

# “Social Media and Political Persuasion: A Field Experiment”

## Pre-analysis Plan

Ro’ee Levy\*

The main objective of the study is to understand the effects of news consumed through social media on political opinions and political behavior. The effects will be measured using a field experiment manipulating news content in subjects’ social media feeds. The following document details the pre-analysis plan for the experiment. The plan was submitted before the baseline survey took place and a revised version will be published between the baseline and follow-up surveys.

The study aims to answer two main research questions:

- **Does social media news consumption affect political opinions and behavior?**

According to a 2014 Pew Research Center survey “*Among Millennials, Facebook is far and away the most common source for news about government and politics*” (Pew, 2014). The consumption patterns of young Americans indicate a general change in news consumption habits as the share of Facebook users getting news through the site increased from 47% in 2013 to 67% in 2017 (Gottfried and Shearer, 2017). Furthermore, news consumed on social media may be different from traditional news sources and there is growing concern that the filter bubbles and echo chambers characterizing social media are polarizing the electorate.<sup>1</sup> I will measure whether and when people are persuaded by news they consume online and whether increased exposure to partisan news which matches the ideology of consumers, affects polarization.

- **How do consumers establish political opinions and what behavioral biases affect the learning process?**

There have been various experiments attempting to identify how people form beliefs based on information they receive, however the majority of these experiments take place in a lab.<sup>2</sup> While a lab provides a clean setting to investigate mechanisms, it also suffers from known limitations: usually short term effects are studied, the setting is often not natural, the subjects may be a captive audience and Hawthorne effects may bias the results. I propose exploiting social media’s infrastructure to naturally and gradually distribute news to subjects to understand how they process information to form political opinions.

## 1 Background

The study will compare theories explaining how individuals learn when consuming news from an outlet with an ideological slant. In each model consumers have limited information and a prior  $\theta^{C_0} \sim N(\theta^C, \frac{1}{h^C})$  on the state of

---

\*Yale University. This project is supported by the Institute for Social and Policy Studies Field Experiment Initiative, by the Yale Economics Department and the Yale Program in Applied Economics and Policy.

<sup>1</sup>For example, in President Obama’s farewell address he mentioned the “*splintering of our media into a channel for every taste*” as a threat to democracy.

<sup>2</sup>Examples include: Arceneaux et al. (2012); Levendusky (2013); Guess and Coppock (2016); Nyhan and Reifler (2010).

the world, e.g. how is crime affected by immigration, how big of a threat is climate change, what is the quality of a candidate running for office. Media outlets receive a signal  $\theta^F = \theta + \varepsilon^F$  on the true state of the world,  $\theta$ , with the error term distributed normally around zero  $\varepsilon^F \sim N(0, \frac{1}{h_F})$ . Outlet  $j$  reports the state of the world with an ideological slant:  $r_j^F = \underbrace{\theta^F}_{\text{signal}} + \underbrace{s_j^F}_{\text{slant}}$ . Due to the slant, a conservative outlet will have a more positive report on a

conservative candidate than a liberal outlet. The ideological slant of outlets can be explained by owner incentives (Anderson and McLaren, 2012), or by an attempt to maximize market share (Gentzkow and Shapiro, 2006).

I will compare three demand-side models that explain how consumers interpret the report,  $r^C(r^F)$ , and formulate a posterior.

