Pre-analysis plan: Social Influence and News Consumption

Alex Moehring* Carlos Molina†

March 26, 2023

Abstract

Populations in many countries have become decidedly more polarized over the last decades. Many believe that social media, which creates echo chamber-like interactions, is partly to blame. In principle, more intense communication between like-minded individuals can have two distinct impacts on political beliefs. First, individuals may receive a slanted diet of political news shared by their like-minded friends. Second, they may purposefully slant their own news consumption, and beliefs, in order to remain more in line with these friends. Despite the importance of these questions, there is little evidence for either of these two types of influences. This paper designs a unique field experiment on Twitter to separately identify both mechanisms. In our sample, politically-active individuals consume a highly slanted news diet. By varying what an individual's social media followers see about her news diet and tracking their news consumption and sharing behavior, we test if (1) this news diet has an impact on individual behavior and beliefs; and more importantly, (2) whether individuals manipulate their news diet in order to remain more in line with their friends when they believe that their choices will be observed by these friends. This is either because they would like to signal to their friends via their news diet choices or because they are afraid of certain reactions when they deviate from a news diet aligned with their friends' ideological position.

Keywords: news diet, disclosure, learning, Twitter.

JEL Classification: C93, D72, D83, L82.

*Massachusetts Institute of Technology, Sloan School of Management, 50 Memorial Drive, Cambridge, MA 02142. E-mail: moehring@mit.edu

†Massachusetts Institute of Technology, Department of Economics, 50 Memorial Drive, Cambridge, MA 02142. E-mail: camolina@mit.edu

1 Motivation

In the last few decades, the world has witnessed a structural change in how individuals interact with their peers. When it comes to in-person social capital, individuals now seem to be "bowling alone". Americans for example have reduced most forms of off-line social intercourse and political involvement[1] upon which they used to found, educate, and enrich the fabric of their social lives [Putnam et al., 2000]. On the other hand, new information technologies have fostered on-line interactions and potentially contributed to the propagation of homophily and bias in the information diffusion toward like-minded peers [Lelkes et al., 2017, Allcott et al., 2020, Boxell et al., 2020].

While there is suggestive evidence that this transition from offline to online communication due to the introduction of new information technologies is one important driver of the dramatic raise in affective polarization that we observe today [Sunstein, 2018, Levy, 2021], we know very little about the mechanisms through which interactions with others shape political beliefs and ultimately behavior.[2] How instrumental is the information that someone learns from their peers to determine their preferences for biased news?

In this study, we propose an online field experiment to overcome these challenges in the context of demand for biased information. Our experiment allows us to separate between two main channels through which peers can influence an individual's news diet and ultimately their political views and polarization: an *internal* one, where the news peers consume impacts the news diet of an individual through discovery or signaling quality and a *external* one, where individuals alter their news diets when observed by their peers.

We organize the rest of the document outlining the pre-analysis plan as follows. Section 2 describes the data we use to construct the relevant measures in our experiment. Section 3 outlines our experimental design. Section 4 discusses our hypothesis as well as our proposed empirical strategy and measures. Finally, Section 5 details the main outcomes.

2 Data

Our experimental design provides participants with a summary of the slant of their news diet. As we discuss in the next section, a subset of participants will also observe a summary of the news diet of their peers. We construct these summaries based on real-time data from the hard-news outlets the participants follow on Twitter and the outlets followed by their peers. In this section we start by discussing how we create our database of hard-news publisher and then discuss the construction of the news diet summary.

2.1 US hard-news publishers on Twitter

The population of publishers we consider are generated from a data set of scraped online news publishers. We consider the first three months of 2021 which contains 4,412 publishers and then exclude publishers that published fewer than 10 articles per month during this period, resulting in 3,049 publishers. We also exclude

[1]Including work with political parties, service on committees and attendance at public meetings.

[2]This is also challenging to estimate because it poses two challenges. First, an individual's preferences and her or his network exposure are typically jointly determined. Second, even considering the possibility of a random allocation on the network, in many settings individual choices can be either partially or completely observed by peers, which could cause individuals to alter their behavior.

publishers that do not have a ".com" or ".net" top-level domain and who do not primarily publish hard news.

