The Dissertation Committee for Yujin Kim Certifies that this is the approved version of the following Dissertation:

How Language Use on Facebook Drives Affective Polarization

Committee:

Natalie Jomini Stroud, Supervisor

Roderick P. Hart

Sharon Jarvis

Josephine Lukito

How Language Use on Facebook Drives Affective Polarization

by

Yujin Kim

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2022

Dedication

This dissertation is dedicated to my parents, who love me unconditionally.

Acknowledgements

First and foremost, I am extremely grateful to my advisor, Talia Stroud. Your guidance, support, and patience during my PhD have been invaluable. Your mentorship is a true blessing. I would like to extend my deepest appreciation to my committee members who were critical to completing my dissertation. To Rod Hart, thank you for your patience and insights for broadening my perspectives in academia which numbers are not able to explain. To Sharon Jarvis, thank you for your encouragement and friendly advice throughout the graduate school experience. To Jo Lukito, thank you for helping me keep my research on track and move it forward.

My dissertation is deeply indebted to many people at the Center for Media Engagement and the Department of Communication Studies at the University of Texas at Austin. To Caroline, Nicholas, Natalie, and Mary Margaret, your efforts contributing to my research projects are appreciated. To Ellery, Kat, Victoria, and Melody I gratefully appreciate your help and support. To Anne, Aida, and Jennifer, thank you for helping me get through my PhD program. I am also grateful that I have met enthusiastic scholars during my PhD. To Alex Curry, Ashley Muddiman, Cynthia Peacock, Emily Van Duyn, Jay Jennings, Jessica Collier, and Josh Scacco, thank you for your motivational work that has inspired me.

I very much appreciate my friends and family for supporting me. To Joanne Click, thank you for your valuable advice in life and for proofreading my drafts. To Taeyoung Lee, thank you for being a genuine friend who understands the challenges in graduate school. To Jin Seo, Jinmi Song, Eunbin Jo, and Chang Sun Oh, thank you for always encouraging me in difficult times. To Hyeok Jin Kim, thank you for taking care of our parents while I'm in the U.S. I appreciate that you're my younger brother. To my parents, Young Bok Kim, Mi Kyung Lee, who always believe in me, thank you for the unending support and love.

Finally, the survey results reported here were obtained from searches of the iPOLL Databank and other resources provided by the Roper Center for Public Opinion Research.

Abstract

How Language Use on Facebook Drives Affective Polarization

Yujin Kim, Ph.D.

The University of Texas at Austin, 2022

Supervisor: Natalie Jomini Stroud

Social media has changed people's experience with political language. Social media platforms have become places where political tolerance is rarely preserved and expressions of high levels of intense dislike proliferate among the American public toward the opposing political party. Focusing on political content on Facebook, I argue that partisan language, when combined with a highly charged affective aspect, exacerbates affective polarization. I further propose that the type of language that increases affective polarization is promoted by social media algorithms, hence making partisan content more visible. To examine these dynamics, I provide both longitudinal and cross-sectional evidence of how partisan language, combined with affect, influences users' political attitudes, leading to increases in affective polarization in society-at-large. At the aggregate level, I look at the effects of affective political language on engagement metrics and also examine time-series analysis to see how partisan affective language correlates with feelings toward the political parties. Using a computational social scientific lens, I apply natural language processing techniques such as sentiment analysis, incivility identification, and the determination of partisan in-/out-group entities to identify affective language in Facebook hyper-partisan pages. I use data consisting of 14 million posts from 1,493 hyper-partisan pages on Facebook between January 2008 and mid-November 2020 were as well as aggregated public opinion surveys corresponding to those timelines. At an individual level of analysis, I test the effects of partisan affective language on

polarization using a mock Facebook setting. The survey experiment provides causal evidence about the relationship between language and polarization. Results confirm two routes to polarization: negative and uncivil language attracts more engagement leading to more visibility and such language corresponds with how people feel about the political parties. This dissertation further emphasizes how incivility and negativity related to polarization differently and how the presence of political targets influences affective polarization. Partisan differences in the mechanisms of engagement with political posts and affective polarization are discussed. The dissertation overall shows that political language can shape affective polarization, a process that is facilitated by social media.

