ICA 2019 Computational Methods Interest Group Preconference

Understanding Partisan Bias in Media Use

How Computational Methods Can Inform Theory

JungHwan Yang

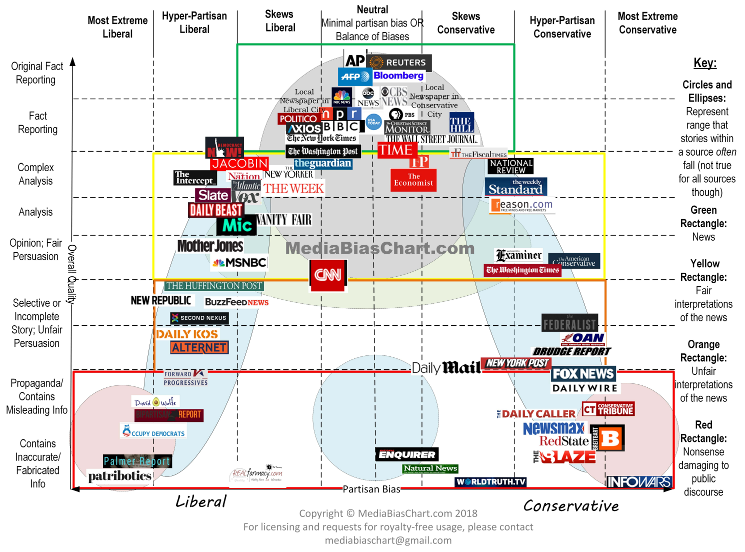
May 24, 2019

Department of Communication
University of Illinois at Urbana-Champaign

Outline

Introduction

High-choice information environment



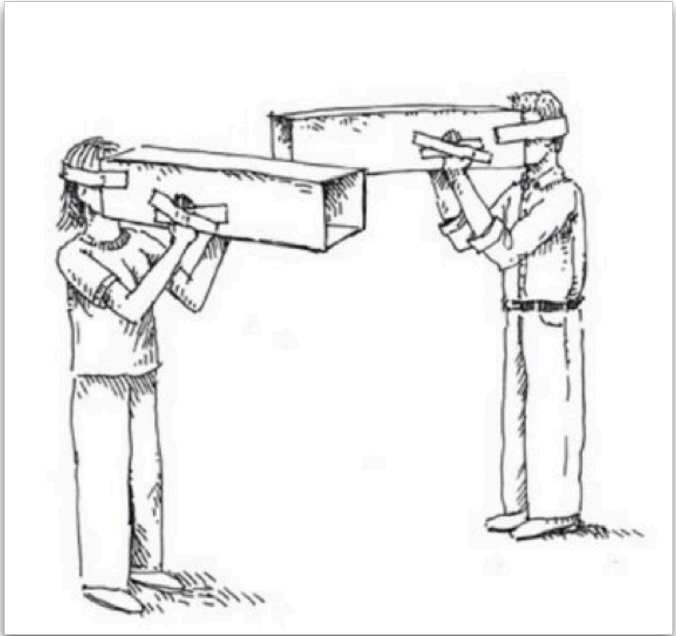
Personalized information consumption

Information curation is inevitable

- A growing number of sources
- A wider range of information
- Recommendation algorithms

Political selective exposure

- Individual's tendency to favor information which aligns with their pre-existing political views while avoiding contradictory information.
- Psychological factor: Cognitive dissonance
- Environmental factor: *De facto* selective exposure



Evolving discourse on selective exposure

The Washington Post

Democracy Dies in Darkness

In Theory • Opinion

Confirmed: Echo chambers exist on social media. So what do we do about them?

By **Christine Emba** July 14, 2016 Email the author

MOSTAFA M. EL-BERMAWY BUSINESS 11.18.16 05:45 AM

YOUR FILTER BUBBLE IS DESTROYING DEMOCRACY

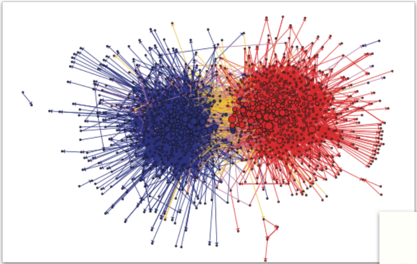
2016 Presidential Election – Digital Analysis by the Numbers

