Is It All Just Politics?

A Bayesian Analysis of Public Opinion On Regulating Social Media Companies

George Perrett

Contents

Introduction

Calls to increase government regulation of social media tech companies have been elevated in the national political dialogue. Social media companies have an inherent interest in limiting increased regulations which can disrupt many aspects of business models and growth strategies.

Public opinion is an important motivator of politicians actions, and the public's views towards social media companies and the extent to which they are regulated can either motivate or prevent politicians from altering existing frameworks. Political ideology and identities are inherently related to individuals opinions about regulation and government intervention, with those identifying as more conservative favoring less government involvement and those identifying as more liberal favoring a more active government. While political persuasion is likely remains a critically important determinant in individuals views on regulating the tech sector, individuals experiences directly interacting with social media sites and products may also influence their views.

This analysis applies Bayesian statistical methods to test is public opinion on regulating social media companies can be explained entirely as a function of ones political views or if the types of content users see and

1

the user experience on social medias sites is also related to views towards regulation. Put another way, are opinions about regulating big tech all just politics, or is the relationship more complex?

If opinions about regulation are associated with specific aspects of the social media experience, social media companies may be able to restructure existing algorithms to down-prioritize these types of content as a means of increasing their favorability and preventing increased government intervention.

The Data

In 2018 the Pew Research Center surveyed a representative sample of Americans on their use or and feeling toward social media. Participants were asked whether the government should regulate social media companies more than they already are, about the same as they are now or less than they are now. This item was made dichotomous by coding respondents who support current levels or reduced levels of regulation with 0 and respondents favoring increased regulation with 1. This coding was used in all subsequent analyses.

Demographic items were respondents indicated their political identity, race and age were used as measures of political persuasion and identity. Political ideology was described as either very conservative, conservative, moderate (reference class), liberal or very liberal. Age was divided by users who were 18-29, 30-49 (reference class), 50-64 or 65 and above. Respondents could indicate their race as wither white (reference class), black, asian, mixed race or other.

User experiences with social media content was measured through three separate variables: the frequency of encounter posts that increased negative affect, the frequency of encountering politically charged, triggering and controversial content and the frequency of encountering disinformation.

Factor analysis was used to combine 3 separate items about the frequency of seeing content that made users feel angry, depressed and lonely into a single continuous latent variable representing negative affect. Similarity, a factor scores were used to combine a series of 4 separate items asking about the frequency that users encountered posts about race relations, sexual harassment/assault, gun control/gun violence and immigration into a single continuous latent variable representing politically charged content. A single item where respondents were asked to describe if they saw more posts promoting deception, saw more posts trying to point out misinformation or an equal amount each (reference class) was included as a measure of disinformation.

To understand the relationship between political identity, social media content and views on regulation 3 candidate models were created and compared against each other.

Hypotheses: comparing 3 separate models

Model 1:

Views on regulation as a function of political ideology, age and race

Model 1 is the uses only political identification, race and age to predict whether a given individual factors expanding existing regulations of social media companies or believes current regulatory practices are sufficient. If model fits the data the best, this implies that opinions on the regulation of social media companies are an extension of already existing political beliefs.

$$regulation_i \sim Binomial(n, p_i)$$
$$logit(p_i) = \alpha + \beta veryconservative_i + \beta conservative_i + \beta liberal_i + \beta veryliberal_i +$$
$$\beta age : 18 - 29_i + \beta age : 50 - 64_i + \beta age : 65andup_i +$$
$$\beta black_i + \beta asian_i + \beta mixedrace_i + \beta race : other_i$$

Model 2:

Views on regulation as a function of political ideology, identity and features of social media sites

Model 2 includes the same political and identity predictors as model 1 abut adds the measures of social media based features (negative affect, charged content and deceptive vs corrective information).

If model 2 fits the data better than the other models this supports the hypothesis that views towards the regulation of social media companies is not a purely political issues but is also related to the content and types of experiences encountered by users.

$$regulation_i \sim Binomial(n, p_i)$$
$$logit(p_i) = \alpha + \beta veryconservative_i + \beta conservative_i + \beta liberal_i + \beta veryliberal_i +$$
$$\beta age: 18-29_i + \beta age: 50-64_i + \beta age: 65andup_i +$$
$$\beta black_i + \beta asian_i + \beta mixedrace_i + \beta race: other_i +$$
$$\beta negativeaffect_i + \beta chargedcontent_i + \beta deception_i + \beta correctingmisinformation_i$$

Model 3:

Views on regulation as a function of identity and features of social media sites with the role of user experience varying across different political ideologies

Model 3 allows the relationship between social media content and opinions on regulation to vary across different political ideologies. The opinions of those with less extreme political opinions may be more malleable that those who identify as highly ideological. By estimating different slopes for each political identity, model 3 allows for these differential effects.If model 3 fits the data best, this implies that for some political identities, views on social media is are only a function of political identity, but for other political identities views on regulation vary with different experiences on social media sites.