Table of Contents

List of Tables

List of Figures

Chapter 1: Does Social Media Exacerbate Polarization and, If so, How?

If you noticed that political spaces on social media, such as Facebook or Twitter, are more hostile and harsher in terms of language usage than what you experience in face-to-face contexts, you are not alone. A Pew Research study asked respondents for their views about social media and found that a considerable number of the users were of the opinion that these platforms are angry, disgraceful, and uncivil venues that do not offer space for serious engagement about politics (Duggan & Smith, 2016). In fact, similar percentages of partisans on both sides of the aisle reported feeling "worn out" by the amount of political content they come across on social media. Social media platforms have become places where political tolerance is hardly preserved, as evidenced by the use of language.

Political language, in particular partisan language, includes markers of group membership that convey the superior "us" and/or the inferior "them." Often divisive, insulting, and dehumanizing terms describe "them." A notorious example is "libtard" (a neology of "liberal" and "retard") that degrades "people on the left" as intellectually disabled. Even various legacy online dictionaries define "libtard" as an offensive term used by extreme right-wing people to describe a person with left-wing political views (e.g., "Lexico," n.d.; "Cambridge Dictionary," n.d.). Although not a direct counterpart of "libtard," "trumpanzee" (a combination of "Trump" and "chimpanzee") is another example of a political portmanteau describing supporters of Donald Trump as un-evolved and animalistic. Given that these types of political slang pervade all manner of online political communication, what does political language tell us about the current political atmosphere? Does political language cause polarization, or does political language reflect polarization? My dissertation aims to examine such dynamics.

PROBLEMS OF POLITICAL DISCOURSE ONLINE AND POLARIZATION

Why might language matter in the current political environment? More than 70 years ago, George Orwell (1946) noted a decline in the quality of political discourse and its political causes in his essay "Politics and the English Language," in which he observed, "When the general atmosphere is bad, language must suffer… But if thought corrupts language, language can also corrupt thought" (p. 228). Although Orwell's assessment of language in political discourse reflected his pondering about politics in the late 1940s, his observations suggest a direct link between language usage and unfavorable public sentiment. This feeling seems to resonate, with many Americans' believing that now is a time when "the general atmosphere is bad." Considering that a majority of Americans relies on digital platforms for news and political information (Shearer, 2021; Shearer & Mitchell, 2021), it is likely that users have witnessed toxic language that has the potential to affect how they evaluate the political landscape.

Partisan affective polarization – described as intense dislike toward the opposing political party and attachment toward one's own party (Iyengar et al., 2012) – characterizes the public's political sentiment. This form of polarization is considered problematic because it threatens core components of democracy, such as tolerance for opposing views, compromise, and shared opinions about which issues are important. My dissertation focuses on the U.S. context where there are currently high levels of affective polarization, documented by research based on explicit and implicit measures of individuals' attitudes and perceptions toward co- and opposing partisans (e.g., Garrett et al., 2014; Iyengar & Westwood, 2015). Research also implicates the media as related to higher levels of affective polarization (e.g., Lelkes et al., 2017; Levendusky, 2013; Stroud, 2010) however, we have little understanding of how the media drive affective polarization in the current media environment. Against this background, my dissertation proposes to examine whether high levels of affective polarization proliferate as a function of language embedded in the new media environment.

The Internet and social media have changed people's experience with political language and, as I investigate here, contribute to increasing polarization. Two possible rationales include: (a) selective exposure and filter bubbles can polarize by limiting people's exposure to only their own views (e.g., Iyengar & Hahn, 2009; Pariser, 2011; Stroud, 2008, 2010; Sunstein, 2007) and (b) inadvertent exposure to counter-attitudinal information can also polarize (e.g., Bail et al., 2018; Weeks, Lane, et al., 2017). Although some studies have proposed that incidental exposure to online political information leads people to encounter cross-cutting views and might subsequently reduce polarization (Barberá, 2014; Brundidge, 2010), others have found that, in general, exposure to cross-cutting content on social media is limited (Bakshy et al., 2015; Bessi et al., 2016) and cross-cutting content exposure can be polarizing (Bail et al., 2018). Building on this research, my dissertation asks what it is about political content that yields polarization. I argue that social media, Facebook in particular, contributes to and accelerates people's exposure to content on hyper-partisan pages and that the language contained in this content plays an important role in polarizing the public.