Hillary Clinton	Criteria	Donald Trump
238.5M	Total Social Media Shares	256.5M
16.633	Average Shares per Post	17.894
8.3M	Facebook Page Likes (Official Page)	12.2M
10.3M	Twitter Following (Official Page)	13.1M
32.9K	Number of Referring Domains	21.4K
1.51M	Number of Backlinks to Website	960K
550	Alexa Rank in The US	681
"There are five living U.S. presidents. None of themsupport Donald Trump." 1.7 M SHARES (by hillaryclinton.com)	Most Shared Article (Full Name)	"Why I'm Voting For Donald Trump." 1.5 M SHARES (by kellyquelette.wordpress.com)
165.000	Average Monthly Searches (Full Name)	7.480.000
$2.50	Google's Suggested Keyword Bid (Full Name)	$4.25

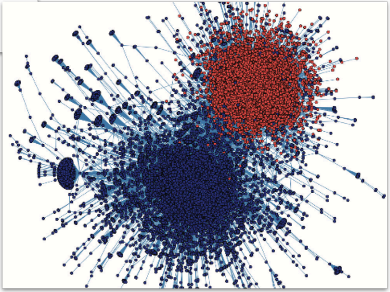
Sources: Ahrefs, Alexa, BuzzSumo, Facebook, Google AdWords, Twitter WIRED

SHARE

ON NOVEMBER 7, 2016, the day before the US election, I compared the number of social media followers, website performance, and Google search statistics of Hillary Clinton and Donald Trump. I was



Adamic and Glance (2005)



Conover et al. (2011)

Conflicting Arguments

Yes, selective exposure!

- There is a clear preference for like-minded information
- Fragmentation of news and filter bubble
- Political polarization

No, there is no clear pattern

- There is selective exposure but no selective avoidance
- No patterns of selective exposure in online news consumption

There are **conceptual** and **methodological** issues

Conceptual issues of selective exposure research

The power of metaphor

- Echo chambers, filter bubbles

Simplified interpretation of the findings

- Find evidence (60% vs. 40%) and conclude "there's a partisan selective exposure; thus, echo chambers exist."
- Statistical significance over statistical power

Lacking discussion about the *magnitude* of selective exposure

Measurement issues of selective exposure research

How long is the coastline of Britain?

Measurement issues of selective exposure research

How long is the coastline of Britain?

It depends on the length of your ruler.

Limitations in measurement

Experiment

- A small number of predefined set of options
- Low external validity

Survey

- Still small number of options (time=cost)
- Measurement errors in self-reports

They are likely to overestimate selectivity!

Potential benefits of computational methods

How can computational methods help?

CSS $\neq$ Big Data

Time-stamped measures

- Disaggregation of data

Data with higher external validity

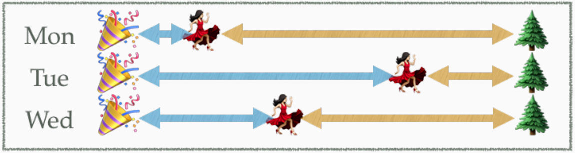
- Digital trace data from the real world
- Little restrictions in the number of options

Combining datasets

- Other classic methods are still valid
- We can integrate digital trace data with the survey/experiment data to understand the impact of information use on people

Study 1: Enhanced experiment

Media use is not an one-time event.



Strategy 1: Calculate the mean distance

Strategy 2: Sum of the choices

Selectivity = +1 -1 +1 = +1

Two different conclusions can be drawn from the same data.

BUT, WHICH ONE REFLECTS THE TRUTH?

Sequence of choice

Classic measures

- First choice: provides a set of choices, allows choice of one
- Main sources: Ask the most important source(S) people use
- Provide diverse options and count the choices

If we measure *information choice* with a time stamp:

- With a little help of *programming*, we can analyze data in a different angle
- Sequence of choice: the patterns of information choice over-time