- **Rational and informed consumers** - If consumers are rational and have information on the slant of the outlet, they will adjust any signal they receive from a media outlet according to the slant of the media outlet:  $r^C(r^F) = r^F - s_j^F = \theta^F$ . The consumer's posterior is the weighted average of her signal and the adjusted report:  $N(\frac{h^C\theta^C + h^F\theta^F}{h^C + h^F}, \frac{1}{h^C + h^F})$ . According to this model, the slant of an outlet has no effect since the consumer takes the slant into account and only learns from the signal the outlet actually observed. Therefore, it does not matter if households are exposed to more conservative or liberal outlets (assuming both outlets observe the same signal and that the only difference between the outlets is their slant).
- **Credulous Bayesians** (Glaeser and Sunstein, 2009) - Consumers may not be fully informed of the slant of the media outlets or fail to account for the slant. More generally, consumers may fail to account for correlated signals and repetition of information (DeMarzo et al., 2003). One simple way to model this behavioral failure is to assume consumers suffer from persuasion bias (DellaVigna and Kaplan, 2007).<sup>3</sup> In this model consumers individuals only partially adjust for the outlet's slant and infer that the signal received by the outlet is  $r^C(r^F) = r^F - (1 - \lambda)s_j^F$  where  $\lambda \in [0, 1]$  represents the persuasion bias. When  $\lambda = 0$ , consumers are fully aware of the outlet's slant and act according to the rational informed consumers model, and when  $\lambda = 1$ , they are completely unaware of the slant and treat each report of an outlet as being sampled from a distribution around the true state of the world. This model predicts that outlets with different slants lead to diverging opinions. For example, Fox News readers may become more conservative than *New York Times* readers.
- **Confirmation Bias** - A third option is that consumers suffer from confirmation bias and they uncritically accept news from outlets with a slant similar to their ideology and challenge news from opposing outlets (Taber and Lodge, 2006). One explanation for this behavior is that consumers' utility depends not only on the accuracy of the report they receive but also on reaching their desired conclusion or a position that strengthens their current opinion (Kunda, 1990; Mullainathan and Shleifer, 2005). If consumers suffer from persuasion bias, but they can partially overcome this bias with some cognitive cost,<sup>4</sup> they will trade-off the effort in overcoming the persuasion bias with the cost of changing their opinion.<sup>5</sup>

Consequently, consumers are more likely to be persuaded by signals that are closer to their current opinion.

One way to model this behavior is:  $r^C(r^F) = \begin{cases} r^F & |r^F - \theta^C| \leq \tau \\ \theta^C & |r^F - \theta^C| > \tau \end{cases}$  In this interpretation, consumers ignore news if the distance between their opinion and the signal they receive is above a certain threshold,  $\tau$  which

<sup>3</sup>Alternatively this type of behavior could be explained by lack of information or limited attention (Eyster and Rabin, 2010).

<sup>4</sup>The cognitive cost can be explained if disbelieving is a System 2 operation while believing is a system 1 operation (Gilbert, 1991).

<sup>5</sup>Confirmation bias can also be explained by the fact that people more easily and naturally recall memories of evidence supporting a current hypothesis, instead of evidence refuting it so perhaps even the costs of overcoming the persuasion bias are higher when consuming news from cross-cutting media outlets.

depends on the costs and benefits from changing one's opinion substantially, and naively believe news when the signal is close to their current opinion.

## 2 Design Overview

Subjects will be recruited to the study using Facebook ads. Individuals who decide to begin the survey will login using their Facebook account and complete a baseline survey. In order to randomize news exposure, towards the end of the survey subjects will be given an option to "like" four Facebook news pages. "Liking" a page on Facebook is similar to subscribing to a (free) newspaper, and once a page is "liked" its content may start appearing in the subjects' feed (the words "like" and "subscribe" will be used interchangeably throughout the rest of the document). Subjects will be assigned to a conservative, liberal or control treatment. In the conservative treatment four pages of conservative-leaning media outlets will be offered (e.g. Fox News); and in the liberal treatment the four pages offered will be of liberal-leaning media outlets (e.g. the *New York Times*). In the control treatment, subjects will not be offered any pages. This is an encouragement design and in each treatment, subjects will choose whether to "like" each of the pages.

Approximately 6-8 weeks after the baseline survey, subjects will be invited to a follow-up survey to measure any self-reported changes in their opinions and political knowledge.

### 2.1 Sample

Facebook users living in the US who are over 18 years old will be recruited to the study using Facebook ads. Two main ad sets will be used: one set of ads will emphasize the survey and gift card lottery, while the second set of ads will also mention that the survey may be of interest to people who follow politics. Several ads will be used to ensure that there is a large sample of subjects completing the survey on desktop and mobile. After clicking an ad, potential subjects will be referred to the online survey. Only individuals who login to Facebook, provide permission to access the list of pages they "like" and reach the final section of the survey, will be included in the sample.

### 2.2 Randomization

Each participant will be assigned to a control, conservative or liberal treatment, based on a randomized block design by participants' baseline ideology. Since the sample frame arrives over time, the entire sample will not be stratified in advance, but will be split gradually into groups, according to subjects' ideology.