Overarching Questions for Polarization, Social Media, and Language

After the advent of Web 2.0, the Internet made various new technological phenomena possible, such as social networks and interactive user communication (Cormode & Krishnamurthy, 2008). Such characteristics have changed the way that people are exposed to political information through greater customization and selectivity. The networked algorithms of social media are known to surface news that supports an individual's existing beliefs, an effect that is enhanced by clusters of "friends" who consume and disseminate similarly likeminded news content. An algorithmic process of content distribution is further influenced by users' interests, preferences, and engagements (DeVito, 2017; Thorson & Wells, 2016). In the pages that follow, my dissertation directs attention to Facebook, the most popular social media platform in the U.S., along with the partisan content available on that platform. I examine how political language from Facebook posts affectively polarizes the mass public and, additionally, how users' engagement with political content fosters hyper partisanship through Reactions (i.e., Like and the other emotional reactions such as Love, Haha, Wow, Sad and Angry) as well as Shares, and Comments.

Moreover, I pursue another less well understood topic about how toxic political communication and partisan identity influence how people process political language in ways that lead to affective polarization. Language usage not only serves as a cognitive and subtle way to deliver intergroup bias (Maass et al., 1989), but its valence is associated with intergroup linguistic bias (Leyens et al., 2000; Matsumoto et al., 2016). Language contains rich and high-dimensional information, making it possible for partisan language to reflect, intentionally or unintentionally, a person's attitudes, perceptions, behaviors, and emotions toward political in- and out-groups. In the context of affective polarization, political language containing certain (often derogatory) terms associated with partisanship can be measured by examining specific linguistic characteristics of political discourse. Studies have indicated, for example, that negativity and incivility operate as manifestations of polarized attitudes (Gervais, 2014, 2015; Hwang et al., 2014; Rains et al., 2017). My dissertation treats negativity as a type of valence (positive, negative, and neutral affectivity) and an emotional expression (e.g., sadness, anger, fear, disgust, remorse, and rancor are regarded as negative emotions that can be conveyed through language) and incivility as a sub-category of negativity with several unique features, a distinction detailed later in this document. This study further investigates the relationship between negative and uncivil words posted on political

Facebook pages and partisan polarization. My argument is that language can prime partisan identity and, when combined with a highly charged affective aspect (i.e., negativity and incivility), motivates public affective polarization. The pattern between partisan language and affective polarization, I further propose, is facilitated by social media where users' engagement via Reactions, Shares, and Comments with political content leads the Facebook algorithms to prioritize such content.

To do this work, I used computational approaches to extract meaningful linguistic features and applied time-series analysis at the aggregate level and, additionally, a survey experiment that focuses on the individual level. Widespread availability of massive and cumulative textual data as well as advances in statistics and computational science offer support for research conducted on the effects of partisan language based on issues or ideology to understand political polarization. By using these technological advances, I examine negativity and incivility of textual content by applying sentiment analysis and machine learning approaches to political posts on highly partisan pages of Facebook. The survey experiment provides stronger causal evidence about the relationship between language and polarization.

OVERVIEW OF DISSERTATION

This dissertation focuses on language as a driver for engagement on social media and polarization through an examination of both aggregated and individual data. In particular, I pay attention to negativity, incivility, and presence of political targets in Facebook posts. I propose two routes for polarization. First, negative and uncivil language on political content attracts more engagement, which affects social media algorithms and results in this content being prioritized and gaining more visibility. Higher visibility means more exposure leading to polarization. Second, the presence of negative and uncivil content polarizes the public. Both aggregated and individual data were gathered to investigate why polarization may occur in the social media context. Chapter 2 reviews the literature of what we know about political polarization, identifies causes of political polarization, and examines the role of social media in affecting polarization. I argue that language is a possible driver for political polarization and discuss the important features of language, negativity, incivility, and political targets in the context of polarization and social media. Research questions and hypotheses at both the aggregated and individual levels follow.