Study context and participants

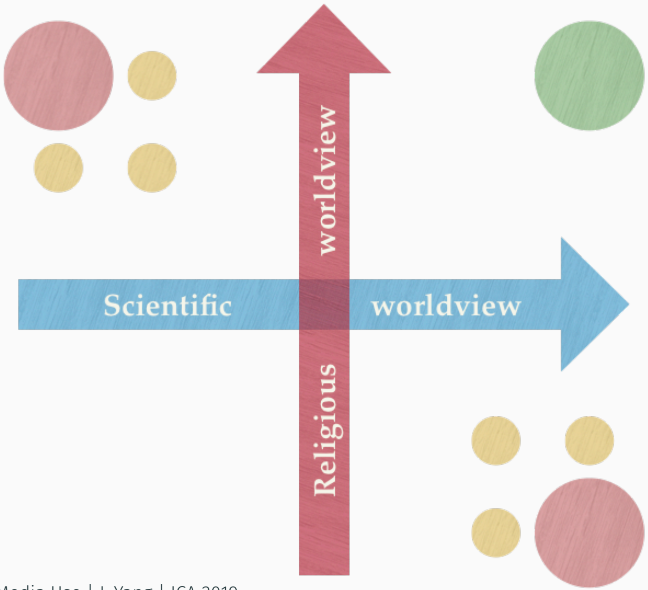
Conflict between religion and science

- The issue is heavily polarized ≈ partisan conflict

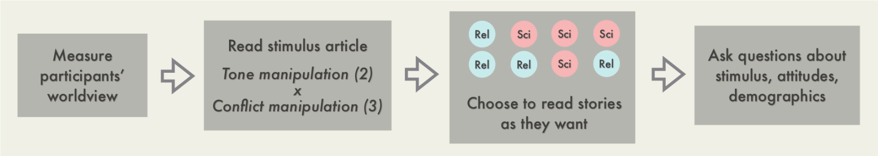
Participants

Partisans from both sides		National sample
Freedom From Religion foundation	The BioLogos Foundation	SSI Survey Sampling International
N=260	N=252	N=257

Partisan classification

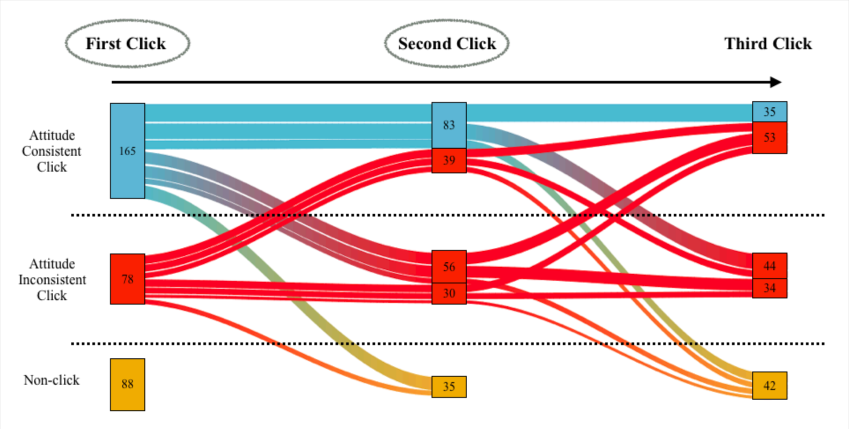


Experimental procedure

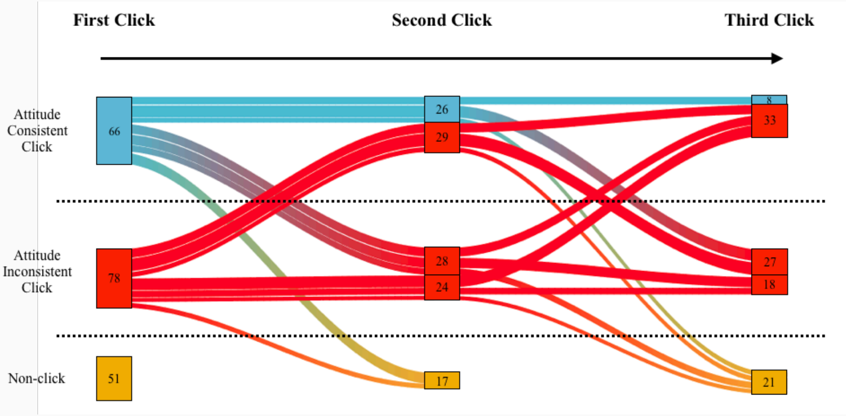


All selections are time-stamped.

