity, about the truth. Social media companies need to empirically investigate whether the concerns raised in these experiments are relevant to how their users are processing information on their platforms.

Opinion

The Right Way to Fight Fake News

Social media platforms need to make sure their anti-misinformation strategies are empirically grounded.

By Gordon Pennycook and David Rand

Dr. Pennycook and Dr. Rand are cognitive psychologists.

March 24, 2020

The obvious conclusion to draw from all this evidence is that social media platforms should rigorously test their ideas for combating fake news and not just rely on common sense or intuition about what will work. We realize that a more scientific and evidence-based approach takes time. But if these companies show that they are seriously committed to that research — being transparent about any evaluations that they conduct internally and collaborating more with outside independent researchers who will publish publicly accessible reports — the public, for its part, should be prepared to be patient and not demand instant results.

The obvious conclusion to draw from all this evidence is that social media platforms should rigorously test their ideas for combating fake news and not just rely on common sense or intuition about what will work. We realize that a more scientific and evidence-based approach takes time. But if these companies show that they are seriously committed to that research — being transparent about any evaluations that they conduct internally and collaborating more with outside independent researchers who will publish publicly accessible reports — the public, for its part, should be prepared to be patient and not demand instant results.

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer^{a,1}, Jamie E. Guillory^{b,2}, and Jeffrey T. Hancock^{b,c}

^aCore Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of ^bCommunication and ^cInformation Science, Cornell University, Ithaca, NY 14853

Let's see some interventions that can be tested

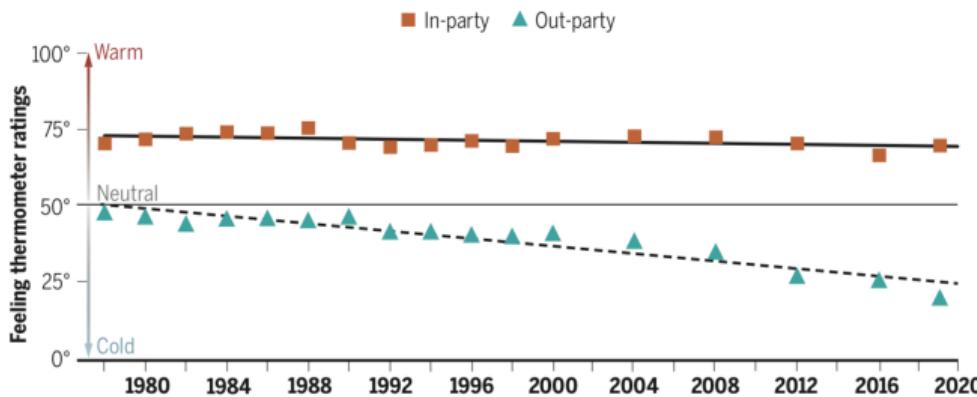
SOCIAL SCIENCE

Political sectarianism in America

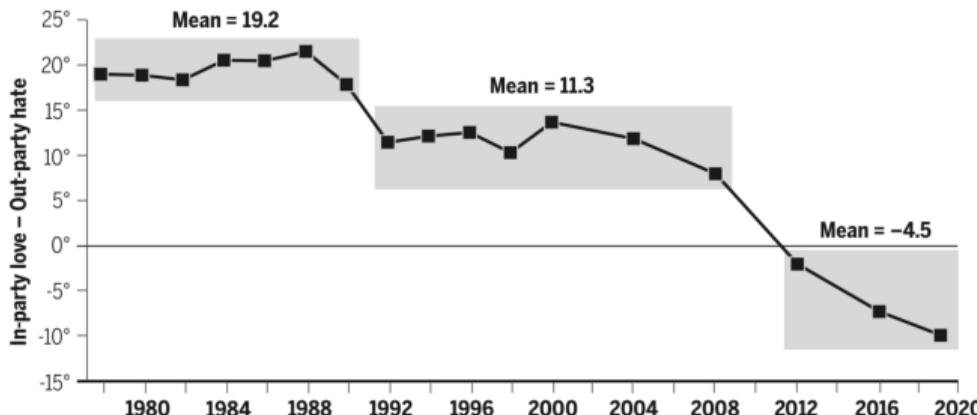
A poisonous cocktail of othering, aversion, and moralization poses a threat to democracy

By Eli J. Finkel¹, Christopher A. Bail², Mina Cikara³, Peter H. Ditto⁴, Shanto Iyengar⁵, Samara Klar⁶, Lilliana Mason⁷, Mary C. McGrath¹, Brendan Nyhan⁸, David G. Rand⁹, Linda J. Skitka¹⁰, Joshua A. Tucker¹¹, Jay J. Van Bavel¹¹, Cynthia S. Wang¹, James N. Druckman¹