Chapter 3 outlines my methodology of using two aggregate-level analyses and one individual-level analysis. At the aggregate level, I first look at the relationship between the language attributes (negativity and incivility) of political posts and social engagement using computational approaches and social scientific methods. I use CrowdTangle, a Facebook tool, to gather the text of public posts made on the platform and data on how people engaged with the posts. Next, I inspect the relationship between the language attributes (negativity, incivility, and presence of political targets) of political posts and affective polarization at the aggregate level using computational approaches and time series analysis. The post data again comes from CrowdTangle and I pair it with over time aggregated survey results. The individual-level analysis is an experiment in a mock Facebook setting to see the direct effects of language attributes on social engagement and affective polarization.

Chapter 4 reports on the findings of whether negative and uncivil language from hyper-partisan pages on Facebook invites more engagement than posts not using this language. I compare the effects of negativity and incivility on social engagement. Partisan differences are also examined to evaluate whether the negativity/incivility and engagement relationship varies for posts from liberal and conservative Facebook pages.

Chapter 5 evaluates whether negative and uncivil language on Facebook political posts is related to polarization among the public. I also analyze whether presence of partisan entities (in-groups only, out-groups only, both in- and out-groups, and non-explicit entities) alongside negative and uncivil content corresponds with affective polarization in the public. Comparisons of negativity vs. incivility and partisan differences are also explored.

Chapter 6 tests the causal relationship of negativity and incivility in Facebook posts on both social engagement and affective polarization at the individual level, reporting the results of the survey experiment. Comparisons of negativity vs. incivility and of in-group vs. out-group follow.

Finally, Chapter 7 summarizes the findings making sense of the three studies and highlights the implications of the findings. Contributions to the theoretical literature and suggestions for future research are discussed. Limitations of this dissertation are addressed subsequently. The aggregated and individual levels of analysis provide complementary answers for understanding the role of political language in furthering affective polarization.

Chapter 2: Polarization, Its Causes, and the Role of Language

Some people may avoid talking about politics in person or online. Others think that the incivility contained in some social media discussions is not problematic. Both happen according to the Pew Research Center (Jurkowitz & Mitchell, 2020; Mcclain et al., 2021). The implication may be that social media use is unrelated to polarization. Even if the language on social media is affectively polarizing, if most people avoid political engagement, such language may not necessarily lead to polarization. For the latter, if people think the current language of online political discourse is not worrisome, exposure to uncivil language may not affect polarization.

Yet other scenarios are possible: (1) even those who want to avoid politics may encounter it incidentally, resulting in affective polarization, and (2) polarization happens through mere exposure to political content, even those who think online discussions are acceptable may become polarized. These possibilities suggest that language can still affect the public even if they don't seek it out or find it acceptable. Therefore, we need a comprehensive understanding of affective language on social media to examine the dynamics of political polarization, which this dissertation pursues.

This chapter provides an overview of what we know about political polarization along with possible causes that scholars have identified. I review the literature about social media and polarization and argue that language can serve as a driver for polarization. By proposing research hypotheses and questions, this chapter details the possible roles and features of language in producing political polarization.

ROLE OF POLITICAL POLARIZATION AND TYPES OF POLARIZATION

Major forms of political polarization include party-ideology sorting, issue positions, and affect-based partisan identity. Each type of political polarization is discussed, along with an explanation for why this dissertation focused on affective polarization.

Party-Ideology Sorting

Compared to the 1970s, American politics has experienced a large shift to higher ideological homogeneity within parties at the mass and elite levels. This homogeneity has been documented by low levels of split voting and increased levels of party loyalty (e.g., Jacobson, 2000; Levendusky, 2009). Party-ideology sorting, a term used to describe this form of polarization, refers to the degree to which political ideology aligns with party identification (Fiorina & Abrams, 2008; Hill & Tausanovitch, 2015; Levendusky, 2009). It is characterized by increasing ideological distance between the parties as well as increasing ideological homogeneity within parties (Druckman et al., 2013; Levendusky, 2010). That is, conservative Democrats have become liberals and liberal Republicans have become conservatives.