Result: Strong partisans



Result: Weak partisans



Takeaway

- Some limited evidence of selective exposure
- The browsing pattern gives an interesting story + generate more questions
- An example of an application of computational methods to a small scale classic experiment

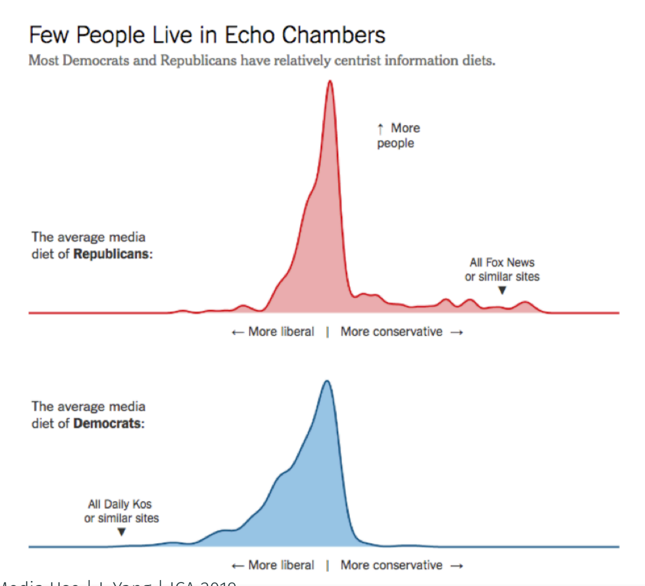
Study 2: Network personalization on Twitter

Digital trace data

Benefits of digital trace data

- Bypass measurement errors of self-reported question
- Real-world activity
- Comprehensive, large-scale data exist
- E.g., Web-tracking data, social media data, TV viewership data

Recent evidence: not-so-strong selective exposure



Are there echo chambers?

Missing points

- Popular sources draw audience from everywhere
- Are people really selective?
- The degree of selectivity in question

These points are related to the conceptual/measurement issues we discussed earlier.

Network personalization on Twitter

Who is following whom?

- The list of Twitter accounts users follow

Actual pattern of network curation

- Twitter users' decisions to follow a particular accounts
- Twitter as a heavily political space

No constraints in number of accounts you can observe

- The entire list of following network is available
- The long-tail, niche accounts

Goals of this study

- Questioning the echo chambers argument
- Understand the magnitude of partisan selective exposure
- Understand the role of the long tail

Research process (abbreviated version)

Find the right users

- Politically attentive, relevant to U.S. politics

Identify political preferences of those users

- Twitter users' decisions to follow particular accounts

Examine political slant of network personalization

Find political users and estimate political preferences

Political dictionary

- Combination of unigrams, bigrams, and Twitter specific texts (N≈700)
- Political tweets (N≈45,000)

Building the ground truth

- Human coded ground truth data (N≈1,800)
- Boosting the size of the data by automated process (N≈40,000)

Naïve Bayes Classifier

- Balanced accuracy of 0.84+
- Identified 13,000+ users' political preference: 29.1% liberal, 37% moderate, 33.8% conservative

Reconstructing Twitter network

Getting the following network

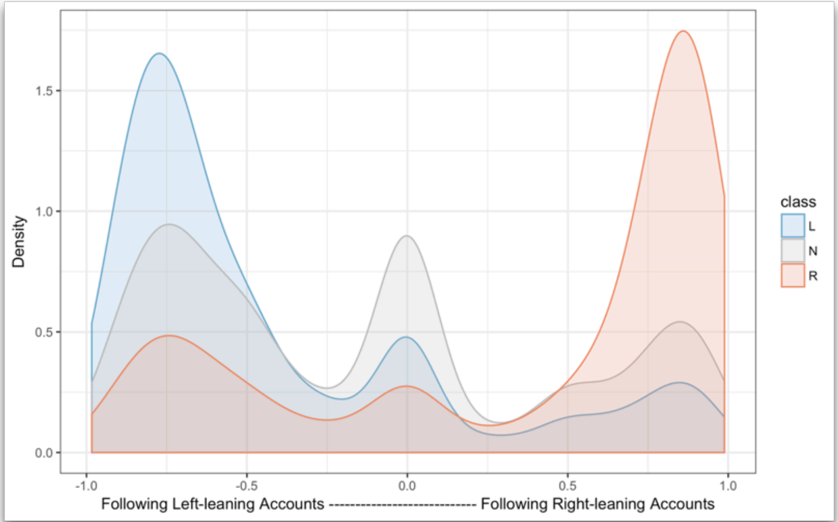
- Getting the entire accounts the 13,000+ Twitter users are following through Twitter REST API
- The 13,000+ users follow 4.4 million users (1.6 million unique users)