$$regulation_i \sim Binomial(n, p_i)$$
$$logit(p_i) = \alpha_{ideology_i} +$$
$$\beta_{ideology_i} negative\ affect_i + \beta_{ideology_i} charged\ content_i +$$
$$\beta_{ideology_i} deception_i + \beta_{ideology_i} correct\ misinformation_i +$$
$$\beta age: 18-29_i + \beta age: 50-64_i + \beta age: 65andup_i +$$
$$\beta black_i + \beta asian_i + \beta mixedrace_i + \beta race: other_i$$
$$\begin{bmatrix} \alpha_{ideology} \\ \beta_{ideology} \end{bmatrix} \sim MVNormal \left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, S \right)$$
$$S = \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} R = \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix}$$
$$\alpha \sim Normal(0, 2)$$
$$\beta\ negative\ affect \sim Normal(0, .25)$$
$$\beta\ charged\ content \sim Normal(0, .25)$$
$$\beta\ deception \sim Normal(0, 1.5)$$
$$\beta\ correct\ misinformation \sim Normal(0, 1.5)$$
$$(\sigma_{alpha}, \sigma_{beta}) = Exponential(1)$$
$$R = LKJcorr(1)$$

Analysis

Selecting Priors and Drawing from the Prior Predictive Distribution

Specifying priors is the first step of a Bayesian analysis. Given that there already exists a large body of evidence connecting political ideology and views on regulation, informative priors were given to political identification.

In contrast, less is known about relationship between age, race and social media related variables. This uncertainty was reflected with less informative priors that were all centered at 0 (showing equal likelihood of favoring or not favoring increased regulation) and given wider standard deviations.

The specific assigned priors are shown in the code block printed below. The following code not only assigns priors but draws from the prior predicative distribution to ensure that the selected priors make plausible predictions.

```r
## Set priors
# for demogrpahic only model
base_prior <- normal(c(-1, # prior for very conservative
                       -.25, # prior for conservative
                       .25, # prior for liberal
                       1, # prior for every liberal
                     rep(0, 7)),  #  weaker priors for race and age
                   c(rep(.5, 4), # tigter scale for political ideology
                     rep(1, 7)), # more flexibility for other demogrpahics
                    autoscale = F)


## Model 1

# demographics only
M1 <- stan_glm(
  regulation ~
    very_conservative + conservative + liberal + very_liberal +
    age_18_29 + age_50_64 + age_65_up +
    black + asian + mixed_race + other,
  data = model_data,
  family = binomial,
  prior_intercept = normal(0, 2, autoscale = F),
  prior = base_prior, # informative priors for political identification
  prior_PD = T, # samples from prior predictivee distribution
  seed = 2,
  cores = 4,
)

# Model 2
## Adding priors for social media variables
full_prior <- normal(c(-1, -.25,.25, 1, # informative priors for political ideology
                     rep(0, 11)),  #  weaker priors for race, age adn social media
                   c(rep(.5, 4), # tigter scale for political ideology
                     rep(1, 6), # more flexibility for other demogrpahics
                     rep(.25, 2),
                     rep(1, 3)),
                    autoscale = F)

M2 <- stan_glm(
```

```r
  regulation ~
    very_conservative + conservative + liberal + very_liberal +
    age_18_29 + age_50_64 + age_65_up +
    black + asian + mixed_race + other +
    # adding UX varaibles
    Negative_Affect + Charged_Content + deception + correct_misinformation,
  data = model_data,
  family = binomial,
  prior_intercept = normal(0, 2, autoscale = F),
  prior = full_prior,
  prior_PD = T,
  seed = 2,
  cores = 4,
)

# Model 3
M3 <- stan_glmer(
  regulation ~
    # Population Estimates:
    Negative_Affect + Charged_Content +
    deception + correct_misinformation +
    age_18_29 + age_50_64 + age_65_up +
    black + asian + mixed_race + other +
    # Varying Slopes by ideology:
    (
      Negative_Affect + Charged_Content +
        deception + correct_misinformation |
        ideology
    ),
  family = binomial("logit"),
  model_data,
  prior_intercept = normal(0, 2, autoscale = F),
  prior = normal(0,c(rep(.25, 2),rep(1.5,9)), autoscale = F),
  prior_PD = T,
  QR = TRUE,
  adapt_delta = .99,
  seed = 1234,
  chains = 2,
  cores = 2
)
```

Visualizing the Prior Predictive Distribution

The plot shown below provides predictions drawn from the priors for each of the 3 models. Predictions are grouped by political ideology to inspect informative prior given to different ideologies.

The plot suggests that all predictions are in a reasonable range and that the assumptions of the prior specifications are reflected in draws from the prior predictive distribution. Different political ideologies are grouped together in model 3 due to the multi-level structure of the data.