Warmth toward the opposing party (out-party) has diminished for decades



Out-party hate has emerged as a stronger force than in-party love



The Filter Bubble

What [REDACTED] the [REDACTED]
[REDACTED]
[REDACTED] Internet [REDACTED]
[REDACTED]
[REDACTED] Is [REDACTED]
[REDACTED]
[REDACTED] Hiding [REDACTED]
[REDACTED]
[REDACTED] From [REDACTED]
[REDACTED]
[REDACTED] You [REDACTED]

Eli Pariser

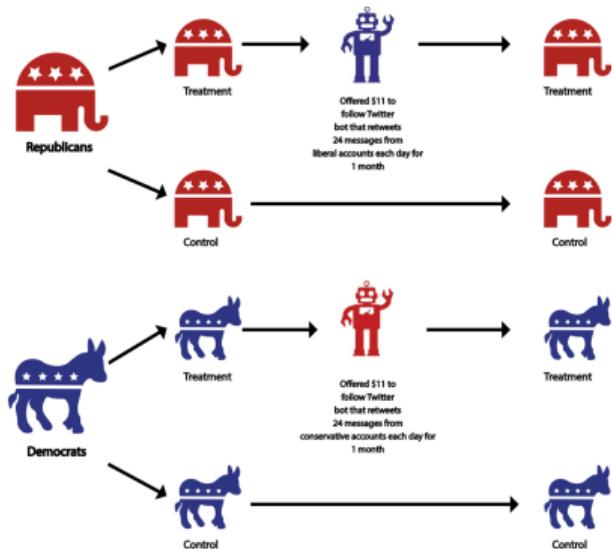
Filter bubbles can be created by algorithms or choices of people

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

^aDepartment of Sociology, Duke University, Durham, NC 27708; ^bDepartment of Political Science, Brigham Young University, Provo, UT 84602; ^cDepartment of Political Science, Duke University, Durham, NC 27708; ^dDepartment of Sociology, New York University, New York, NY 10012; and ^eDepartment of Statistical Science, Duke University, Durham, NC 27708

Initial Survey	Randomization	Weekly Surveys	Post-Survey
Respondents were offered \$11 to provide their Twitter ID and complete a 10-minute survey about their political attitudes, social media use, and media consumption habits (demographics provided by survey firm).	One week later, respondents were assigned to treatment and control conditions within strata created using pre-treatment covariates that describe attachment to party, frequency of Twitter use, and overall interest in current events.	Respondents in treatment conditions informed they are eligible to receive up to \$6 each week during the study period for correctly answering questions about the content of messages retweeted by Twitter bots.	Respondents were offered \$12 to repeat the pre-treatment survey one month after initial survey.



Preregistered hypotheses

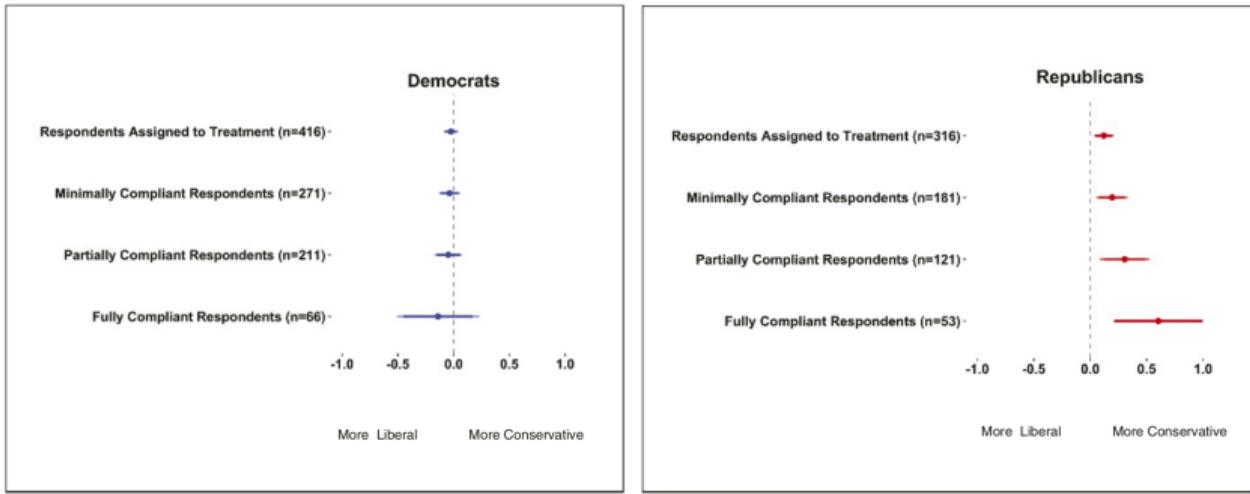
- ▶ disrupting selective exposure to partisan information will decrease political polarization