At the beginning of the survey, respondents will be asked where they position themselves ideologically on a 7-point ideological scale from very liberal to very conservative, with an additional option of "I haven't thought about it much." Subjects will be assigned to a treatment based on where they position themselves on the scale and when they answered the question. Each block (stratum) will be composed of three sequential subjects who chose the exact same answer among the eight options in the ideology scale survey question. The first three subjects who gave the same answer in the ideology question will be randomly split to the conservative, liberal and control groups. The next three individuals with the same ideology will be also be split into the three groups, and the process will continue with each additional respondent. As a result of this process, there will be approximately the same number of liberal, conservatives and moderates in each treatment group. In practice, since subjects arrive gradually, the first subject in each stratum will be randomly assigned to one of the three treatments, the second

subject will be randomly assigned to one of the two remaining treatments and the third subject will be assigned to the final remaining treatment.<sup>6</sup> If there are any technical issues in determining the strata for a particular subject, the subject will be randomly assigned to one of the three treatments.

Responses from the baseline survey will be used to test for balance across treatment assignments. Subjects will be compared across treatments according to their sources of news consumption, number of Facebook “likes”, answers to several questions on ideology (approval of President Trump, party identification, ideology) and standard demographic variables (age, education, gender, income and race). The questions on income, race and education will be asked after the treatment and could theoretically be influenced by the treatment.<sup>7</sup> Therefore, as a robustness check, balanced will also be tested by imputed income and education, based on the subject’s location, which is not affected by the treatment.

## 2.3 Setting

The experiment will take place in the United States in February, 2018. The following pages will be offered:

- Conservative page:
- Liberal pages:

If a subject is already subscribed to one of the liberal or conservative pages being offered in the survey, an alternative page will be offered instead. The following backup pages will be offered in case a subject already likes one of the pages:

- Conservative backup pages: The Blaze, Washington Times, Townhall, The Weekly Standard, The Conservative Tribune
- Liberal backup pages: Washington Post, NPR, The New Yorker, The Atlantic, PBS

The news outlets were chosen according to several criteria. First of all, they have a relatively clear ideological slant. Secondly, preference was given to popular outlets (Fox News and the *New York Times* are the second and third most popular news pages on Facebook). Finally, I attempted to include outlets of varying quality and extremity of opinions in order to allow subjects more variety when choosing which page to “like” and in order to increase the likelihood that subjects engage with at least one of the pages offered.

## 3 Data

The following data sets will be used:

- Baseline survey - The survey will include questions on news consumption habits, political opinions and demographics. Subjects will also be asked what they thought is the ideological slant of various outlets.

---

<sup>6</sup>The main risk with this randomization design is that the treatment for some subjects is pre-determined and this may create selection bias. However, I believe this potential issue is not a cause for concern since subjects do not know the randomization method, even if they knew the method, they would not know what is the expected treatment they will receive at a given time (since participants are being recruited constantly throughout the US), and they do not have an incentive to receive a specific treatment.

<sup>7</sup>While it is possible to include all the questions before the experiment, I believe it makes more practical sense to end the survey with the demographic questions, since some subject may stop completing the survey when presented with these questions. In addition, I prefer that the intervention (being offered to like a Facebook page) will not be the very last thing subjects are asked to do in the survey, in order to decrease the association of the survey with this particular question.

- Follow-up survey - Subjects will be asked about their political opinions on issues and figures which were recently discussed in the news (e.g. tax reform), in addition to questions related to partisan hostility and various news outlets.
- Facebook posts and likes - Subjects will be asked to provide permissions to access Facebook data on the pages they “liked” and their posts. The pages subjects’ “like” on Facebook will be used to measure which pages each subject subscribed to in the experiment and whether and when subjects unsubscribed from pages. Facebook posts will be used to analyze the effect of the treatments on political behavior.
- Browsing and Facebook feed data - Subjects who completed the baseline survey using Google Chrome on a desktop computer will be offered a small reward in exchange for installing a Chrome extension which will provide data on their news-related browsing behavior and Facebook feed. Any estimates using this data will also be analyzed while excluding data from the first couple of days after the extension was installed, since Hawthorne effects are less likely to affect browsing behavior as subjects become less conscious of the extension.