Prior Predictive Draws
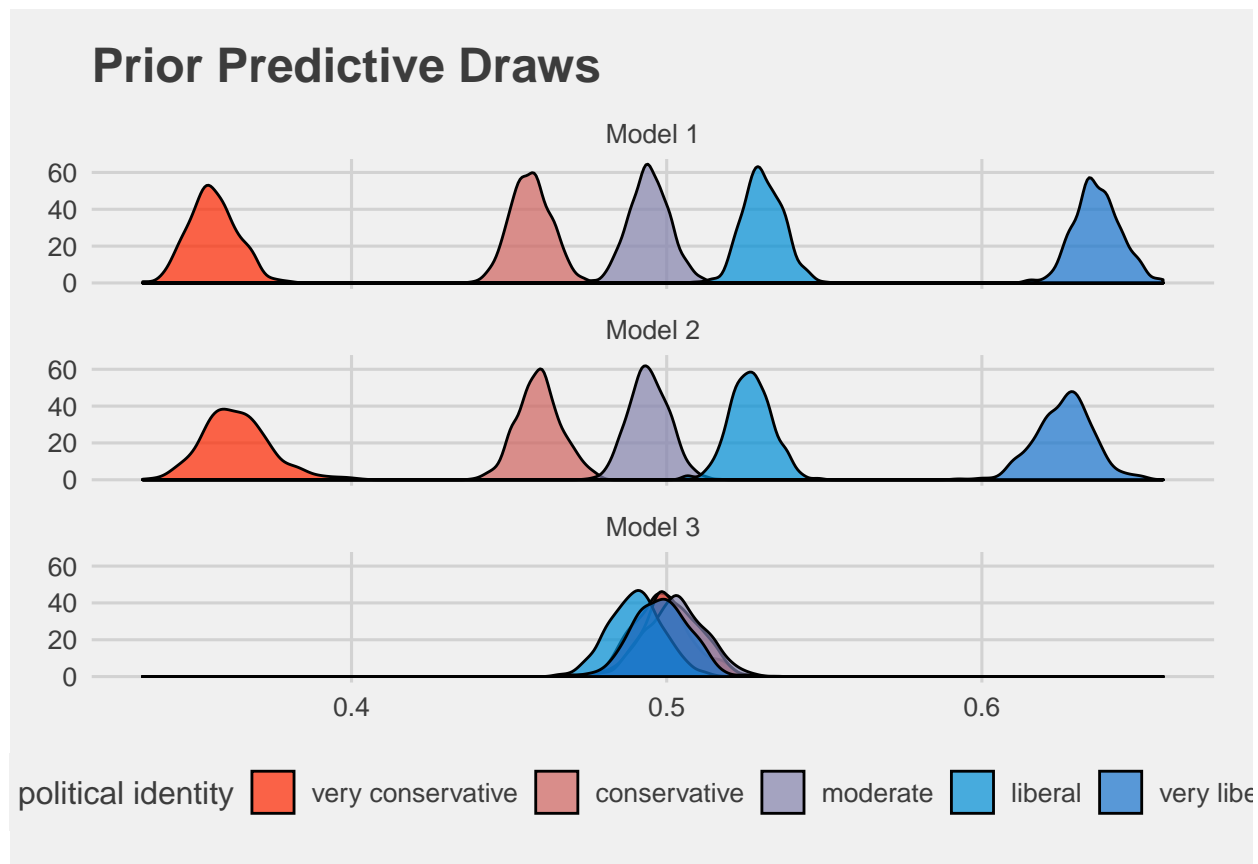
Conditioning on the Data: Fitting and Comparing Models

After assessing the prior distributions, the code chunk below conditions on the data and fits the models. These results can be used to conduct inference and evaluate the established hypotheses.

```
# Fitting to the data
M1 <- update(M1, prior_PD = F)
M2 <- update(M2, prior_PD = F)
M3 <- update(M3, prior_PD = F)
```

Leave one out cross validation (LOO) was used to compare the three models. This approach tests model on unseen data and prevents over-fitting. The model that maximized elpd provides the best fit to the data. Shown below in Table 1. Model 2 which includes social media experience variables with an equal effect across different political ideologies produced the best results.

Table 1: Model Comparisons

	elpd_diff	se_diff	elpd_loo	se_elpd_loo	p_loo	se_p_loo	looic	se_looic
M2	0.000000	0.000000	-2009.843	15.56040	15.56006	0.2620555	4019.687	31.12079
M3	-5.185577	2.192971	-2015.029	14.93688	23.59048	0.4268576	4030.058	29.87375
M1	-24.858873	7.573431	-2034.702	13.87719	11.87562	0.2280922	4069.405	27.75437

Diagnostics

The LOO comparisons shown above indicated that model 2 was the best model relative to the other candidate models, however, several diagnostic tests are necessary to ensure that the model is accurately specified. The LOO comparison shown above is a relative measure and without proper diagnostic checks, it is possible that all 3 models were poorly fit. Assessment of various diagnostic tests suggests that the final model is well specified.