We map publisher domains to Twitter handles by attempting to scrape each publishers' landing page and if a valid Twitter handle is linked to from the landing page, select this handle. If no handle is found, we then use the Twitter API to search for the publisher domain and choose the Twitter handle with the most followers out of the top 20 handles returned. This is an error-prone algorithm, so we manually verified the handles selected are correct and if not, we manually search Twitter and the publisher's landing page and update the handle accordingly. We find a Twitter handle for 87% of the publishers in the sample.

We assign a slant score to each domain using the data from Robertson et al. [2018]. These data consists of nearly 19,022 of the most popular domains with an associated slant score. To generate this score, Robertson et al. [2018] collect recent tweets containing urls from known Democrats and Republicans. The slant measure is then calculated as the difference in the probability of sharing a domain conditioned on being republican who has shared at least one domain less the same conditional probability for democrats, normalized to be between -1 and 1. Formally, the slant measure is define as

$$\text{bias-score}(i) = \frac{\frac{r_i}{\sum_{j \in I} r_j} - \frac{d_i}{\sum_{j \in I} d_j}}{\frac{r_i}{\sum_{j \in I} r_j} + \frac{d_i}{\sum_{j \in I} d_j}}$$

where r_j (d_j) is the number of unique Republicans (Democrats) who shared domain j and I is the set of all domains. This measure is equal to 0 if shared by equal shares of Republicans and Democrats and equal to 1 (-1) if it was shared only by Republicans (Democrats). Robertson et al. [2018] shows this measure of domain slant agrees with several existing measures of publisher slant [Bakshy et al., 2015, Budak et al., 2016].

After filtering our list of publishers to exclude those who do not appear in the Robertson et al. [2018] data, we are left with 1,170 US hard-news outlets on Twitter.

2.2 News diet summary

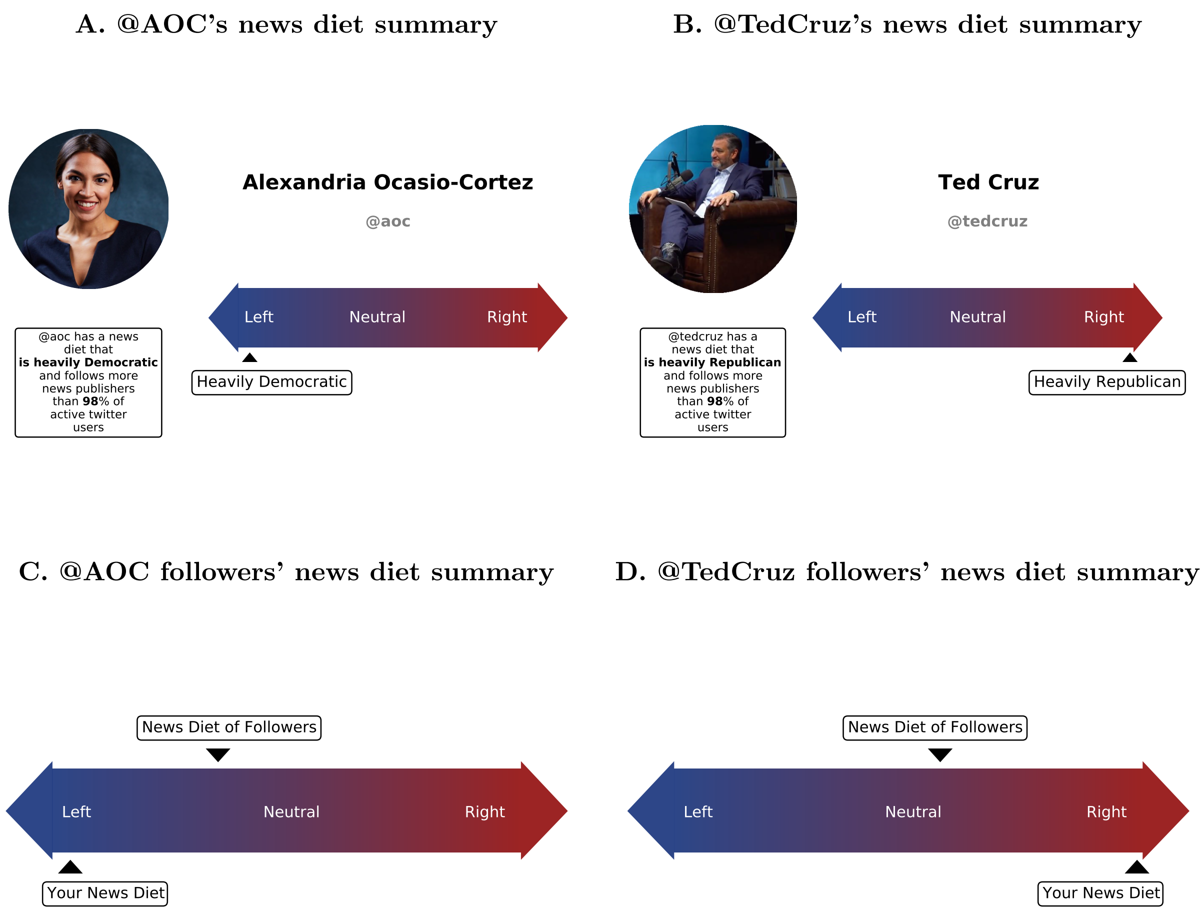
We define the slant of a given Twitter account as the average slant of the publishers that the account follows at the time of the experiment. We also calculate the number of publishers a user follows and report a measure of the user's engagement with publishers relative to a random sample of active Twitter users. Using these two pieces of information we create an infographic. Panels A and B of Figure 1 show two examples of known politicians.