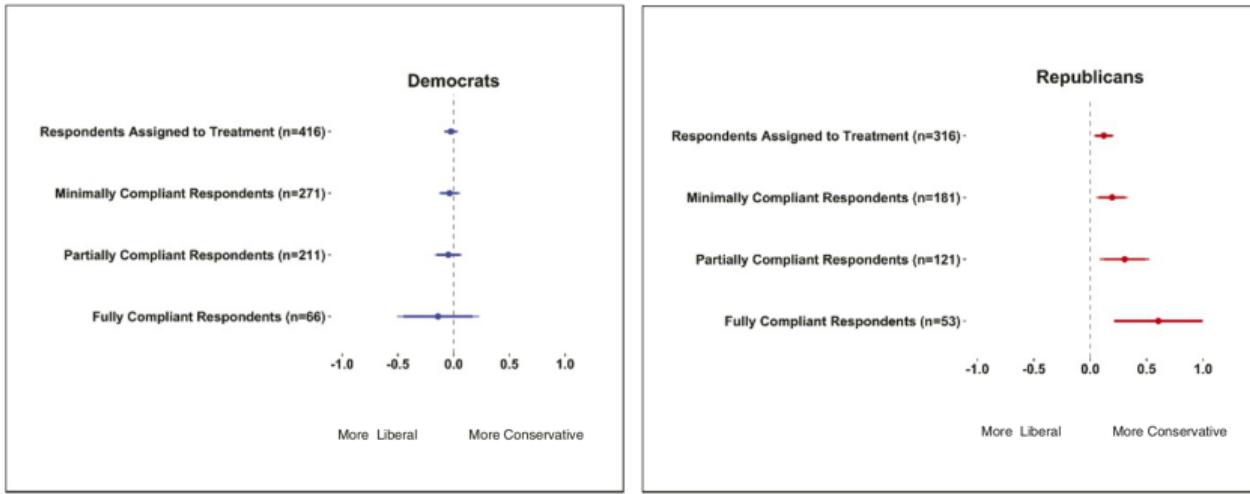
Preregistered hypotheses

- ▶ disrupting selective exposure to partisan information will decrease political polarization
- ▶ exposure to opposing political views can increase political polarization

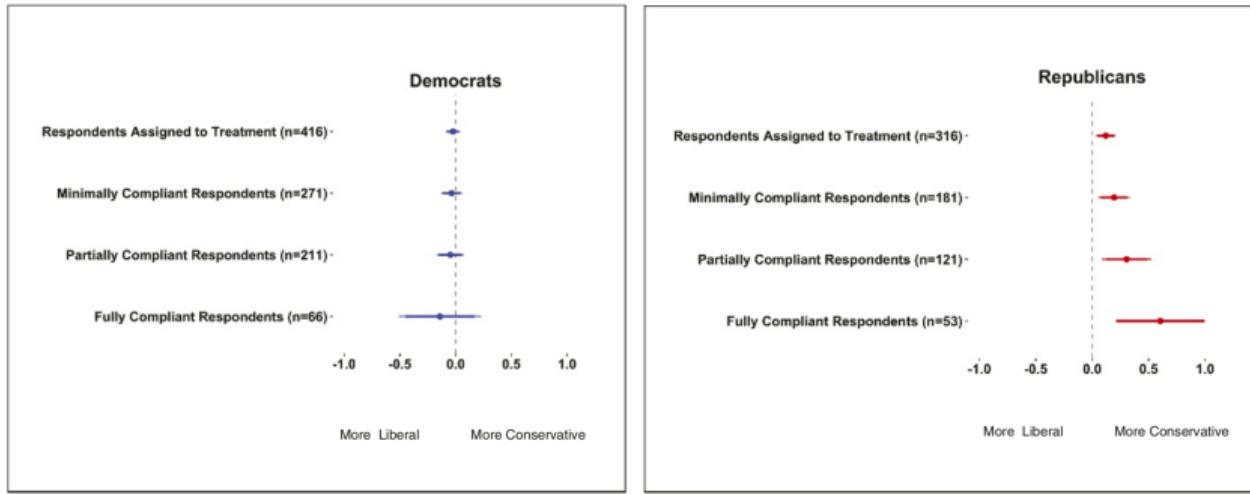
Preregistered hypotheses

- ▶ disrupting selective exposure to partisan information will decrease political polarization
- ▶ exposure to opposing political views can increase political polarization
- ▶ backfire effects are more likely to occur among conservatives than liberals