**Posterior draws compared to the observed data**

The plot below shows the observed data plotted over each draw from the posterior distribution. Results suggest the posterior distribution strongly fits the observed density of the actual data points and implies good model fit.
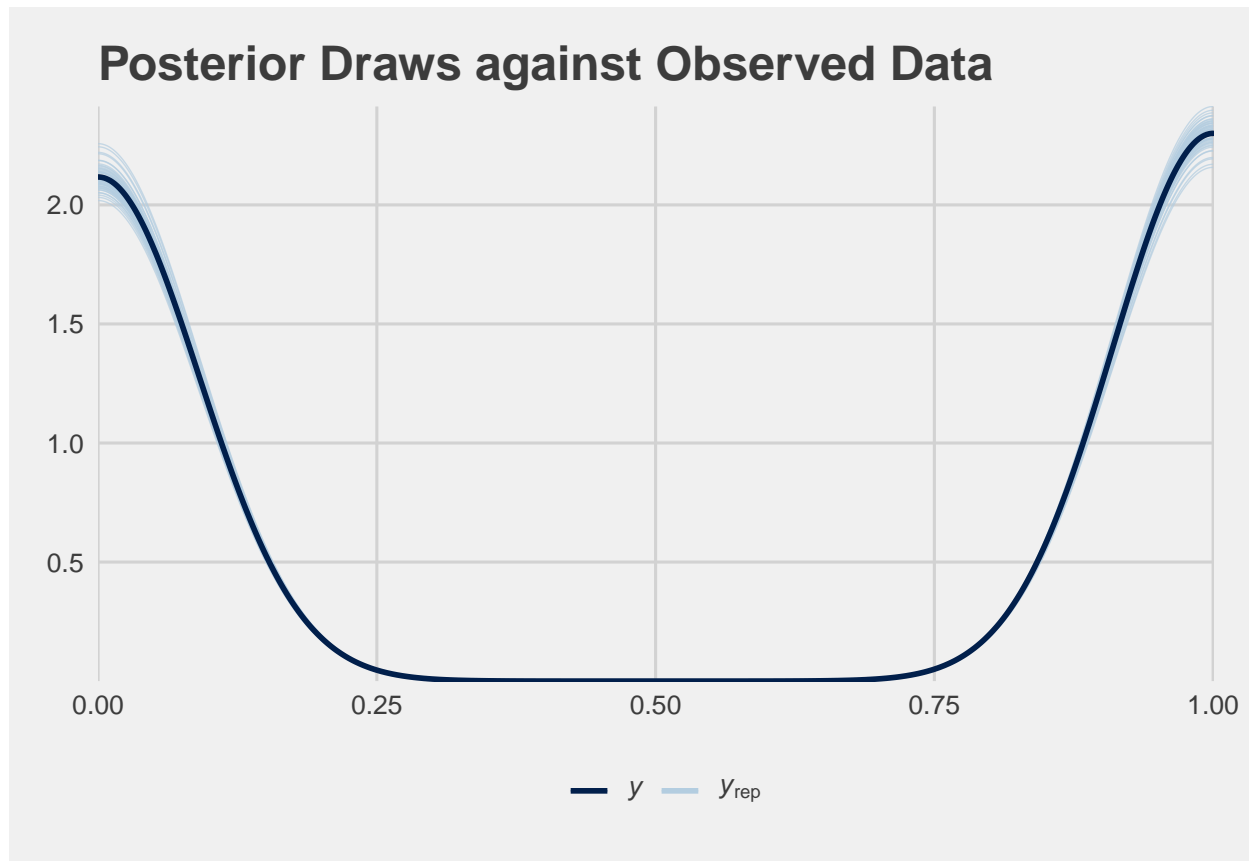
Checking for outliers and leverage points

As with any general linear model, outliers and high leverage points can bias the regression line and are problematic. The plot below shows there are no outliers or high leverage points that bias the realizations of model parameters.