We also estimate the average slant of the peers of a given Twitter account by taking the average slant of a random sample of an acocunts' peers. We compute this average based on a random sample instead of the full sample of users due to Twitter API restrictions. In particular, we estimate the slant of the peers of a given account as the slant of a random sample of five Twitter peers. Panels C and D of Figure 1 show the news diet summary of the peers of the same politicians.

3 Survey instrument

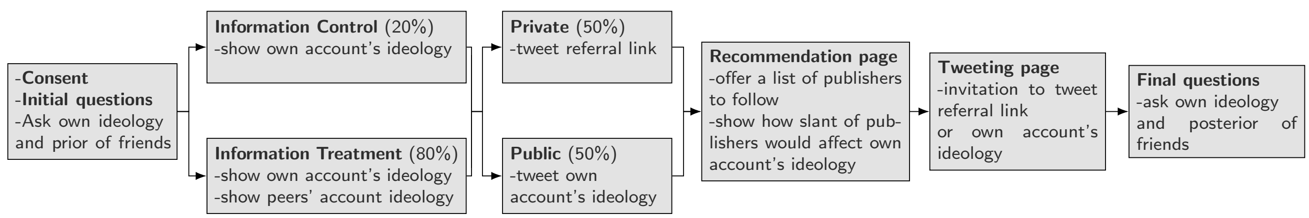
We recruit American adults through Twitter advertisements and invited them to participate in our experiment.

Figure 1: News diet summary



Subjects that click our ads are redirected to a website that contains our experimental design. Figure 2 shows the general structure that participants encounter during the experiment. In the first page we describe our experiment in detail and ask participants for their Twitter accounts if they consent to participate. Participants are then asked some baseline demographic questions as well as a set of questions aiming to gauge participants' interest in politics, political knowledge, and priors about their own ideology and the ideology of their peers.

Figure 2: Timeline, experimental design



Users are then randomized into two different treatments: an information control/treatment group and a sharing control/treatment group. In the information arm, all participants are shown a news diet summary of their Twitter account. This summary reveals real-time information about the average political ideology of the hard-news publishers followed by the subject's account (see Section 2). The only difference between the information control and the information treatment group is that the latter receive information about the news diet summary of their peers (followers). This news diet summary of the peers is computed based on a random sample of their peers.[3]

Participants are subsequently randomized into two groups: the control sharing group is informed that at the end of the experiment, they will be asked to share with their peers–by tweeting–a referral link to promote our survey. Importantly, individuals are made aware that no personal information will be disclosed in this tweet. Users in the sharing treatment group are informed that at the end of the experiment, they will be asked to share with their peers their own news diet summary. Importantly, users are informed that they will have the chance to alter their news diet summary before sharing this information with their friends. All users are informed that sharing this information is optional for both groups and they will receive a chance at a reward upon completion.