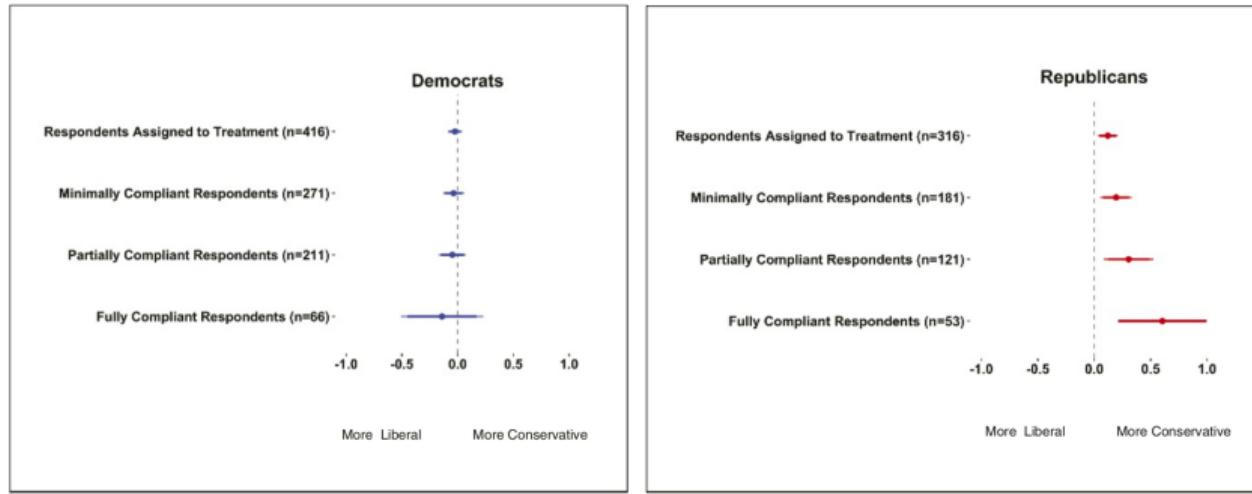




- ▶ Democrats appear to become slightly more liberal, Republicans become more conservative



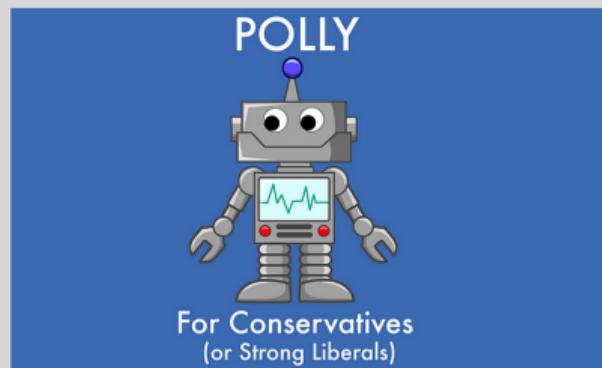
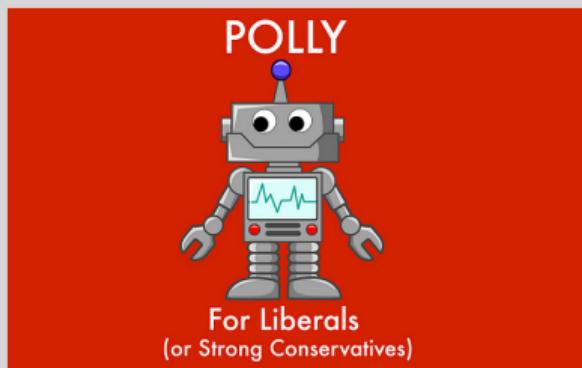
- ▶ Democrats appear to become slightly more liberal, Republicans become more conservative
- ▶ Higher levels of compliance show larger effects



- ▶ Democrats appear to become slightly more liberal, Republicans become more conservative
- ▶ Higher levels of compliance show larger effects
- ▶ unclear about exactly why this happened and whether it is specific to the way they constructed their bots (perhaps related to elites)

OUR BOTS

Our research suggests that stepping outside your echo chamber can make you more polarized not less. One of the reasons why this happens is that we become captivated by extremists on the other side who overshadow more moderate voices. The bots below retweet messages 12 by moderates each day who our research indicates resonates with members of the other party. We also screen out messages that score high on toxicity/incivility.



OUR TOOLS

In order to **fight back** against political polarization, we all need to learn more about how our own behavior is perceived by others, how to avoid trolls and other extremists, and how to identify other users with whom we can find common ground.

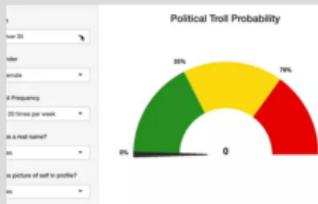
What Do Your Posts Say About Your Politics?



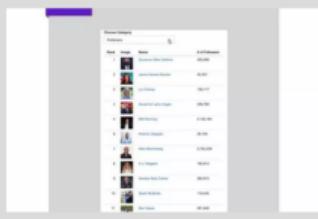
How Do Your Political Views Compare To Others?



Are You Being Trolled?



Social Media Bipartisanship Leaderboard



<https://www.polarizationlab.com/our-tools>

Adapting Security Warnings to Counter Online Disinformation

Ben Kaiser
Princeton University

Jerry Wei
Princeton University

Eli Lucherini
Princeton University

Kevin Lee
Princeton University

J. Nathan Matias
Cornell University

Jonathan Mayer
Princeton University

A search input field with a blue "Search" button to its right.