It is worth noting that the concept of sorting can be distinguished from polarization in this way: sorting describes how closely partisanship and ideology are related while polarization characterizes a changing state by which the public is taking more ideologically extreme positions (Levendusky, 2009). That is, party-ideology sorting means choosing a side, but not necessarily adopting a more extreme version of a party’s position, whereas polarization entails a tendency toward extremes as well as taking sides. This distinction is important because sorting explains the existence of those remaining in the middle. On the other hand, polarization is defined as both a state, “the extent to which opinions on an issue are opposed in relation to some theoretical maximum,” and as a process, driving “the increase in such opposition over time” (DiMaggio et al., 1996, p. 698). The process approach tends to focus on a dynamic and evolving sense of polarization (e.g., DiMaggio et al., 1996; Lelkes, 2016).

Party-ideology sorting is clearly evidenced among political elites. The established definition of elites refers to those who are actively involved in politics, such as political activists and elected officials, in contrast to the public, or more specifically, voters in American politics (Fiorina et al., 2006; Lee, 2002; Zaller, 1992). For instance, Republican members of Congress have become more conservative as measured by roll call voting records (Hetherington, 2009; Hill & Tausanovitch, 2015; McCarty et al., 2006) while Democratic members of Congress have become more liberal as measured by interest-group ratings (Stonecash et al., 2003). The consequences of party-ideology sorting entail legislative gridlock and stalemate (Jones, 2001).

When it comes to the mass public, most scholars tend to agree that the mass public has become more sorted over the last five decades – Democrats have become more liberal and Republicans have become more conservative (Abramowitz, 2010; Fiorina et al., 2006; Levendusky, 2009). As a possible cause of mass sorting, some scholars see the phenomenon as largely attributable to the elites (e.g., Druckman et al., 2013), echoing Zaller's (1992) logic that ordinary voters rely on elites to make sense of the political world. Levendusky (2010) also argues that elite polarization enhances the American public's attitude consistency, making it easier for individuals to follow their party's clear cues. In other words, elite polarization changes the way ordinary people form their political opinions.

Sorting is not a focus of my dissertation because I aim to identify the dynamics of political polarization resulting from political language. Although it is possible that political language use by those who are correctly sorted may prompt sorting among the mass public, this perspective is unrelated to the attributes of language and the role of the Internet, including social media. This dissertation research is interested in how linguistic features (i.e., negativity and incivility) in partisan content contribute to political polarization as well as to increased visibility of such language on social media.

Issue-Based Polarization

Scholars point out that since the 1970s the American public has been deeply divided with regard to both politics and culture (e.g., Bishop, 2008; Abramowitz & Saunders, 2008). They claim that over time, while the number of ideological moderates has declined, issue position differences between Democratic and Republican identifiers have increased (Abramowitz & Saunders, 2008; Campbell, 2008; Jacobson, 2000). More concretely, the distance between Republican and Democratic party activists on social and economic issues has grown, coinciding with greater divergences as reflected in the party platforms (Layman, 1999; 2001). There are both negative and positive consequences of issue-based polarization among the public (e.g., Layman et al., 2006). Some argue that polarization among the public has decreased trust in government and increased political disengagement (e.g., Fiorina et al., 2006), while others believe that polarization may promote political participation, such as higher turnout (Abramowitz & Saunders, 2008), and increase consistent and cohesive policy attitudes (Carsey & Layman, 2006).

Issue differentiation by parties extends over a wide range of issues including aid to Blacks, abortion, jobs/living standards, health insurance, and presidential approval (Abramowitz, 2010; Abramowitz & Saunders, 2008). Furthermore, students of the “culture war” that focuses on moral issues, such as abortion, gay marriage, and geographical divisions (i.e., red- and blue-states), have observed partisan differences among the mass public (Green et al., 1996; Hunter, 1994; Layman, 2001). This line of research suggests issue polarization among the public is widespread.

Not everyone agrees with the idea that the severity of political polarization on issues is overwhelming. Some scholars contend that the majority of Americans have remained centrists and ideologically moderate (Fiorina et al., 2006). Fiorina and Abrams (2008) did not observe a systematic increase in polarization on political issues. Several scholars reached the conclusion that growing political polarization in American politics is a "myth" (Ansolabehere et al., 2006; Fiorina et al., 2006; Glaeser & Ward, 2006). Related to the earlier discussion of party-ideology sorting, this group of scholars also argues that while the mass electorate is "better sorted" and "more tightly aligned" in partisanship and ideology today than a few decades ago, issue-based polarization is limited because sorting can occur without the public embracing more extreme issue positions (e.g., Levendusky, 2009). Overall, from this perspective, the extent of issue polarization among the public is relatively modest.