Identify most popular accounts

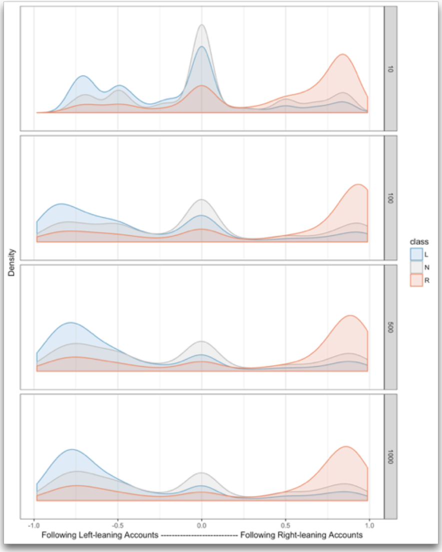
- Find the top 1,000 most followed accounts among the 4.4 million accounts
- Hand-coded these 1,000 accounts' political slant

Rank	Twitter Account	Name	Class	Follwing Number
1	@realDonaldTrump	Donald J. Trump	R	5949
2	@BarackObama	Barack Obama	L	5193
3	@POTUS	President Trump	R	5024
4	@HillaryClinton	Hillary Clinton	L	3874
5	@POTUS44	President Obama	L	3658
6	@seanhannity	Sean Hannity	R	3414
7	@KellyannePolls	Kellyanne Conway	R	3353
8	@FLOTUS	Melania Trump	R	3231
9	@mike_pence	Mike Pence	R	3183
10	@DonaldJTrumpJr	Donald Trump Jr.	R	3082
11	@FoxNews	Fox News	R	3026
12	@nytimes	The New York Times	N	2948
13	@AnnCoulter	Ann Coulter	R	2915
14	@IngrahamAngle	Laura Ingraham	R	2913
15	@cnnbrk	CNN Breaking News	N	2857
16	@SheriffClarke	David A. Clarke, Jr.	R	2816
17	@EricTrump	Eric Trump	R	2779
18	@VP	Vice President Pence	R	2735
19	@JudgeJeanine	Jeanine Pirro	R	2706
20	@IvankaTrump	Ivanka Trump	R	2659
21	@billoreilly	Bill O'Reilly	R	2625
22	@CNN	CNN	N	2614
23	@AP	The Associated Press	N	2606
24	@RealBenCarson	Ben & Candy Carson	R	2590
25	@TuckerCarlson	Tucker Carlson	R	2565
26	@RealJamesWoods	James Woods	R	2514
27	@newtgingrich	Newt Gingrich	R	2502
28	@SenSanders	Bernie Sanders	L	2470
29	@BernieSanders	Bernie Sanders	L	2420
30	@BBCBreaking	BBC Breaking News	N	2404

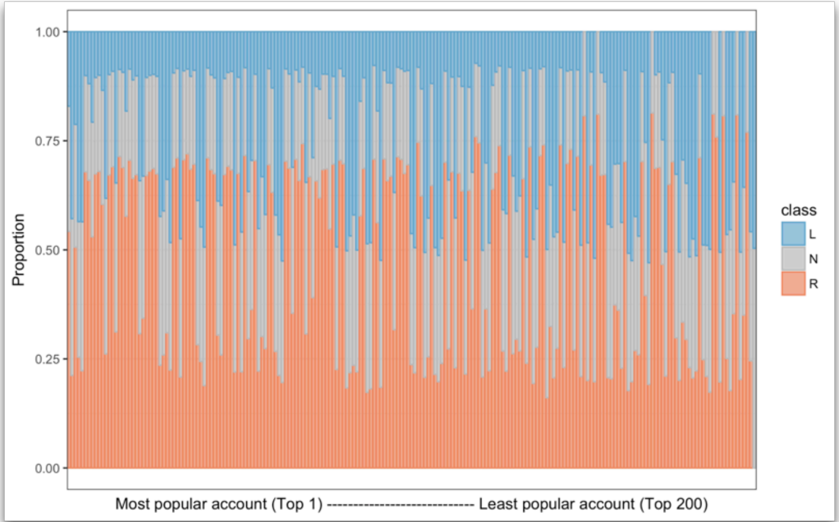
Result 1: overall preference for like-minded information