## 4 Empirical Analysis

Three main types of models will be used throughout the empirical analysis. First I will analyze who complies with the treatment and which pages they choose to “like”. The next two models are more important and analyze the effect of the treatment on opinions on behavior. The persuasion model focuses on the persuasive effects of consuming news and will measure whether subjects changed their opinions according to their treatment assignment, while the polarization model will compare subjects who received news matching their baseline ideology with subjects who received news from opposing sources. Below I will detail the main regressions I intend to run, however it is likely that additional analysis will be conducted based on the final data set, in order to explore mechanisms and further test theories.<sup>8</sup>

### 4.1 Selective Exposure

The first set of results will analyze which subjects complied with the treatment and “liked” at least one of the pages offered. A model such as the following will be used:

$Y_i = \beta_1 T_i^L I_i^L + \beta_2 T_i^L I_i^C + \beta_3 T_i^C I_i^L + \beta_4 T_i^C I_i^C + \alpha X_i + \varepsilon_i$ , where:

- $T_i^L \in \{0,1\}$  is whether subject  $i$  was assigned to a liberal treatment, and  $T_i^C \in \{0,1\}$  is whether subject  $i$  was assigned to a conservative treatment.
- $I_i^L \in \{0,1\}$  is whether subject  $i$  has a liberal ideology, and  $I_i^C \in \{0,1\}$  is whether subject  $i$  has a conservative ideology. The ideology of each subject will be determined according to where subjects place themselves on the ideological scale. For moderate subjects or subjects who do not know where to place themselves, the party they identify with will be used to determine their ideology, and if they do not lean towards either party, their ideology will be determined according to the candidate they supported in the 2016 elections. I predict that by using this method more than 90% of individuals will be assigned an ideology.
- $X_i$  are standard demographics: gender, age and squared age. In addition,  $X$  will control for the set of pages offered.

---

<sup>8</sup>Olken (2015) provides a detailed discussion on the advantage and limitations of pre-analysis plan and exploratory analysis in economics.

- $Y_i$  is whether the subject complied with the treatment and “liked” at least one page. As a robustness test I will run an identical regression measuring compliance two months after the initial treatment. The result for this outcome could be different if subjects “unliked” some of the pages they initially subscribed to in the experiment.

The main coefficient of interest is the difference between the effect of being offered a page that matches your ideology and being offered a page with opposing ideology. I will test if  $\beta_1 = \beta_3$  and  $\beta_2 = \beta_4$

To better understand the choice for each specific page the following model can be used:

$Y_{ij} = \beta_1 Know_{ij} + \beta_2 Know_{ij} * (Ideology_i - Slant_{ij}) + \alpha X_i + \varepsilon_{ij}$  where  $i$  is an individual and  $j$  is a page offered to the individual:

- $Know_i = \begin{cases} 0 & \text{The subject does not know the slant of the outlet} \\ 1 & \text{Otherwise} \end{cases}$
- $Slant_i$  is the perceived ideology of outlet  $j$  according to subject  $i$  on a  $[-3, 3]$  scale, where  $-3$  is most liberal and  $+3$  is most conservative
- $Ideology_i$  is the self-reported ideology of the individual on the same scale.

Note that this regression is only meant to provide a simple model for how people choose outlets, but the relationship are not necessarily causal, since it does not rely on the exogenous variation generated by the intervention. This regression can be used with and without page fixed effects.

## 4.2 Persuasion

The following intention-to-treat regressions will be used to measure the average effect of the intervention.

$Y_i = \beta_0^L T_i^L + \beta_0^C T_i^C + \alpha X_i + \varepsilon_i$  where:

- $X_i$  is a set of covariates from the baseline survey. Since the effects of the experiment are expected to be relatively small, and since the baseline coefficients will have high predictive power, it is important to include coefficients to increase precision. The covariates included will be the strata, self-reported ideology, party affiliation and approval of President Trump. Additional covariates will be included for outcomes where a lagged variable exists (for example, in regressions where the outcome is the slant of subjects’ posts, the model will also control for the baseline slants of posts). Using a nested F-Test I will check if a model which include these covariates along with age, age squared and gender is significantly different than a model without the demographics covariates when predicting the outcomes of policy questions in the baseline survey. If demographics add valuable information they will be included in the main analysis as well.
- $Y_i$  is the outcome, where a higher  $Y_i$  indicates a more conservative position. The following outcomes will be used:
  - Opinions - Opinion on topics and figures currently discussed in the news are the primary outcome which will be studied. The treatment in this study will have a relatively small immediate effect on subjects’ news feed. Therefore, I do not expect the subjects’ ideology to change substantially, while their opinions on topical issues are more likely to be influenced by the treatment. The challenge is that it is difficult to anticipate in advance which topics will be the most relevant during the study period.

Therefore, I will finalize the list of follow-up questions according to issues debated in major news outlets during the study period, and those questions will be used to measure the primary outcome. For example, in a previous study conducted in November 2016 through January 2017, subjects were asked about the presidential transition and their opinion on Steve Bannon.