Shaping Europe's digital future

[Home](#) [Policies](#) [News](#) [Library](#) [Funding](#) [Calendar](#) [Consultations](#)

[Home](#) > [Policies](#) > Tackling online disinformation

POLICY | 16 MARCH 2021

Tackling online disinformation

The Commission is tackling the spread of online disinformation and misinformation to ensure the protection of European values and democratic systems.

<https://digital-strategy.ec.europa.eu/en/policies/online-disinformation>

3 common ways platforms deal with it

- ▶ derank

3 common ways platforms deal with it

- ▶ derank
- ▶ remove content and people

3 common ways platforms deal with it

- ▶ derank
- ▶ remove content and people
- ▶ add warnings

3 common ways platforms deal with it

- ▶ derank
- ▶ remove content and people
- ▶ add warnings

These approaches have different implications for free speech norms



Donald J. Trump ✅ @realDonaldTrump · 9h

There is NO WAY (ZERO!) that Mail-In Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, ballots will be forged & even illegally printed out & fraudulently signed. The Governor of California is sending Ballots to millions of people, anyone.....



[Get the facts about mail-in ballots](#)



22K



24.3K



86.7K



Donald J. Trump ✅ @realDonaldTrump · 9h

...living in the state, no matter who they are or how they got there, will get one. That will be followed up with professionals telling all of these people, many of whom have never even thought of voting before, how, and for whom, to vote. This will be a Rigged Election. No way!



[Get the facts about mail-in ballots](#)

(a) Contextual warning



Donald J. Trump ✅ @realDonaldTrump · 1h

Some or all of the content shared in this Tweet is disputed and might be misleading about an election or other civic process. [Learn more](#)

View



(b) Interstitial warning

What effects might these have?

<https://techcrunch.com/2020/05/26/twitter-trump-labels-fact-checking-tweet/>

<https://abcnews.go.com/Technology/twitter-facebook-slap-labels-trumps-misleading-election-posts/story?id=74020537>

Adapting Security Warnings to Counter Online Disinformation

Ben Kaiser Princeton University	Jerry Wei Princeton University	Eli Lucherini Princeton University	Kevin Lee Princeton University
	J. Nathan Matias Cornell University	Jonathan Mayer Princeton University	

- ▶ 2 studies: 1 lab study with Princeton students, 1 online study on Mechanical Turk workers

Adapting Security Warnings to Counter Online Disinformation

Ben Kaiser
Princeton University

Jerry Wei
Princeton University

Eli Lucherini
Princeton University

Kevin Lee
Princeton University

J. Nathan Matias
Cornell University

Jonathan Mayer
Princeton University

- ▶ 2 studies: 1 lab study with Princeton students, 1 online study on Mechanical Turk workers
- ▶ no cooperation from any social media platform

Adapting Security Warnings to Counter Online Disinformation

Ben Kaiser
Princeton University

Jerry Wei
Princeton University

Eli Lucherini
Princeton University

Kevin Lee
Princeton University

J. Nathan Matias
Cornell University

Jonathan Mayer
Princeton University

- ▶ 2 studies: 1 lab study with Princeton students, 1 online study on Mechanical Turk workers
- ▶ no cooperation from any social media platform
- ▶ does not address the challenge of deciding what is misinformation

Study 1: Lab study of Princeton students

 News - Latest - The News Lens International Edition
<https://international.thenewslens.com/news> ▾
This site may contain misleading information. Consider using another site.
News Latest Related Tags: Han Kuo-yu, Tsai Ing-wen, Daguan, Taiwan 2020 presidential ... Former Taiwan Premier William Lai Enters 2020 Presidential Race.

(a) Contextual warning

 thenewslens.com may contain disinformation

The website you are trying to navigate to has been flagged as potentially containing disinformation.

If you trust the credibility of the website, you can disregard this warning. Otherwise, try finding a different website for the information you are seeking.

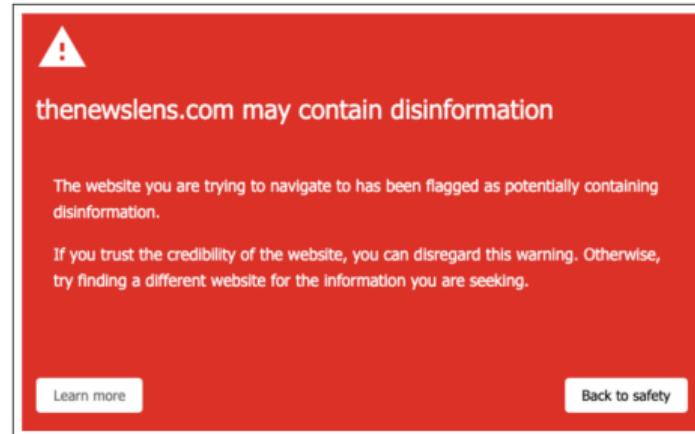
[Learn more](#) [Back to safety](#)