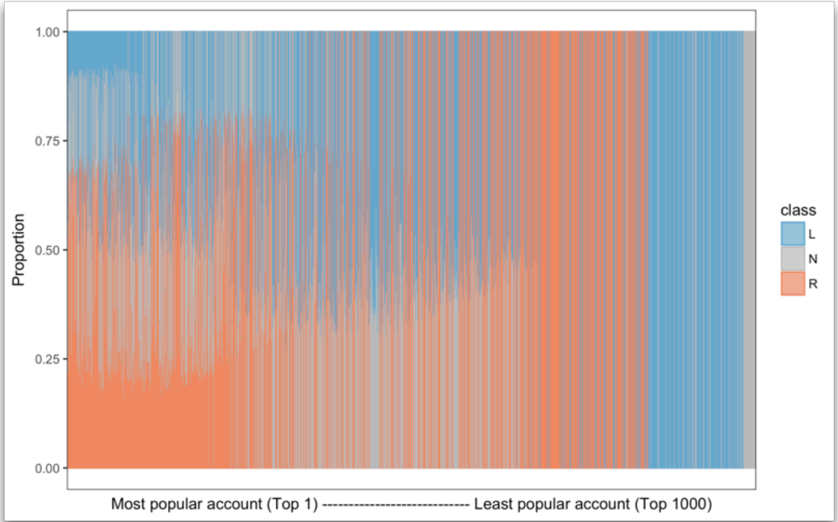
Result 2: the long-tail contributes partisan selctivity



Result 3: Ideological composition of the top 200 accounts



Result 4: Ideological composition of the top 1000 accounts



Takeaway

- Overall slant of personalized political Twitter network
- However, it is far from *Echo Chambers*
- The role niche news in network selectivity–It looks more like Echo Chambers in the long-tail

To see the full presentation, come to ***Computational Approaches to Political Communication (11 am, May 25 at Morgan)***

Study 3: An experiment on online partisan media exposure

Studying real-world effects of online partisan media

- Concerns about online echo chambers leading to extremism (Sunstein 2001)

Studying real-world effects of online partisan media

- Concerns about online echo chambers leading to extremism (Sunstein 2001)
- Is the internet fueling affective polarization? (Lelkes, Sood & Iyengar 2017; Settle 2018)

Studying real-world effects of online partisan media

- Concerns about online echo chambers leading to extremism (Sunstein 2001)
- Is the internet fueling affective polarization? (Lelkes, Sood & Iyengar 2017; Settle 2018)
- Are partisan media reducing trust in *other* media? (Ladd 2012)

Media effects and self-selection

- Lots of evidence on these questions come from lab or online survey experiments

Media effects and self-selection

- Lots of evidence on these questions come from lab or online survey experiments
- People can select into/out of congenial content but also into/out of politics altogether (Prior 2007; Arceneaux & Johnson 2013)

Media effects and self-selection

- Lots of evidence on these questions come from lab or online survey experiments
- People can select into/out of congenial content but also into/out of politics altogether (Prior 2007; Arceneaux & Johnson 2013)
- *Most* evidence focuses on effects of cable TV

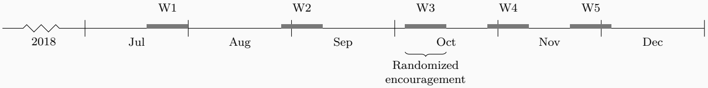
Media effects and self-selection

- Lots of evidence on these questions come from lab or online survey experiments
- People can select into/out of congenial content but also into/out of politics altogether (Prior 2007; Arceneaux & Johnson 2013)
- *Most* evidence focuses on effects of cable TV
- *But* does this still hold today?