PSIS diagnostic plot
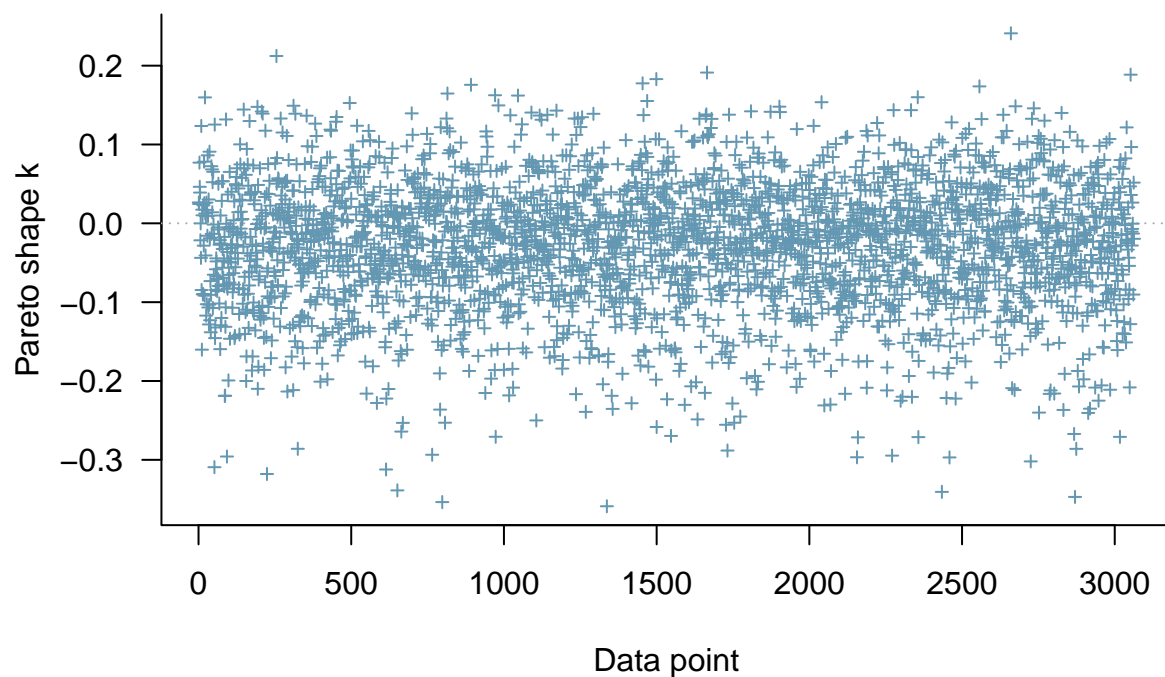
```
## NULL
```

Calibration

Logistic regression models return predictions in the form of percentages. The higher the percentage the more likely a data point is to be in class 1, the lower the percentage the more likely it is to be in class 0. Calibration plots compare predictor probabilities against empirical probabilities. When calibration is good, an event that is predicted to haven 70% time actually happens 70% of the time. In calibration plots, "good" calibration is reflected when points are close to the signal line. Shown below, the final model is well calibrated and, in general, higher predicted probabilities coincide with higher empirical probabilities.
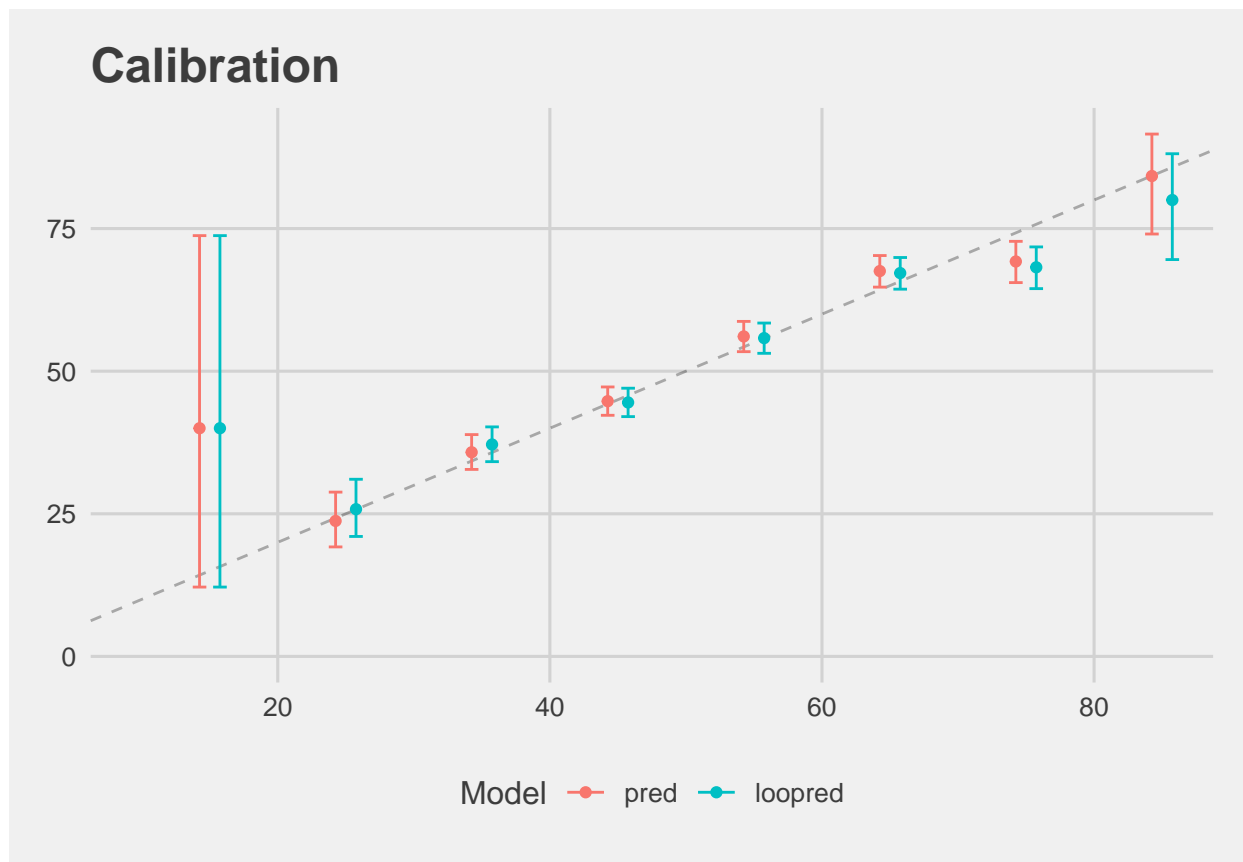
Inference: interpreting results