- Behavior - The slant of each article or post shared by an individual will be determined according to a database based on articles shared on Facebook (Bakshy et al., 2015).<sup>9</sup> I will measure whether the treatment had an effect on the difference between the number of conservative and liberal posts. I will also test whether individuals were more likely to share posts specifically from the pages that they were randomly offered. This outcome will measure whether the treatments had any effect on actual political behavior (sharing partisan articles with friends), in addition to the effect on self-reported political opinions. Depending on the project's final budget and the length of the follow-up survey, additional behavioral measures may be used to measure opinions in the follow-up survey, such as asking subject where they would prefer donating a small amount of money.
- $\beta_0^C, \beta_0^L$  are the coefficient of interests, which measure the effect of being offered conservative vs. liberal news pages. For example,  $\beta_0^C > 0$ , will imply that a conservative treatment made subjects more conservative.

All the regressions will use robust standard errors. The main estimations will not be weighted. Ordinal variables based on survey questions will usually be converted first to a binary variable indicating whether the answer is above/below the median before being used as an outcome.

The primary test is whether  $\beta_0^L < \beta_0^C$ , that is whether people randomly offered liberal news become more liberal than people randomly offered conservative news. This will determine if consumers suffer from persuasion bias and tend to change their opinions according to an outlet they are randomly exposed to, or whether they are rational and adjust for the slant of the outlet. Testing whether  $\beta_0^L < 0$  and  $0 < \beta_0^C$  may provide further evidence on which subjects were persuaded by the slant of the outlet, by comparing each treatment arm to the control group. However, these tests require more power and therefore the probability that an effect will be detected is smaller. In addition these tests are problematic in cases where subjects have extreme opinions and as a result even being persuaded by news that matches their opinion, makes them more moderate.

If subjects are indeed persuaded by outlets, the model will also be estimated only for subjects who "liked" pages for which they were able to estimate the slant of the outlet. When subjects are exposed to news from outlets they are not familiar with, they may be affected by the slant of the outlet since they are not aware of the slant. However, if subjects suffer from persuasion bias also when they are familiar with the outlet, the results will suggest that persuasion bias cannot be explained only as a consequence of missing information.

#### 4.2.1 IV Regression

Any result from the intention-to-treat estimate is relevant for a policy offering people exposure to new opinions, but will not estimate the effect of actually subscribing to news outlets. Therefore, I will use the following IV regressions based on two separate measures of compliance:

$$Y_i = \beta_0^L C_i^L + \beta_0^C C_i^C + \alpha X_i + \varepsilon_i \text{ with } T_i^L, T_i^C \text{ as instruments and}$$

$$Y_i = \beta_0^L E_i^L + \beta_0^C E_i^C + \alpha X_i + \varepsilon_i \text{ with } T_i^L, T_i^C \text{ as instruments. Where:}$$

- $C_i^L, C_i^C$  is whether subject  $i$  initially complied with the treatment and "liked" liberal or conservative pages.

<sup>9</sup>A database created by BuzzFeed may be used to identify the slant of additional outlets: <https://github.com/BuzzFeedNews/2017-08-partisan-sites-and-facebook-pages>

- $E_i^L, E_i^C$  is whether subject  $i$  was exposed to the liberal/conservative pages as a result of the treatment. Exposure to news pages will be defined based on self-reported answers in the follow-up survey. For the subset of users who install the chrome extension, exposure will be defined according to actual news feed data. This measure is important, since subjects may “like” a page but not see its content in their feed, either because they “unliked” it, Facebook’s algorithm determines that the content offered by the page would not interest the subjects or the subjects do not engage with the page and its content stops appearing as a result. It is unlikely that a page will have an effect on opinions if it does not appear in the subject’s feed.
- $\beta_0^L, \beta_0^C$  are the LATE estimators, the average effect of conservative/liberal pages among individuals who are willing to subscribe to these pages when they are offered, or among people who are exposed to articles from those pages as a result of the pages being offered.

These regression will estimate local average treatment effects, the average effects for compliers. Therefore, it is important to understand who are the compliers. Since the persuasion in the experiment occur over a roughly two-month period, it is possible that some subjects would have complied with the treatment without being treated (“always-takers”), i.e. some subjects might have “liked” one of the pages following the experiment and would have been exposed to its content, even if the would not have been offered. Thus, it is necessary to disentangle the characteristics of compliers from always-takers. The treatment groups will be used to identify the proportion of never-takers, and data from the control group will be used identify the share of always-takers in order to formally identify the characteristics of compliers (Abadie, 2003).