Panel setup

Survey

- initially 1,500 respondents
- Quota sampling (gender, age, education) + weights
- Multiple waves (3 pre-election, 4 post-election)
- passive tracking data from YouGov Pulse (web + apps + social media)
- Treatment encouragement in wave 3



Preregistration at EGAP under `https://egap.org/registration/5311`

Randomized encouragement

Three encouragements

1. Change of browser's homepage to news website
2. Following the media outlet's Facebook page
3. Subscribing to media outlets' newsletters

Treatment conditions

- (1) Fox News ($N = 361$), (2) Huffington Post ($N = 360$); Control: no encouragement ($N = 377$)
- Blocked on browser used in Wave 2 (pre-treatment)

Outcomes under scrutiny

Political attitudes

- Issue positions
- Affective polarization
- Perceived polarization
- Trump approval
- Approval of Congress

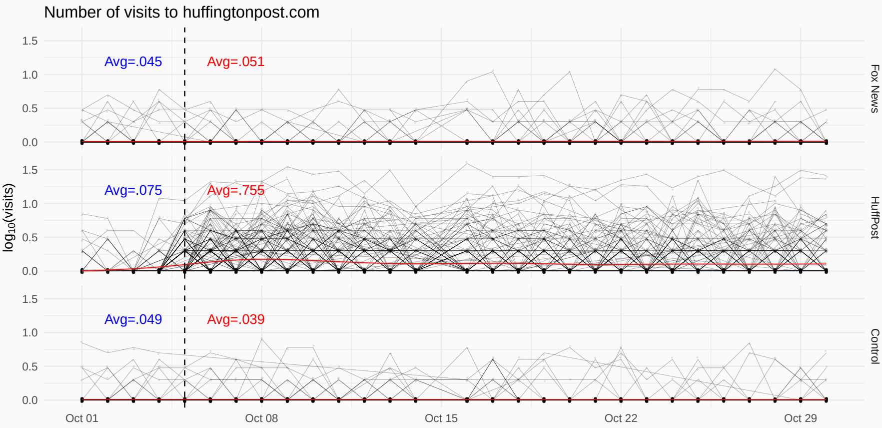
Political behavior and knowledge

- Self-reported turnout in midterms
- Political knowledge
- Event knowledge
- Social media sharing

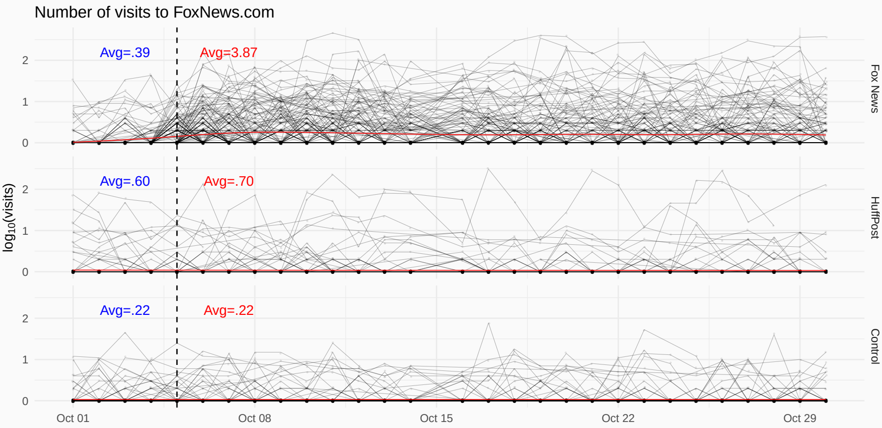
Media attitudes and consumption

- Media trust
- Perceived media bias
- Partisan media consumption
- Issue prioritization