After the randomization takes place, all participants are redirected to the recommendation page where they are given a chance to change the publishers they follow: we provide information on at least three conservative and three liberal hard-news publishers and explain the impact that following any of these publishers would have on their political slant. Subjects are allowed to follow any or none of the recommended publishers in addition to following (or unfollowing) publishers that are not listed.

On the next page (Tweeting page), users are presented with an updated version of their news diet summary. On this page users decide whether to tweet their news diet summary or the referral link.

[3]This calculation is based on a random sample of peers rather than all peers because, as all our information is scraped in real-time, we have to adhere to Twitter API restrictions.

Finally, users are invited to complete an endline survey that aims to understand whether users update their beliefs on the slant of their own news diet as well as the news diet of their peers.

In summary, our experiment exploits two sources of variation. First, whether participants learn information about the news diets of their peers. Second, the visibility of the participants' own news diet. In particular, participants in the sharing treatment group face the same experimental structure as the sharing control group with the sole exception that additional information is provided before they are offered the chance to modify their news diet: they are incentivized to share–via a tweet–the updated version of their summary statistic. The comparison of these two groups allows us to observe individuals' differential behavior due to changes in their expectations of how visible their news diet is to their friends.

4 Econometric Models

4.1 Hypothesis

Our goal is to test the relevance of the following two hypotheses.

1. Disclosing: Individuals purposefully alter their news diet when their peers observe their news consumption choices.

2. Learning: When deciding what news to consume, individuals incorporate information they learn about their peers' preferences for slanted news.

4.2 Reduced-Form

Our main econometric model is a Reduced-Form ITT approach. It analyzes the effect of our treatments on a range of outcomes discussed below.

To improve power, we will include controls based on variables collected in the first part of the survey (e.g. demographics). We will use the double machine learning approach from Chernozhukov et al. [2018a] (algorithm DML1) to select control variables. In all regressions, robust standard errors will be reported.

4.2.1 Disclosing

To test whether individuals purposefully alter their news diet when observed by their peers, we exploit the random variation created by the assignment to the treatment sharing (as opposed to the control sharing) group and estimate its effects on (1) whether subjects share their news diet summary (first stage), (2) whether subjects alter their news diet, and (3) longer-term changes of the news diet in the weeks after the intervention. See Section 5 for more detail on the outcomes.

4.2.2 Learning

To test whether individuals incorporate information they learn about the news their peers consume when deciding their own news diet, we exploit random variation in the information we provide to participants about their friends. We operationalize this under two different methods.

In our main specification, we estimate the effect of offering imprecise information about the slant of peers (due to Twitter's API restrictions described in Section 2). In particular, we exploit the variation induced by our noisy estimate of the average peer's slant. After the experiment ends, we compute a more accurate measure of the average peer's slant which is based on a larger number of friends.[4] We exploit the difference between the noisier measure that individuals observe in the experiment and the more accurate measure that we calculate after the experiment as a random source of variation to instrument subjects' beliefs about the ideological position of their peers. Again, focused on the reduced form, we estimate the effect of this instrument on (1) change in beliefs about the position of their peers (first stage), (2) whether subjects make changes to their news diets, and (3) longer-term effects of the changes in the news diet in the weeks after the intervention.

As a complementary method, we will also test this hypothesis by exploiting the variation induced by the information treatment vs the information control group. That is, we will compare individuals that were informed about the slant of their peers versus those do not. While we will report the average treatment effect of this assignment, we see this effect as secondary outcome, as we expect to systematically find smaller effects among individuals that have accurate initial beliefs about their peers (this is something we will test) given the information is less relevant in this case. Therefore, our main analysis when implementing this method is to estimate heterogeneous treatment effects by the level of accuracy of the initial beliefs about the peers. We will focus on the same three sets of outcomes defined above: (1) change in beliefs about the position of their peers (first stage), (2) whether subjects alter their news diet, and (3) longer-term changes in the news diet in the weeks after the intervention.

