

Class 21: Fixing social media

Matthew J. Salganik

Sociology 204: Social Networks
Princeton University

2/2 Possible effects of interventions to fix social media



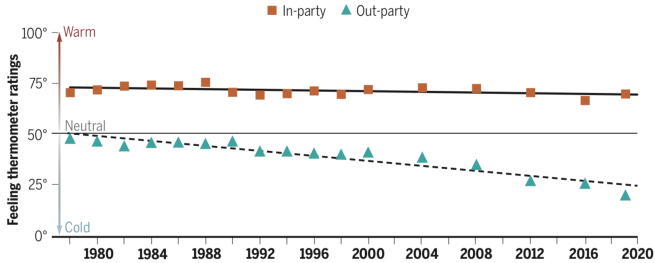
SOCIAL SCIENCE

Political sectarianism in America

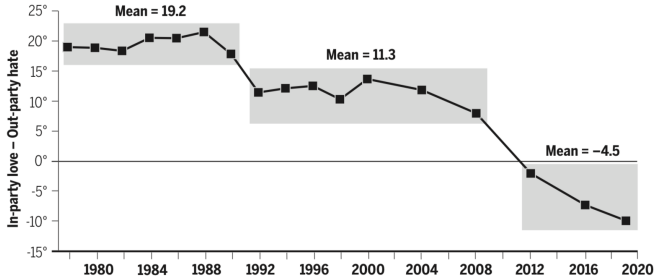
A poisonous cocktail of othering, aversion, and moralization poses a threat to democracy

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

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).

Randomization

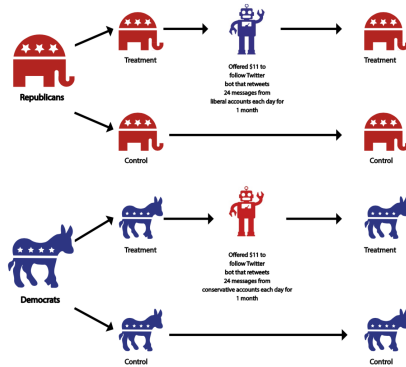
One week later, respondents were assigned to treatment and control conditions within strata created using pre-treatment covariates that describe attachment to party, frequency of Twitter use, and overall interest in current events.

Weekly Surveys

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.

Post-Survey

Respondents were offered \$12 to repeat the pre-treatment survey one month after initial survey.



Preregistered hypotheses

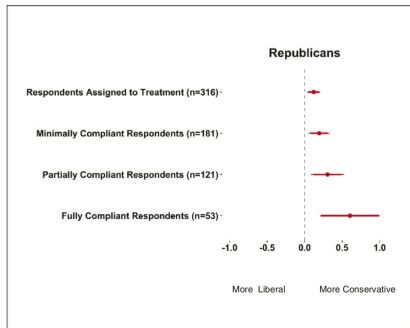
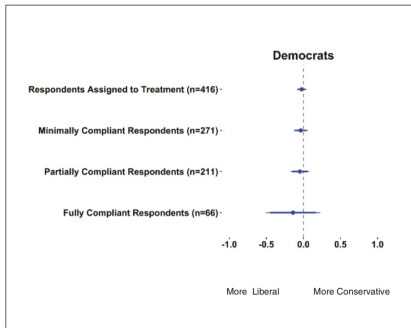
- ▶ disrupting selection exposure to partisan information will decrease political polarization

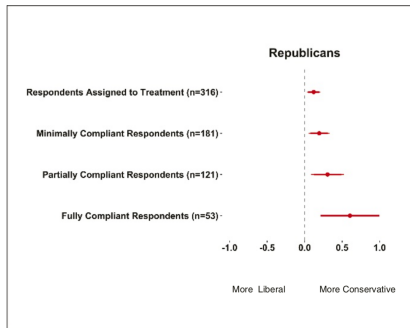
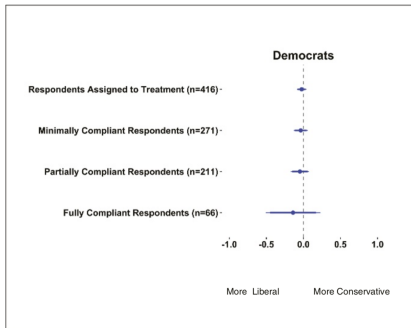
Preregistered hypotheses