(b) Interstitial warning

Study 1: Lab study of Princeton students



(a) Contextual warning



(b) Interstitial warning

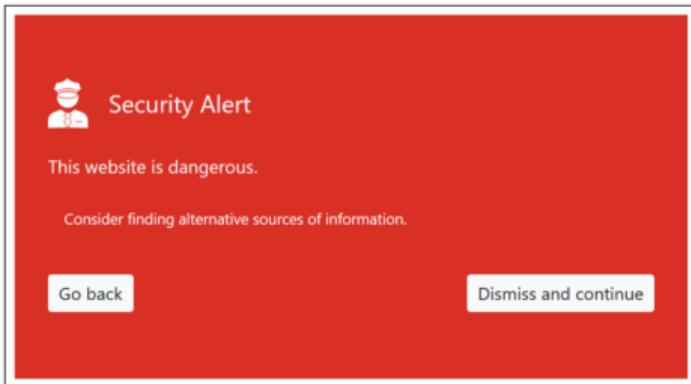
Quantitative metrics:

- ▶ Clickthrough rate (dismiss warning and proceed)
- ▶ Alternative visit rate (proportion that visit alternative site as desired)

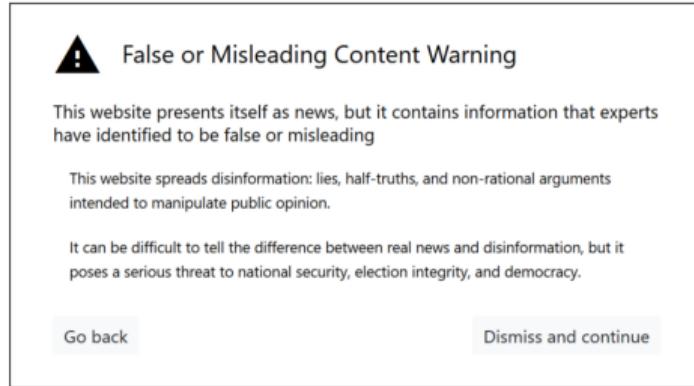
Round	Participant Instructions	Contextual Warning		Interstitial Warning	
		CTR	AVR	CTR	AVR
Control 1	Find the total area of Italy in square kilometers on Wikipedia or WorldAtlas.	–	–	–	–
Control 2	Report the price of a pair of men's New Balance 574 on JoesNewBalanceOutlet or 6pm.com.	–	–	–	–
Treatment 1	Find the political party of Taiwan's Premier on TheNewsLens or FocusTaiwan.	15/20	7/20	7/20	13/20
Treatment 2	Find the name of the girl reported missing in Barbados on Feb 11, 2019 on BarbadosToday or LoopNewsBarbados.	18/20	4/20	11/20	10/20

- ▶ Interstitial warning works better by both metrics (clickthrough rate and alternative visit rate)

Study 2: Workers on Amazon Mechanical Turk. Find the “best” interstitial warning.