Model 2 had the best fit to the data which provides evidence that opinions about the regulation of social media companies is not only a function of political identity but is related to the types of content users see while on social media sites. Moreover, Model 2 holds that the association between social media content and views on regulation is equal across different political ideologies.

Parameter realizations

Table 2 presents the median value of each parameter along with a corresponding 90% intervals.

Social media content

The plot below visualizes variables related to experience on social media sites. Holding political ideology constant, users who see content that increases negative affect are more likely to favor increasing regulation. A similar pattern emerged for charged content, where reports of seeing more politically charged and controversial posts was associated with favoring increased regulation.

There was more uncertainty surrounding the association of deception and correcting misinformation, this is reflected in the wider distributions. That being said, users who reported seeing more deceptive posts consistently favored increasing regulation. The association between seeing more posts that try to correct misinformation was the most uncertain of the social media variables, however, over 95% of draws from the posterior distribution signaled that users who saw more posts attempting to correct misinformation were against expanding regulation.

Table 2: Model Output

|	Median	5%	95%
(Intercept)	-0.26	-0.40	-0.12
very_conservative	-0.63	-0.85	-0.40
conservative	-0.49	-0.65	-0.33
liberal	0.49	0.32	0.66
very_liberal	0.85	0.65	1.05
age_18_29	-0.04	-0.24	0.16
age_50_64	0.39	0.24	0.55
age_65_up	0.51	0.34	0.68
black	0.11	-0.11	0.33
asian	0.42	0.05	0.80
mixed_race	0.50	0.17	0.84
other	0.22	-0.01	0.45
Negative_Affect	0.20	0.13	0.27
Charged_Content	0.16	0.09	0.23
deception	0.31	0.13	0.47
correct_misinformation	-0.18	-0.34	-0.01
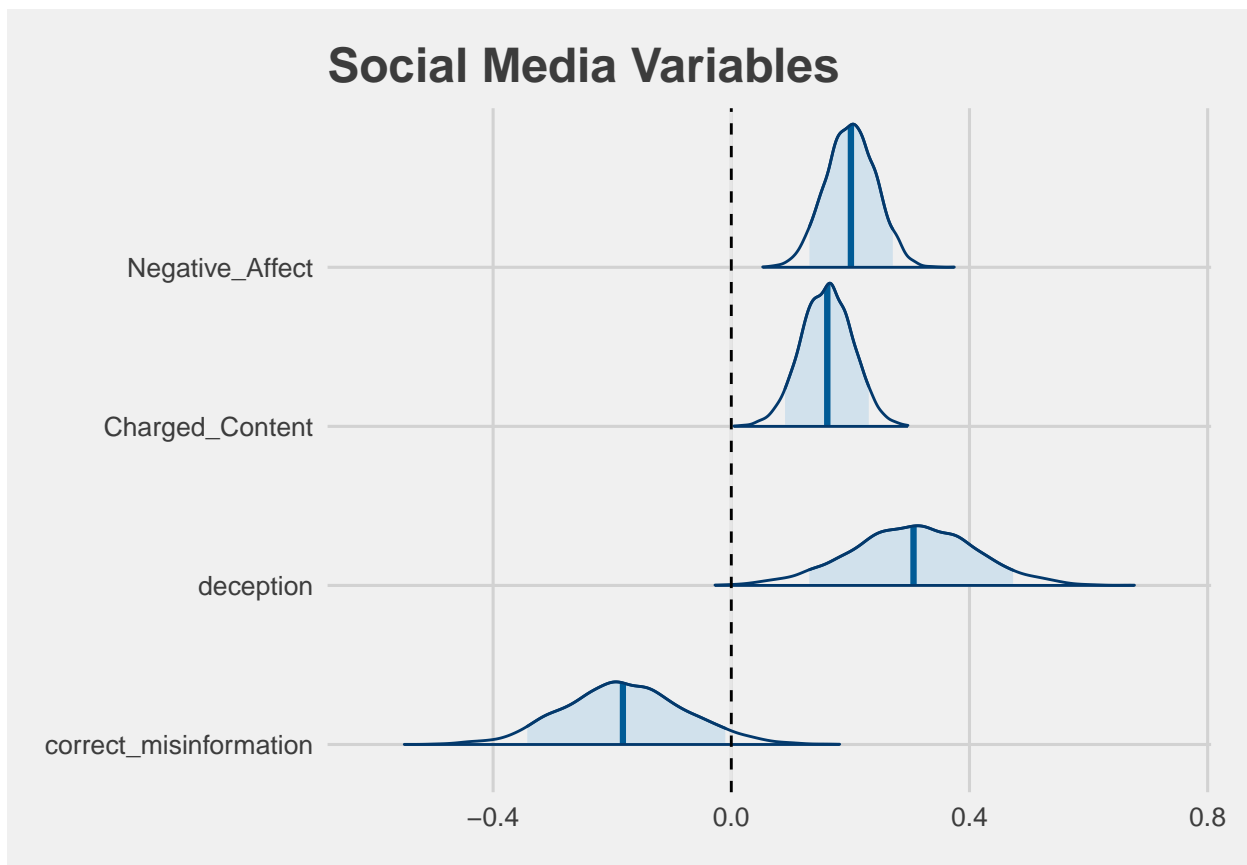
Political ideology