Issue-based polarization is one type of political polarization, but this concept is not a good fit for answering the question that I raise: specifically, how affective and partisan language drives political polarization. It is acknowledged that issue and ideology positions can have both cognitive and affective components (e.g., Mason, 2016; Webster & Abramowitz, 2017). However, I argue that the role of language is more likely to influence forms of polarization that capture people's social and emotional attitudes toward partisans. In general, affect – distinct from cognitive perception – refers to any feeling that is experienced with or without awareness and is an essential component of more complex emotions (Huitt, 2003). Related to the political world, affect involves emotional responses that incorporate reasoning to guide one's interpretation of politics (Marcus et al., 2000). My dissertation argues that affective and partisan language influences the way people engage with polarized content posted on social media. Also, I argue that people's exposure to such language increases affective polarization, the topic of the next section.

Affective Polarization

Distinguished from sorting and issue-based political polarization, recent studies provide evidence about how mass political divisions are driven by affective polarization based on social identity (Iyengar et al., 2012; Mason, 2015; 2018). Affective polarization is a concept of both positive sentiment and attachment toward co-party members and negative sentiment and hostility toward opposing party members (Iyengar et al., 2012; Iyengar & Westwood, 2015). Building on social identity theory (Tajfel, 1978; Tajfel & Turner, 1986), scholars argue that polarization comes with an in-group vs. out-group mindset and that affective polarization can occur regardless of extremity of issue positions (DiMaggio et al., 1996; Iyengar et al., 2012; Mason, 2015, 2018; Mason & Wronski, 2018). According to this view, party identification, one of the most important forms of social identity in American politics (Campbell et al., 1960; Green et al., 2002; Greene, 1999), represents strong group affiliation and, therefore, forms the basis for affective polarization. In this dissertation research, affective polarization refers to increasing social psychological distances among partisans in the mass public based on emotional and visceral responses.

Much empirical evidence supports the view that affective polarization among the mass public stems from group-centric feelings. Since the 1980s, negative ratings of the out-party and its supporters have increased significantly among partisan identifiers (Haidt & Hetherington, 2012; Iyengar et al., 2012). Both positive stereotyping of the in-party and negative stereotyping of out-party affiliates have increased significantly over time (Iyengar et al., 2012). Measurements recorded by the ANES feeling thermometer scales, which are recognized as a major indicator and involve assessing how warmly or coolly people feel toward the Democratic Party and the Republican Party, have documented increasing affective polarization among those who identify with a party and ideology. Along with feeling thermometers of parties and ideologues, research by Iyengar et al. (2012) found evidence in the U.S. (and in the U.K.) of manifestations of affective polarization by measuring respondents' thoughts about their child's inter-party marriage and partisans' stereotypical trait ratings (i.e., closed-minded, hypocritical, selfish, and mean). Specifically, their research demonstrated that Democrats and Republicans not only increasingly loathe the opposing party but also attribute negative traits to members of the out-party. In a similar vein, Mason (2016) used feelings toward presidential candidates as a measure of affective polarization and found that over time the American public has continued to feel happier about their party's candidates and angrier about their partisan opponent's candidates. This type of polarization, the social attachments of party affiliation that result in political divides among the mass public, is the focus of this dissertation.

It is particularly important to study affective polarization because of its problematic consequences. Because negative emotions such as aversion reduce willingness to compromise and openness to opposing views (MacKuen et al, 2010), dislike and hostility toward opposing partisans potentially threatens integral parts of democracy such as compromise and tolerance (Gutmann & Thompson, 2012; Kingzette et al., 2021). Affective polarization is also influential interpersonally, reinforced by political discussion with the like-minded others (Hutchens et al., 2019).

POSSIBLE CAUSES OF POLARIZATION

What causes political divisions among the mass public? Although there are many possible causes, several focus squarely on the role of the media, in line with my interest in this dissertation. The two main literatures identifying media-related causes of polarization are (1) selective exposure and echo chambers for news and political information and (2) incidental exposure on social media that exposes people to cross-cutting views. Both of these literatures have insights into whether exposure to political information leads to (de)polarization.