(a) Harm

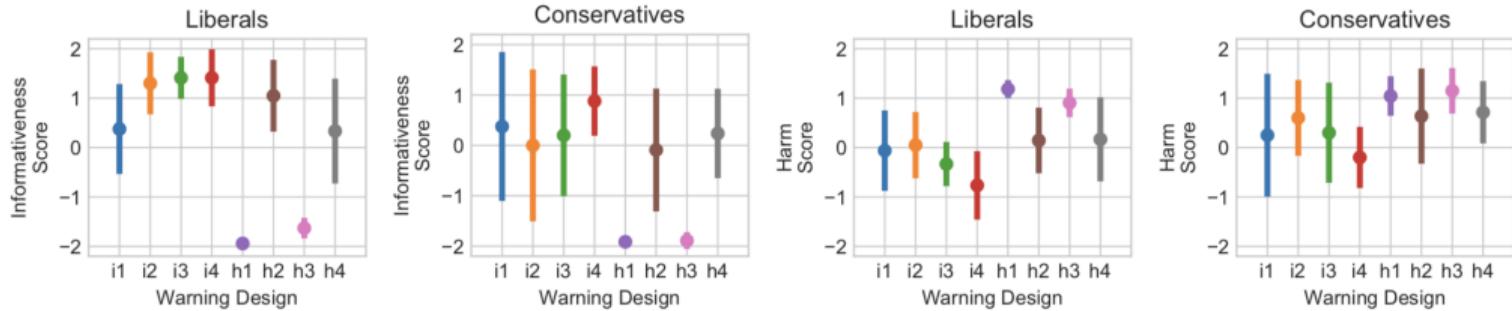


(b) Informative

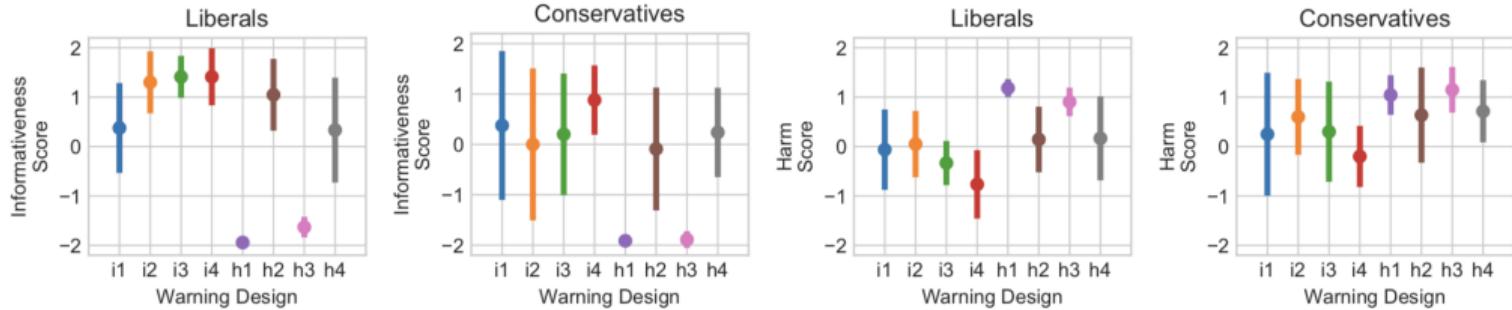
	Harm (white on red background)				Informativeness (black on white background)			
ID	h1	h2	h3	h4	i1	i2	i3	i4
Icon	Skull 	Skull 	Policeman 	Policeman 	Exclamation 	Policeman 	Policeman 	Exclamation 
Title	WARNING	Security Alert	Security Alert	Security Alert	False or Misleading Content Warning	Fake News Warning	False or Misleading Content Warning	Fake News Warning
Primary message	This website is dangerous.	This website contains misleading or false information.	This website is dangerous.	This website contains misleading or false information.	This website presents itself as news, but it contains information that experts have identified to be false or misleading	This website contains misleading or false information.	This website contains misleading or false information.	This website presents itself as news, but it contains information that experts have identified to be false or misleading
Details	None	None	Consider finding alternative sources of information.	None	<p>This website spreads disinformation: lies, half-truths, and non-rational arguments intended to manipulate public opinion.</p> <p>It can be difficult to tell the difference between real news and disinformation, but it poses a serious threat to national security, election integrity, and democracy.</p>	<p>This website spreads disinformation: lies, half-truths, and non-rational arguments intended to manipulate public opinion.</p> <p>It can be difficult to tell the difference between real news and disinformation, but it poses a serious threat to national security, election integrity, and democracy.</p>	Consider finding alternative sources of information.	Consider finding alternative sources of information.

Pre-registered quantitative metrics:

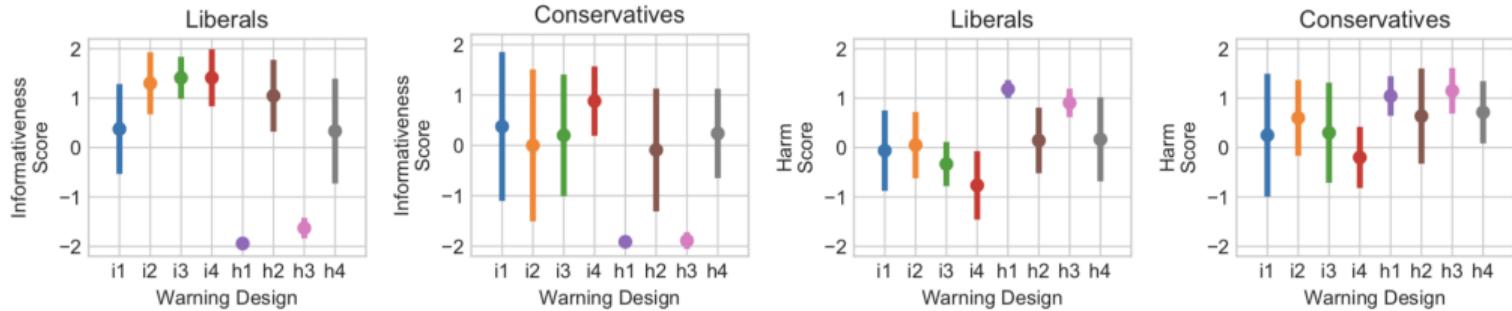
- ▶ Clickthrough rate (dismiss warning and proceed)
- ▶ Alternative visit rate (proportion that visit alternative site as desired)
- ▶ Information score (based on survey)
- ▶ Harm score (based on survey)



- ▶ With a few exceptions, not big differences between warnings, but small differences at scale can matter.