4.3 Two-stage least squares

4.3.1 Disclosing

Using an indicator equal to one if the participant shares the news diet summary as the endogenous variable and an indicator equal to one if the participant is assigned to the sharing treatment group as the instrumental variable, we will provide 2SLS estimates of the effect of disclosing public information on the slant of news consumption (same set of outcomes as in our reduced-form approach).

4.3.2 Learning

We will also report two-stage least squares estimates using changes between posterior and priors beliefs as the endogenous variable in either of the two methods, thus estimating the impact of changing individuals' beliefs about their peers on the slant of news consumption (again, the same set of outcomes as in our reduced-form approach).

[4]The average slant of the peers that subjects observe in our setting is calculated from a random sample of five followers. The more accurate version of this slant is computed based on 20 rather than 5 followers. Our instrument is defined as the difference between the latter and the former. If feasible given current Twitter API uncertainty, we will sample more than 20 followers as the more accurate version.

4.4 Mechanisms and heterogeneous effects

Under the assumption that subjects purposefully make changes to their news consumption when they either perceive their friends would learn about the news they consume (disclosing) or learn information about the political bias of the news their peers consume (learning), we also aim to understand the mechanisms explaining these changes. In other to do so, we plan to explore the following analysis:

- **directions towards the center/peers:** the two primary mechanisms that we will explore are whether individuals either make changes in their ideological consumption of news relative to (i) the ideological position of their peers (this is, whether peers resemble to fit their friends or digress to impress them) and (ii) the center (individuals care about keeping a balanced or extreme diet).
- **disentangle the cost of social image concerns:** in our experiment, individuals will be incentivized to participate in a lottery if they complete the task of retweeting either the referral link (sharing control) or the updated news diet summary (sharing treatment). we will randomize the amount of money that individuals will receive in this lottery in order to estimate their willingness to accept in exchange for these tasks. By estimating the differential effect across the treatment and the control group, we can separate two different social image costs: the cost of sending information to others (that both, the sharing control and sharing treatment groups face) vs the cost of sending private information to others (that only the sharing treatment group face). This analysis will be considered secondary.
- **heterogeneous effects:** we will follow recent machine learning approaches (e.g. Chernozhukov et al. [2018b], Athey et al. [2019]) to test for heterogeneous effects. When heterogeneous effects are detected on an outcome we will use the method to determine the characteristics of subjects that are the most and the least affected by the treatment. This specification will be considered secondary.

4.5 Correlation Analysis

We also plan to show descriptive statistics and correlation analysis between (i) prior self-reported beliefs on ideology and the slant of the Twitter accounts, (ii) raw outcomes, and main endogenous variables (willingness to share the summary statistic and changes in beliefs), (iii) the noisier and more accurate measure of the slant of the peers. This analysis will be considered secondary.

4.6 Additional analysis

We will consider the possibility of implementing two complementary exercises.

- **follow up-analysis:** we will implement a follow-up survey at least four weeks after the main intervention. In this survey we will inquire about the subject's beliefs on political polarization, political knowledge, and political beliefs that can change as a function of a different exposure on news after our main intervention. The outcomes of this survey would be based on the questions in Levy [2021] and Allcott et al. [2020]. We see this analysis as complementary to the longer-term changes in the slant of the news that each user shares on Twitter in the weeks after the intervention. However, we will only conduct this analysis in the paper if the take up rate of this survey is above 40%.

- **model of user behavior:** we will also consider the possibility of estimating a model of user behavior to understand the relative importance of the direct and indirect channels of peer influence. In the simplest case, one could think a user as having preferences given by:

$$u(s_i; x_i, x_i^s, P_i) = -(s_i - x^*(x_i, x_i^s, P_i))^2 \tag{1}$$

where s_i is user i's slant, x_i is individual i's preferred slant absent any information about their peers, x_i^s is user i's posterior beliefs on the average slant of their peers, and P_i is user i's decision to make their slant public. In solving this model, the optimal slant satisfies:

$$s_i^* = x^*(x_i, x_i^s, P_i) \tag{2}$$

where one could parameterize $x^* = x_i + (\alpha + \gamma P_i)x_i^s$. We take this model to the data by estimating the linear regression

$$s_i^* = \beta_0 + \beta_1 x_i^s + \beta_2 P_i x_i^s + \varepsilon_i \tag{3}$$

where we instrument for x_i^s using the sampling noise contained in the peer's average z_i and instrument for P_i using the public randomization.