#### 4.2.2 Confirmation Bias

One of the goals of the study is to determine how respondents are affected by news sources with opposing ideological position: are they persuaded, do they ignore content from opposing outlets or are their original positions reinforced (a “boomerang effect”). Therefore, the most important heterogeneous analysis will be based on the following regression model:

$$Y_i = \beta_1 T_i^L I_i^L + \beta_2 T_i^L I_i^C + \beta_3 T_i^C I_i^L + \beta_4 T_i^C I_i^C + I_i + \alpha X_i + \varepsilon_i$$

Only subject whose ideology is identified ( $I^L = 1$  or  $I^C = 1$ ) will be included in this regression.

- The primary test for motivated reasoning is whether  $\beta_2 = \beta_3 = 0$ . This will test if individuals are affected by opposing news, i.e. are liberals persuaded by the conservative treatment and conservatives persuaded by the liberal treatment. In addition, I will test if  $\beta_1 = \beta_2$  and  $\beta_3 = \beta_4$  to determine if conservatives and liberals were affected similarly by the treatments.<sup>10</sup>

#### 4.2.3 Additional heterogeneity analysis

Additional heterogeneous effects will be tested, such as the following:

- Exposure - The purpose of the first heterogeneous analysis is to validate any effects detected in the experiment. I expect subject who were exposed to more content from the pages they liked to be affected by the treatment.

<sup>10</sup>If there is a small difference between the effect of matching news and non-matching news, measuring the interaction effects may require more power and the estimates will be imprecise. One possibility to increase the precision of these estimates is to assume that the conservative and liberal treatments have similar effects (in opposite direction) and set  $T_i^L = -T_i^C$



- Subjects' openness - The purpose of this analysis is to test whether people who are open to new political opinions are more likely to be persuaded. Openness will be defined according to two questions taken from the Ten Item Personality Inventory (Gosling et al., 2003). I may also be able to predict openness based on pages the subjects subscribed to before the experiment (Kosinski et al., 2013). Similarly, I will test if people who have changed their minds in the past according to something that appeared in their feed and who are less certain in their opinions are more affected by the treatment.
- Echo chamber - I will test for heterogeneous effects according to whether subjects are more likely to be exposed to similar opinions in their news feed (whether their news feed resembles an echo chamber). This effect will be measured according to a question in the baseline survey and based on Chrome extension data on actual posts appearing in the subjects' news feed.
- Sophistication - Two questions in the baseline survey will be based on the cognitive reflection test (CRT) which attempts to distinguish between two types of cognitive processes: intuitive, quick processes, compared to slower and reflective processes (Shane, 2005). We would expect less sophisticated subjects to suffer from persuasion bias, since they may think less critically on the outlet's slant. On the contrary, more sophisticated subjects are likely to fit the rational consumer model, or even act according to motivated reasoning theory, since they can more easily challenge news they don't agree with (Pennycook and Rand, 2017).

All the analysis of heterogeneous effect will control for the pages the subjects were offered to prevent an omitted variable bias since individuals who are offered backup pages have different characteristics than individuals who were offered the primary pages, and they were exposed to different news as a result of the treatment.

### 4.3 Polarization

Models like the following will be used to measure the effect of variation in news exposure on polarization:

$$Y_i = \beta_4^P P_i + \beta_4^A A_i + \alpha X_i + \varepsilon_i$$

- $P_i$  measures whether the participant was randomly offered pro-attitudinal outlets:

$$P_i = \begin{cases} 1 & (T_i^L * I_i^L = 1) \text{ OR } (T_i^C * I_i^C = 1) \\ 0 & \text{otherwise} \end{cases}$$

- $A_i$  measures whether the participant was randomly offered counter-attitudinal (anti-attitudinal) outlets:

$$A_i = \begin{cases} 1 & (T_i^L * I_i^C = 1) \text{ OR } (T_i^C * I_i^L = 1) \\ 0 & \text{otherwise} \end{cases}$$

The model is based on the same randomization used in the previous model, but the focus is not on whether the treatment is conservative or liberal but on whether it matched or opposed the subject's baseline opinion.