Selective Exposure and Echo Chambers

With the emergence of partisan media, scholars have investigated the fragmented media environment as a cause of political polarization among the mass public (Bennett & Iyengar, 2008; Iyengar & Hahn, 2009; Morris, 2007; Lelkes et al., 2017; Levendusky, 2009; Prior, 2007; Stroud, 2008, 2010, 2011). Selective exposure research, in particular, has sought to unveil the media's role in affecting political polarization. By selective exposure, I mean exposure to messages that match one's political views and the avoidance of views that do not. A substantial amount of research concludes that media choices are driven by preexisting partisan orientations. In terms of partisan news labels, for example, Republicans prefer articles labeled as coming from the more conservative-leaning Fox News, while Democrats chose articles from the more liberal-leaning CNN and NPR (Iyengar & Hahn, 2009). Individuals with stronger political ideology that used liked-minded partisan media disclosed even more polarizing attitudes over time compared to individuals with stronger ideology who did not consume like-minded media (Stroud, 2010). Baum and Groeling (2008) analyzed "top news" from various media outlets during the 2006 election, including cable news websites, talk radio, and political blogs, and found that partisan media (i.e., liberal/left-leaning and conservative/right-leaning) were more likely to report news that was harmful to opposing partisans and helpful to co-partisans. These studies suggest that partisan media use different language, that people are attracted to likeminded media, and that partisan media can polarize.

In the current Internet-based media environment, the variety of political interaction and information gathering options are likely to lead to greater political polarization. The Internet and use of social media increase the probability of exposure to congenial information by creating "echo chambers" (Sunstein, 2007) that are amplified by "filter bubbles" where platforms algorithmically surface personalized content that accords with individuals' interests (Pariser, 2011). Moreover, in the context of social media, it matters not only how individuals choose news media and content, but also how they engage with political content. Settle (2018) notes that social media users face the challenge of “how to sift through the vast amount of information available” rather than “how to use heuristics to form opinions in the absence of information” (p. 200). In doing so, people avoid certain (and most) information and selectively attend to particular cues. Taken together, research on selective exposure and filter bubbles is implicated in this dissertation based on the underlying assumption that people tend to be exposed to congenial information that can polarize.

This dissertation proposes that political language on social media explains and strengthens a selective exposure and polarization process. Specifically, linguistic features of an affective nature can make people more attentive to political content and arouse more affect in response to political information. This process can provoke engagement on social media, such as Reactions, Shares, and Comments on Facebook. The increased visibility of this content on social media can give rise to affective polarization.

Incidental Exposure on Social Media as a Possible Causes

There are ongoing debates over two conflicting possible effects of social media on political polarization. On the one hand, social media is seen as a place where selective exposure to political information both dominates and aggravates political polarization. On the other hand, social media is seen as a place where incidental exposure to cross-cutting views occurs and may mitigate political polarization. A review of this debate is important because it hints at the effects of the distinctive environment that social media may have created. Mutz (2006a) proposes that exposure to cross-cutting views promotes a better understanding of both like-minded and oppositional perspectives and political tolerance, essential for democracy. In that regard, Sunstein (2018) argues that two requirements should be met for a well-functioning democracy: a) incidental exposure to information and b) citizens' diverse experiences that enable them to share with one another as a way to create social glue for resolving social problems. Whether this happens on social media, and the implications for polarization, is unclear.

Scholars have found that social media increases selective exposure, possibly aggravating partisan polarization (e.g., Bail et al., 2018; Conover et al., 2011; Hong & Kim, 2016; Lee et al., 2014). The homogeneity of social media networks suggests the limited likelihood of individuals being exposed to cross-cutting views (Bakshy et al., 2015; Vaccari et al., 2013). The relationship between selective exposure on social media and political polarization is strengthened by other features, such as political discussions (Lee et al., 2014) and algorithms that cause someone's news feed to be more personalized and relevant to his/her interests (Sunstein, 2018). Overall, these findings suggest that social media reinforce existing political divisions in society by exposing people to likeminded content.