We plan to evaluate the possibility of estimating a more complete model (this is work in progress) where we can parametrically estimate the mechanisms discussed in Section 4.4 about digressing to impress, resemble to fit as well as preferences for extreme/centrist positions.

5 Main outcomes

5.1 First stage

Our primary outcomes for the first stage regressions are:

- A dummy equal to one if the participant shares the news diet summary (disclosing).
- Difference between the posterior and the prior beliefs about the ideology of the peers (learning). This variables are asked respectively on the last and first page of the main survey (see Figure 2).

5.2 Outcomes during the main intervention

Our main outcomes of interest aim to understand the changes that subject make to their news diet summary during the experiment (before the retweeting page); see section 3. In particular, we provide participants with information on six different publishers and they are allowed to follow either none, any, or all recommendations. Our first primary outcome measures the extensive margin of whether individuals make any change to the publishers they follow. As a complementary outcome, we also look at the intensive margin (how much individuals change).

Our other set of primary outcomes is the direction of such change (whether individuals move towards/against the right, their peers, the center). When more than one publisher is followed, we based the construction of our outcomes on the average slant of the followed publishers.

5.3 Outcomes post-experiment

We plan to use the Twitter API to monitor user activity in the account of the participants after our experiment takes place. This would allow us to understand whether potential changes made during our experiment either dissipate or exacerbate over time. The probability of the effects dissipating over time ultimately would depend on (i) the probability of an user maintaining the changes done during the experiment and (ii) the sensibility of the user to the content in her news feed (which is affected by the publishers that she might have followed). To test (i), our main outcome of interest is a dummy equals one if the user keep following the publishers k weeks after the intervention (initially, we would consider $k \leq 4$ and extend the window if persistence is high enough by the end of week four). To test (ii), we plan to look at change on Twitter behavior by the user (conditional on the treatment). This includes, the slant of the likes, retweets, mentions and follows of material that the user engage with k weeks after the intervention (where k is defined above).

Depending on the take-up rate of the follow-up survey (see Section 4.6), we will also explore all the complementary outcomes on political knowledge, political polarization and political beliefs in this survey.

References

H. Allcott, L. Braghieri, S. Eichmeyer, and M. Gentzkow. The welfare effects of social media. *American Economic Review*, 110(3):629–76, 2020.

S. Athey, J. Tibshirani, and S. Wager. Generalized random forests. 2019.

E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132, June 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaa1160. URL `https://science-sciencemag-org.libproxy.mit.edu/content/348/6239/1130`. Publisher: American Association for the Advancement of Science Section: Report.

L. Boxell, M. Gentzkow, and J. M. Shapiro. Cross-country trends in affective polarization. Technical report, National Bureau of Economic Research, 2020.

C. Budak, S. Goel, and J. M. Rao. Fair and balanced? quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1):250–271, 2016.

V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters, 2018a.

V. Chernozhukov, M. Demirer, E. Duflo, and I. Fernandez-Val. Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. Technical report, National Bureau of Economic Research, 2018b.

Y. Lelkes, G. Sood, and S. Iyengar. The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science*, 61(1):5–20, 2017.

R. Levy. Social media, news consumption, and polarization: Evidence from a field experiment. *American economic review*, 111(3):831–70, 2021.

R. D. Putnam et al. *Bowling alone: The collapse and revival of American community*. Simon and Schuster, 2000.

R. E. Robertson, S. Jiang, K. Joseph, L. Friedland, D. Lazer, and C. Wilson. Auditing partisan audience bias within google search. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–22, 2018.

C. R. Sunstein. *# Republic: Divided democracy in the age of social media*. Princeton University Press, 2018.