- $Y_i$  is the outcome, where a higher  $Y$  indicates greater polarization. Three main outcomes will be used in this model:
  - Absolute opinions (attitudinal-polarization): absolute opinions will be measured according to a similar set of variables measuring the persuasion effect, but the outcomes will be determined by the strength of opinions instead of their position on a left-right scale. For example, I will test whether the *absolute* slant of posts increased when subjects were assigned to receive matching news pages, compared to opposing news pages. In other words, this variable will measure whether people become more extreme when being exposed to information matching their ideology.

- Partisan hostility (affective-polarization): partisan hostility will be measured according to the difference between the feelings of subjects towards the party they identify with and their feelings towards the opposing party, based on a feeling thermometer question, and according to how subjects state they would feel if their son/daughter married an individual identifying with the opposing party. Affective polarization will also be measured based on political empathy, which will be defined according to an index of two-three survey questions (Reit et al., 2017).
- Perceived polarization: Subject will be asked how liberal and conservative a typical Democrat/Republican is, the difference will be defined as perceived polarization. I do not have a strong prior on how matching vs opposing outlets will affect perceived polarization.

Depending on the project's final budget and the length of the follow-up survey, additional behavioral measures may be used to measure opinions in the follow-up survey, such as having subjects play a dictator game.

- Coefficients of interest

- $\beta_4, \beta_4^P$  measure whether exposure to opposing news mitigates polarization as compared to matching news. For example,  $\beta_4^P > 0$  implies that exposure to more pro-attitudinal news increases polarization. The most important test is whether  $\beta_4^A < \beta_4^P$ : do subjects become more polarized when exposed to matching news compared to opposing news. As usual, it is possible to compare the effect of matching and opposing news to the control group by testing whether  $\beta_4^A < 0, 0 < \beta_4^P$ .

#### 4.3.1 Knowledge

While this study focuses on persuasion and polarization, I will also take advantage of the regression above to test whether exposure to random news increases political knowledge, compared to the control treatment.<sup>11</sup> More interestingly, it is possible that exposure to opposing news will provide subjects with greater knowledge on specific issues that are mostly debated by one side of the aisle. Political knowledge will be measured with the following outcomes:

- Knowledge on events - Questions on events related to the news cycle. Both general news stories and stories that were mostly covered in liberal and conservative outlets will be included.
- Accurate perceptions - Subjects will be asked how common they think certain positions are. For example, how many people they think have a favorable opinion on President Trump.

I will test whether subjects exposed to liberal/conservative news believe the electorate is more liberal/conservative and whether exposure to more opposing news, leads to more accurate perceptions and greater knowledge.

## 5 Mechanisms

Data from the subset of individuals who installed the chrome extension will be used to analyze how news consumption changed as a result of the treatment. The analysis will include several steps:

---

<sup>11</sup>In this specification, I may pool the matching and non-matching treatment, and compare the effect of being offered any news vs the control treatment.

1. Immediate effect on exposure - The first test is whether posts from the pages offered to a subject appeared more often in the subjects' Facebook feeds, immediately following the experiment and a month later, compared to the control group and compared to posts which were not offered due to the treatment assignment (for example for subjects in the conservative treatment group, I will check if the four conservative pages appeared significantly more often than the four liberal pages).
2. Immediate effect on consumption - Real-time and historic news consumption data will be used to test whether individuals were more likely to consume news through Facebook, from pages they were offered in the treatment, compared to the control group while controlling for baseline news consumption. Similarly, the effect of exposure on consumption will be estimated, using pages offered as the instrument.
3. Effect on consumption through other means - If individuals start consuming news from the Facebook pages which were offered to them, they may start consuming articles from the pages' corresponding websites regardless of whether the article appears in their Facebook feed. Alternatively, it is possible that the news consumed through Facebook will replace the consumption of the same content through other means. Therefore, the same analysis described in the previous step will be conducted on news which was not accessed through Facebook.
4. Indirect effects on other news - If indeed the exposure and consumption of the assigned news content increases, it is important to understand how exposure and consumption of other content is affected. It is possible that the exposure to other news content will not change (e.g. news offered in the experiment crowds out exposure of entertainment sources instead of other news), it may increase (if news articles are complements) or decrease (if news articles are substitutes). Furthermore, it is important to understand whether the slant of other news consumed was affected by the experiment. For example, if an individual was randomly offered to like Fox News in the treatment and started consuming content from that outlet instead of other conservative outlets (such as the *Wall Street Journal*) we would expect a different effect than in a case where Fox News crowded-out liberal content (such as the *New York Times*). I will analyze if the total number of conservative and liberal outlets consumers were exposed to and consumed changed as a result of the treatment. Exposure will be analyzed using posts appearing in the subjects' feeds and consumption will be based on news consumed through Facebook.