As expected there was a high degree of evidence relating political ideology to views on regulation of social media. More liberal respondents consistently favored increasing regulations while more conservative respondents were against expanding regulations.
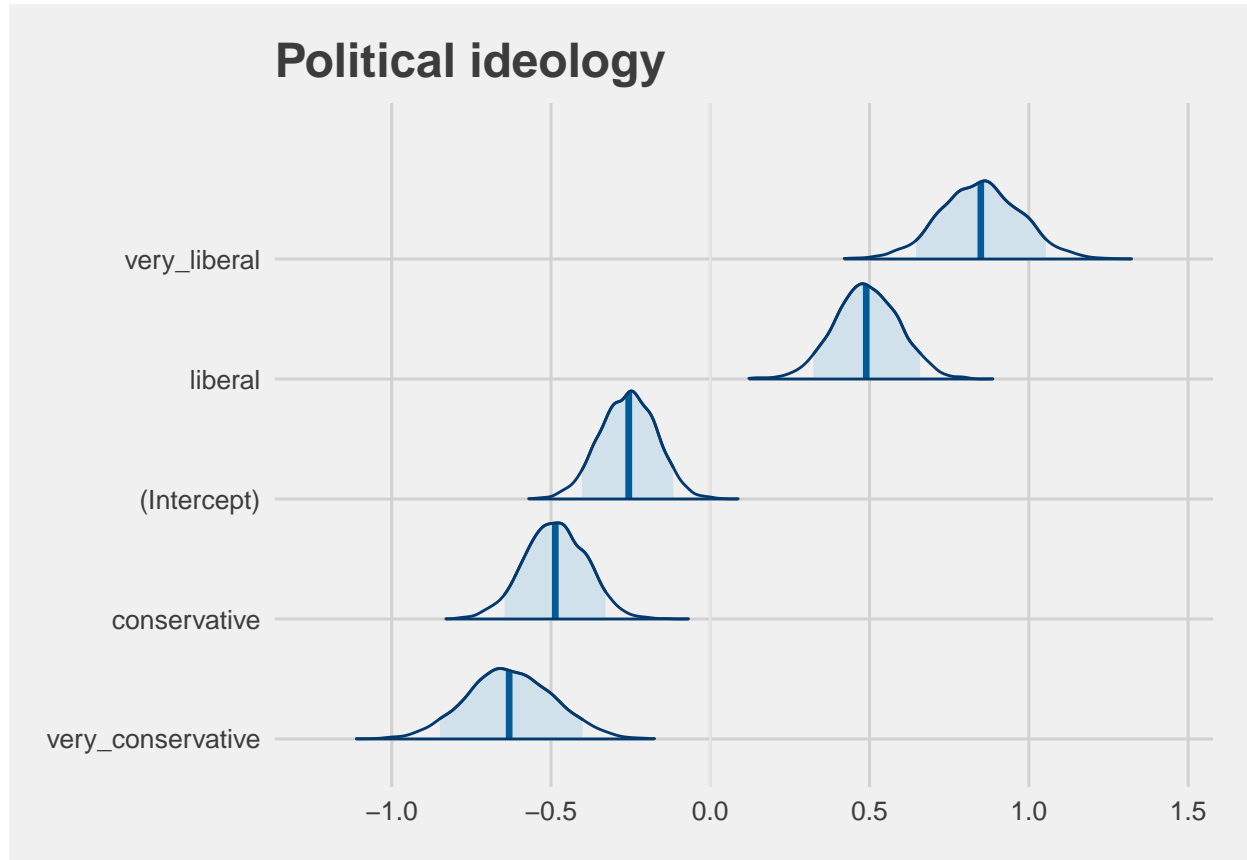
Other demographic predictors

Understanding demographic associations was not the main focus of the present analysis, however, several trends emerged among different age and racial groups. Compared to younger age groups, those 50 (both ages 50-64 and 65 and up) were more in favor of expanding regulations. Compared to other races, whites were more against expanding regulations.