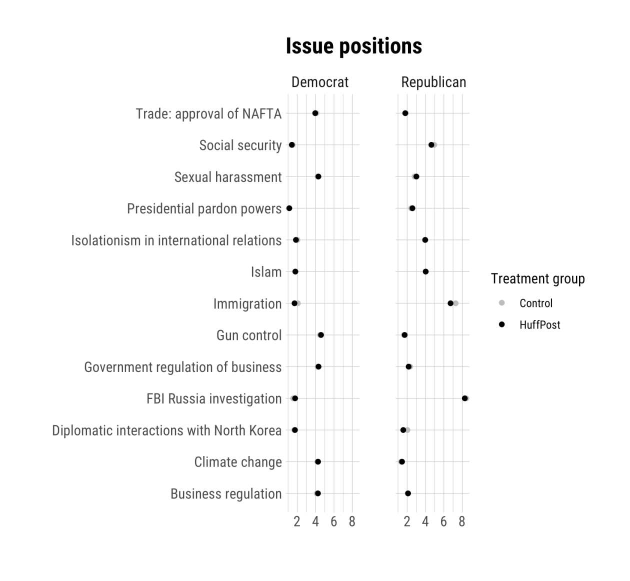
Results: checking compliance



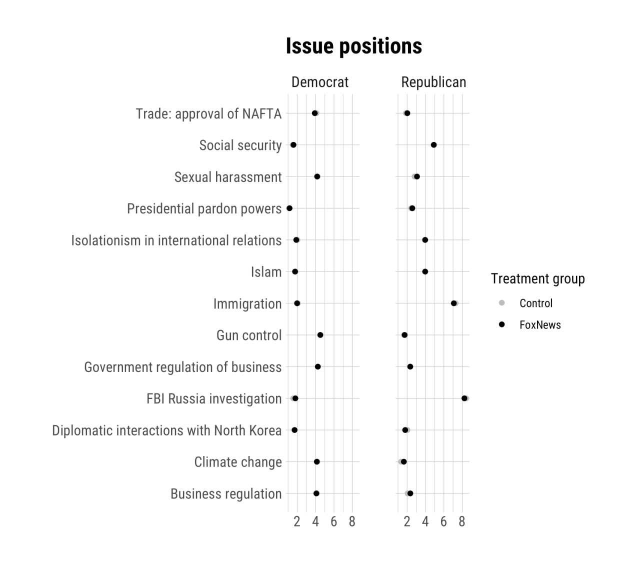
Results: checking compliance



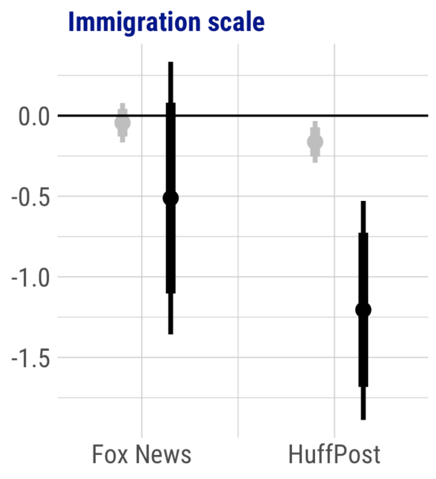
Results: issue positions (Huff Post)



Results: issue positions (Fox News)

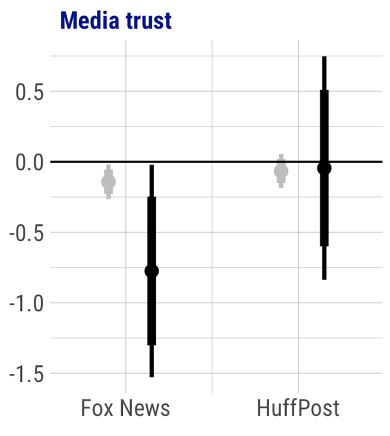


Results: immigration attitudes



- Grey: Intend-to-treat (ITT)
- Black: Local average treatment effect (LATE)

Results: media trust



"How much trust and confidence do you have in the press when it comes to reporting the news about government and politics fully, accurately, and fairly?" (4 = "A great deal")

Other results and takeaway

- Many null results including affective polarization, issue positions, and turnout
- Combining survey, experiment, and digital trace data to test theory more rigorously

To see the full presentation, come to ***Computational Approaches to Political Communication (11 am, May 25 at Morgan)***

Discussion

Summary

Enhanced experiment

- Computational methods are relevant to classic social science methods

Digital trace data from social media

- Data are available. The question is how you use the data.

Integration of survey, experiment, and digital trace data

- Connecting observational data with self-reports to understand the effects

Thank you!

JungHwan Yang (University of Illinois at Urbana–Champaign)
`junghwan@illinois.edu`

Albert C. Gunther (University of Wisconsin–Madison)
David Wise (University of Wisconsin–Madison)
Andy Guess (Princeton University)
Pablo Barberá (USC)
Simon Munzert (Hertie School of Governance)

VolkswagenStiftung