- ▶ With a few exceptions, not big differences between warnings, but small differences at scale can matter.
- ▶ Not big differences between liberals and conservatives, but small differences at scale can matter.



- ▶ With a few exceptions, not big differences between warnings, but small differences at scale can matter.
- ▶ Not big differences between liberals and conservatives, but small differences at scale can matter.
- ▶ If these were to be deployed, you would want to understand differences for many different subgroups.

	Liberal					Conservative				
	#	AVR	CTR	\bar{i}	\bar{h}	#	AVR	CTR	\bar{i}	\bar{h}
Control	318	20%	—	—	—	158	16%	—	—	—
Treatment	318	87%	16%	—	—	158	85%	17%	—	—
Selected treatments										
h1	120	85%	18%	-1.94 ± 0.06	1.18 ± 0.18	46	83%	17%	-1.91 ± 0.11	—
h3	73	84%	18%	—	—	27	81%	22%	—	1.15 ± 0.46
i3	39	87%	13%	1.41 ± 0.43	—	10	90%	10%	—	—
i4	17	82%	12%	—	-0.76 ± 0.69	25	76%	24%	0.88 ± 0.69	-0.2 ± 0.62

- With a few exceptions, not big differences between warnings, but small differences at scale can matter.

	Liberal					Conservative				
	#	AVR	CTR	\bar{i}	\bar{h}	#	AVR	CTR	\bar{i}	\bar{h}
Control	318	20%	—	—	—	158	16%	—	—	—
Treatment	318	87%	16%	—	—	158	85%	17%	—	—
Selected treatments										
h1	120	85%	18%	-1.94 ± 0.06	1.18 ± 0.18	46	83%	17%	-1.91 ± 0.11	—
h3	73	84%	18%	—	—	27	81%	22%	—	1.15 ± 0.46
i3	39	87%	13%	1.41 ± 0.43	—	10	90%	10%	—	—
i4	17	82%	12%	—	-0.76 ± 0.69	25	76%	24%	0.88 ± 0.69	-0.2 ± 0.62

- ▶ With a few exceptions, not big differences between warnings, but small differences at scale can matter.
- ▶ Sample sizes are really small here so we should have caution



Donald J. Trump ✅ @realDonaldTrump · 9h

There is NO WAY (ZERO!) that Mail-In Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, ballots will be forged & even illegally printed out & fraudulently signed. The Governor of California is sending Ballots to millions of people, anyone.....



[Get the facts about mail-in ballots](#)



22K



24.3K



86.7K



Donald J. Trump ✅ @realDonaldTrump · 9h

...living in the state, no matter who they are or how they got there, will get one. That will be followed up with professionals telling all of these people, many of whom have never even thought of voting before, how, and for whom, to vote. This will be a Rigged Election. No way!



[Get the facts about mail-in ballots](#)

(a) Contextual warning



Donald J. Trump ✅ @realDonaldTrump · 1h

Some or all of the content shared in this Tweet is disputed and might be misleading about an election or other civic process. [Learn more](#)

[View](#)



(b) Interstitial warning

What effects might these have?

<https://techcrunch.com/2020/05/26/twitter-trump-labels-fact-checking-tweet/>

<https://abcnews.go.com/Technology/twitter-facebook-slap-labels-trumps-misleading-election-posts/story?id=74020537>

Stepping back

- ▶ Interventions to change social media could be either structural changes or tweaks

Stepping back

- ▶ Interventions to change social media could be either structural changes or tweaks
- ▶ We have solid evidence that tweaks don't always have the intended effects (Bail et al.)

Stepping back

- ▶ Interventions to change social media could be either structural changes or tweaks
- ▶ We have solid evidence that tweaks don't always have the intended effects (Bail et al.)
- ▶ We have solid evidence that the relative effectiveness of tweaks is hard to anticipate (Kaiser et al.)

Stepping back

- ▶ Interventions to change social media could be either structural changes or tweaks
- ▶ We have solid evidence that tweaks don't always have the intended effects (Bail et al.)
- ▶ We have solid evidence that the relative effectiveness of tweaks is hard to anticipate (Kaiser et al.)
- ▶ It will be even hard to anticipate all the effects of structural changes, but that does not have to be a recipe for inaction

- ▶ Lecture 17: Social media and individuals
- ▶ Lecture 18: Social media and society
- ▶ Lecture 19: Social ads in social media
- ▶ [Lecture 20: Fixing social media](#)
- ▶ Lecture 21: Facebook Files

- ▶ Lecture 17: Social media and individuals
- ▶ Lecture 18: Social media and society
- ▶ Lecture 19: Social ads in social media
- ▶ Lecture 20: Fixing social media
- ▶ [Lecture 21: Facebook Files](#)

the facebook files



A Wall Street Journal investigation

Get started early!