Demographic Variables
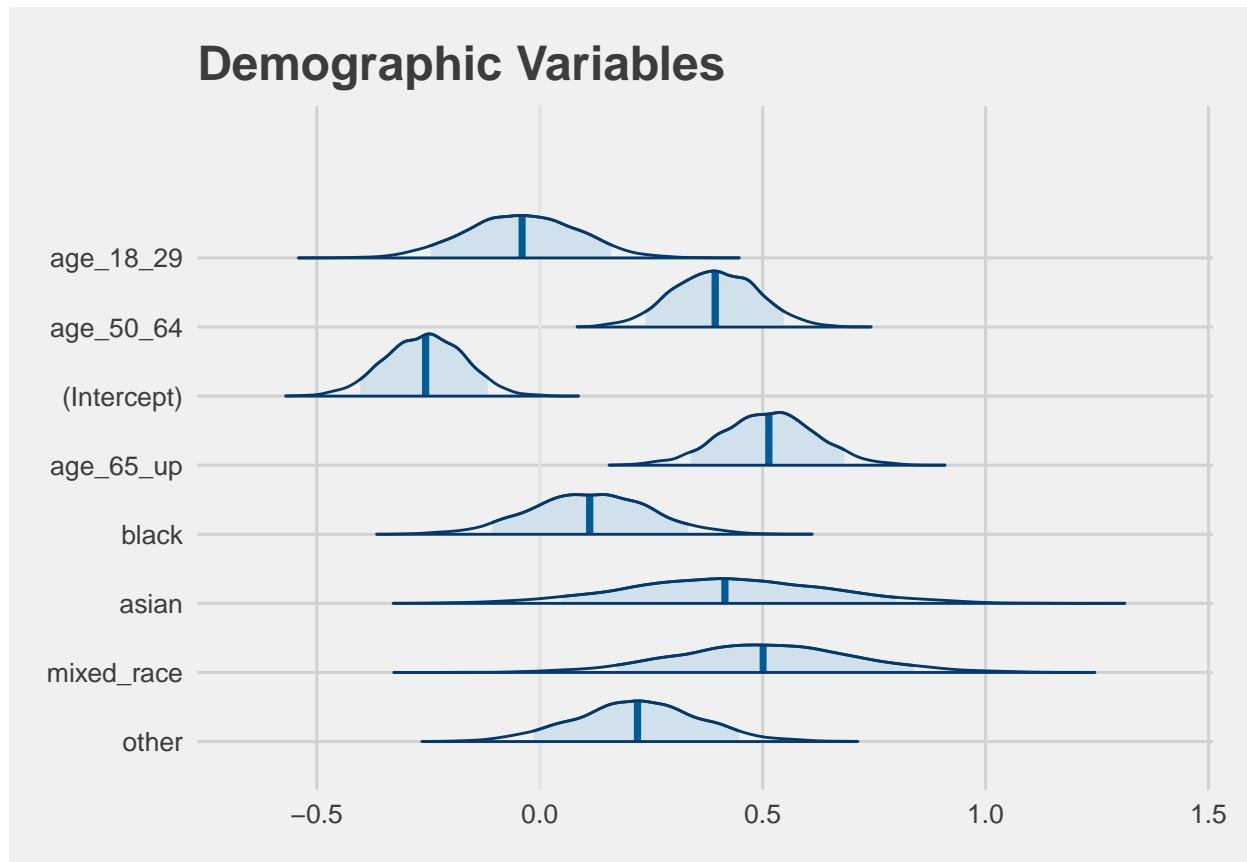
Conclusion and Notes on Causality

This analysis found that the types of content users report seeing is associated with their views on whether or not social media companies need to be more regulated. This is valuable to social media companies in that it identifies that down prioritizing deceptive information and posts that generate negative affect may be an effective stratagy to increase favorability among the general public. This analysis also identifies several political, age and racial demographics were social media companies have an opportunity to improve relations with.

The analysis conducted represent a stating point for more in depth studies, but it needs to be noted that the design used is not causal in nature. The results reported do not imply that social media content causes users to favor more or less regulation. While this is possible, it can not be confirmed through this study design. Subsequent intervention studies, should explore if down prioritizing posts with negative affect, charged content or disinformation directly lead to changes in opinion.