## 6 Implications

New technology generally, and social media specifically, have enabled individuals to consume more news from media outlets that match their ideology. This study can show whether exposure to outlets with clear ideological slants polarizes readers, and which policies may decrease polarization. If individuals act as Credulous Bayesians and are persuaded by the outlets they read ( $\beta_0^L < \beta_0^C$ ), then social media, which is associated with increased segregation in news consumption, is expected to increase polarization. However, perhaps polarization could be mitigated by designing mechanisms that expose individuals to more cross-cutting news (if  $\beta_4^A < 0$ ). Such policies have already been suggested (Sunstein, 2017), but they should first be tested empirically. This experiment provides a nudge that encourages people to receive random news and thus tests how serendipity encounters affect consumers.

## References

- Abadie, A. (2003). Semiparametric Instrumental Variable Estimation of Treatment Response Models. *Journal of Econometrics* 113(2), 231–263.
- Anderson, S. P. and J. McLaren (2012). Media Mergers and Media Bias with Rational Consumers. *Journal of the European Economic Association* 10(4), 831–859.
- Arceneaux, K., M. Johnson, and C. Murphy (2012). Polarized Political Communication, Oppositional Media Hostility, and Selective Exposure. *The Journal of Politics* 74(01), 174–186.
- Bakshy, E., S. Messing, and L. A. Adamic (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science* 348(6239), 1130–1132.
- DellaVigna, S. and E. Kaplan (2007). The Fox News Effect: Media Bias and Voting. *The Quarterly Journal of Economics* 122(3), 1187–1234.
- DeMarzo, P. M., D. Vayanos, and J. Zwiebel (2003). Persuasion Bias, Social Influence, and Unidimensional Opinions. *The Quarterly Journal of Economics* 118(3), 909–968.
- Eyster, E. and M. Rabin (2010). Naïve Herding in Rich-Information Settings. *American Economic Journal: Microeconomics* 2, 221–243.
- Gentzkow, M. and J. M. Shapiro (2006). Media Bias and Reputation. *Journal of Political Economy* 114(2), 280–316.
- Gilbert, D. T. (1991). How mental systems believe. *American Psychologist* 46(2), 107–119.
- Glaeser, E. L. and C. R. Sunstein (2009). Extremism and Social Learning. *Journal of Legal Analysis* 1(1), 263–324.
- Gosling, S. D., P. J. Rentfrow, and W. B. Swann (2003). A Very Brief Measure of the Big-Five Personality Domains. *Journal of Research in Personality* 37(6), 504–528.
- Gottfried, J. and E. Shearer (2017). News Use Across Social Media Platforms 2017. *Pew Research Center*.
- Guess, A. and A. Coppock (2016). The Exception, Not the Rule? The Rarely Polarizing Effect of Challenging Information. *Working Paper*.
- Kosinski, M., D. Stillwell, and T. Graepel (2013). Private Traits and Attributes are Predictable from Digital Records of Human Behavior. *Proceedings of the National Academy of Sciences* 110(15), 5802–5805.
- Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin* 108(3), 480–498.
- Levendusky, M. S. (2013). Why do Partisan Media Polarize Viewers? *American Journal of Political Science* 57(3), 611–623.
- Mullainathan, S. and A. Shleifer (2005). The Market for News. *American Economic Review* (February), 1031–1053.
- Nyhan, B. and J. Reifler (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, 32(2), 303–330.
- Olken, B. A. (2015). Promises and Perils of Pre-Analysis Plans. *Journal of Economic Perspectives* 29(3), 61–80.
- Pennycook, G. and D. G. Rand (2017). Who Falls for Fake News? The Roles of Analytic Thinking Motivated Reasoning, Political Ideology, and Bullshit Receptivity. *Working Paper*.

- Pew (2014). *Political Polarization and Media Habits*. Pew Research Center.
- Reit, E., R. Willer, and J. Zaki (2017). Causes and Consequences of Political Empathy. *Work in Progress*, Stanford University.
- Shane, F. (2005). American Economic Association. *The Journal of Economic Perspectives* 19(4), 25–42.
- Sunstein, C. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.
- Taber, C. S. and M. Lodge (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science* 50(3), 755–769.