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

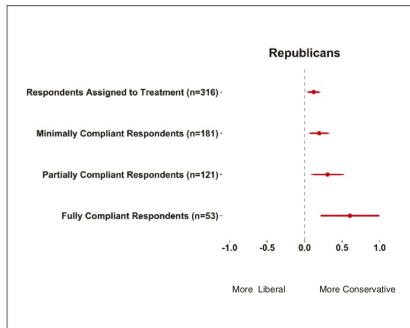
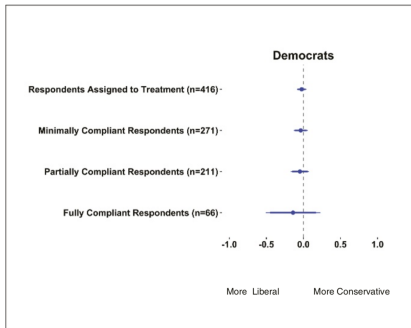
Preregistered hypotheses

- ▶ disrupting selection 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

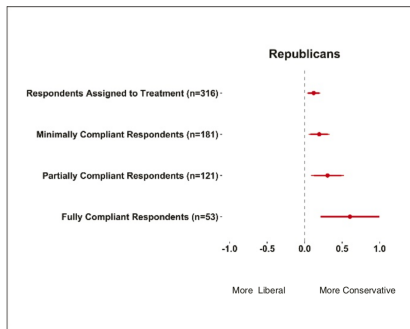
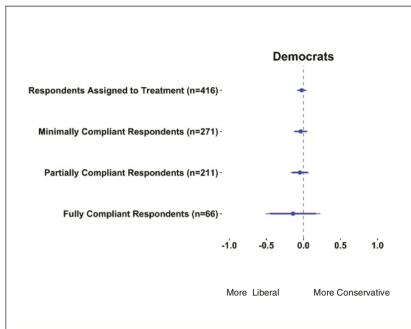




- ▶ Democrats appear to become slightly more liberal, Republicans become more conservative



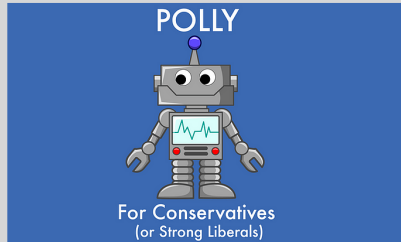
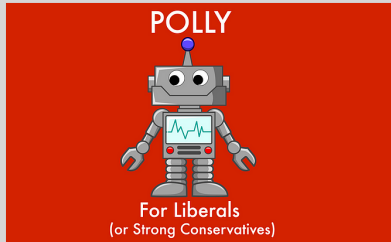
- ▶ Democrats appear to become slightly more liberal, Republicans become more conservative
- ▶ Higher levels of compliance show larger effects



- ▶ 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

OUR BOTS

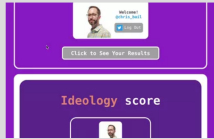
Our research suggests that stepping outside your echo chamber can make you more polarized not less. One of the reasons why this happens is that we become captivated by extremists on the other side who overshadow more moderate voices. The bots below retweet messages 12 by moderates each day who our research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



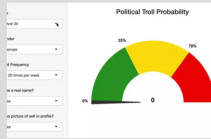
OUR TOOLS

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?



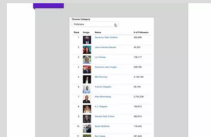
Are You Being Trolled?



How Do Your Political Views Compare To Others?



Social Media Bipartisanship Leaderboard



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

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



(a) Contextual warning



(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

Elena 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

Elena 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

Elena Lucherini
Princeton University

Kevin Lee
Princeton University

J. Nathan Matias
Cornell University

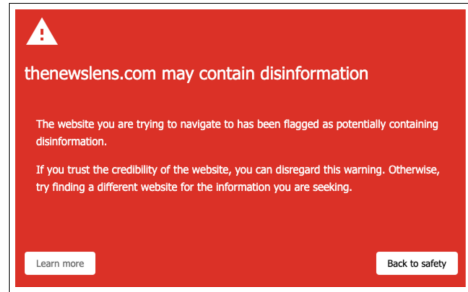
Jonathan Mayer
Princeton University

- ▶ 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



(a) Contextual warning

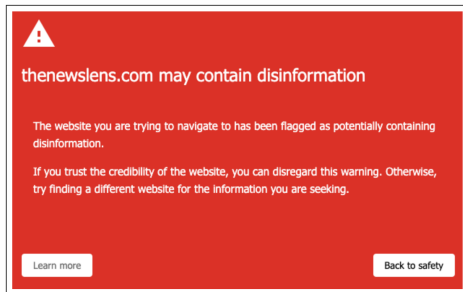


(b) Interstitial warning

Study 1: Lab study of Princeton students



(a) Contextual warning



(b) Interstitial warning

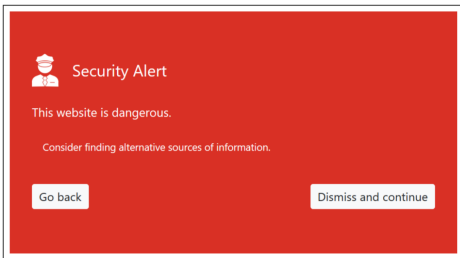
Quantitative metrics:

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

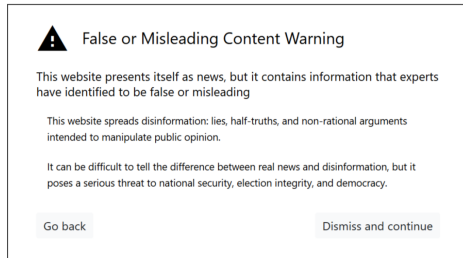
Round	Participant Instructions	Contextual Warning		Interstitial Warning	
		CTR	AVR	CTR	AVR
Control 1	Find the total area of Italy in square kilometers on Wikipedia or WorldAtlas.	–	–	–	–
Control 2	Report the price of a pair of men's New Balance 574 on JoesNewBalanceOutlet or 6pm.com.	–	–	–	–
Treatment 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

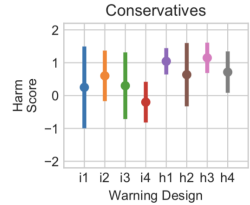
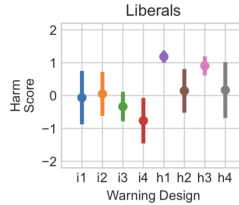
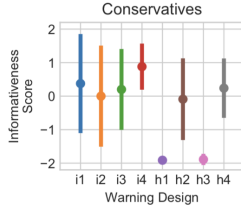
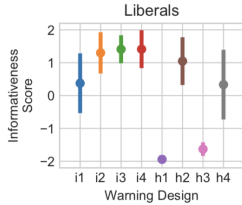


(b) Informative

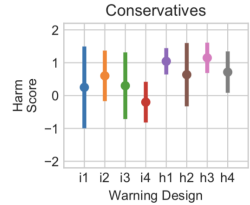
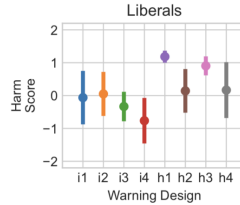
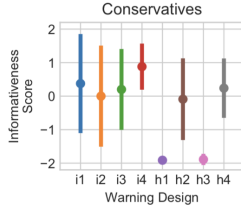
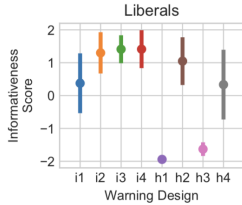
	Harm (white on red background)				Informativeness (black on white background)			
ID	h1	h2	h3	h4	i1	i2	i3	i4
Icon								
Title	WARNING	Security Alert	Security Alert	Security Alert	False or Misleading Content Warning	Fake News Warning	False or Misleading Content Warning	Fake News Warning
Primary message	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.

Quantitative metrics:

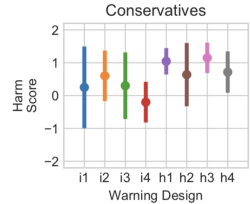
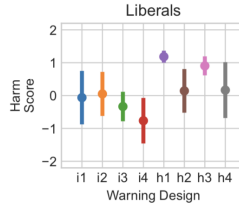
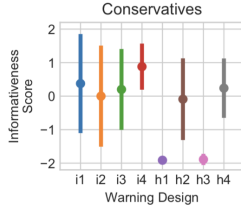
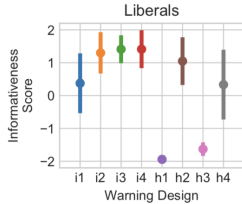
- ▶ Clickthrough rate (dismiss warning and proceed)
- ▶ Alternative visit rate (proportion that visit alternative site as desired)
- ▶ Information score (based on survey)
- ▶ Harm score (based on survey)



- 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



(a) Contextual warning



(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>

Conclusions:

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

Conclusions:

- ▶ 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.)

Conclusions:

- ▶ 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.)

Conclusions:

- ▶ 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

Social media:

- ▶ Social media and individuals
- ▶ Social media and society
- ▶ Social media and social ads
- ▶ Fixing social media

Class 22: Network scale-up method to study groups most at-risk for HIV

- ▶ Feehan, D.M. and Salganik, M.J. (2016). Generalizing the Network Scale-up Method: A New Estimator for the Size of Hidden Populations. *Sociological Methodology*.
- ▶ Feehan, D.M. et al. (2016). Quality vs. Quantity: A survey experiment to improve the network scale-up method. *American Journal of Epidemiology*.