Even though many scholars blame partisan news media as a main driver of partisan polarization, several other studies have found that people are, in fact, exposed to cross-cutting views on Facebook (Anspach, 2017; Beam et al., 2018; Messing & Westwood, 2014), on Twitter (Barberá, 2014), and on a portal site (Kobayashi et al., 2020). This type of exposure is facilitated by the type and volume of social endorsements (Anspach, 2017; Messing & Westwood, 2014). These studies propose that incidental exposure to cross-cutting views alleviates the effects of partisan media and selective exposure on partisan polarization.

Yet encountering politically opposing views may not reduce polarization. Weeks, Lane, et al. (2017) found boomerang effects of incidental exposure to counter-attitudinal information so that those with stronger party identification were more likely to seek pro-attitudinal information after an inadvertent encounter with attitude-challenging political content that, in turn, can lead to subsequent sharing behavior on social media. On Twitter, exposing people to opposing political views that conflict with their own attitudes was found to induce backfire effects because of people's likelihood to counter-argue (Bail et al., 2018). Strong partisans are not only rarely exposed to disagreeable information, but they are also motivated to reinforce their preexisting views when they are presented with counter-attitudinal information. As a result, incidental exposure to non-likeminded content may aggravate political polarization. This assumption points to the need for greater understanding of the underlying mechanisms by which the political content on Facebook affects polarization. The next section addresses Facebook as a major political source for the public and examines user engagement metrics as political signals.

FACEBOOK AND POLARIZATION

This dissertation focuses on political content on Facebook. In the pages that follow, I review the use of Facebook for political information. Facebook engagement metrics are described to understand how they feed into the algorithm and affect what people see, which can have implications for polarization.

Facebook as a Political Information Source

Facebook is one of the main channels offering the public easy access to information and political content. According to recent Pew Research Center surveys, about 70% of U.S. adults have used Facebook and among current adult users, about 74% go to the site more than once a day (Gramlich, 2019; Pew Research Center, 2019). Furthermore, 68% of U.S. adults get news at least sometimes on social media and about 43% of Americans get most of their news on Facebook, which is the most used social media site in the U.S. (Pew Research Center, 2018; Shearer & Matsa, 2018). A substantial body of research supports the assumption that social media, including Facebook, serves as an important source of political information (e.g., Bode, 2016; Cacciatore et.

al, 2018; Feezell, 2018; Oeldorf-Hirsch, 2018; Wells & Thorson, 2017) and political engagement (e.g., Bode, 2012; Conroy et al., 2012; Kahne & Bowyer, 2018; Valenzuela et al., 2009).

Not only does digital media provide a special platform for news and political messages, but it also offers a unique information experience for users in terms of news that is customized. On Facebook, the term for this experience is described as "curated flows" of customized content that are based on one's personal choice/interest, socially mediated information, and algorithms based on a user's activities and relevant data on the platform (Thorson & Wells, 2016). In particular, the core algorithmic values of Facebook's News Feed consist of users' friend relationships, interests, and engagement as well as when content was posted, platform priorities, page relationships, negative preferences, and quality of content (DeVito, 2017). Notably, Settle (2018) defined the unique characteristics of content that are disseminated on social media, including Facebook, as "a personalized, quantified blend of politically informative expression, news, and discussion seamlessly interwoven into a wider variety of socially informative content" (p. 50). As Settle's definition suggests, it is difficult to separate political content that aggravates political polarization from other information contained within a News Feed when studying Facebook's overall effects.

To sum up, Facebook's News Feed consists of political and non-political content that reflects what the audience consumes and can lead algorithms to prioritize likeminded content. My interest in this dissertation is the effects of posts from partisan pages. It is assumed that most posts from political pages are hyper-partisan, but there may be some cases of nonpolitical or less partisan posts depending on the interest of political pages (e.g., issues) and types of the posts (e.g., images). For my purposes, I examine the language used regardless of the content of the post. The next section reviews how individuals on Facebook interact with political content through Facebook's technical features (e.g., Reactions, Shares, and Comments) and what user engagement means in the context of polarization.

Social Metrics on Facebook and its Role in Polarization

As one of the main channels that plays a role in the public's access to political information, Facebook has received much scholarly attention that examines how the flow of political information influences individuals' political behaviors and polarization (Bakshy et al., 2015; Bessi et al., 2016; Del Vicario et al., 2016; Johnson et al., 2020). Relevant studies have focused on different Facebook metrics to understand the dynamics of the public's interactions. This section examines the metrics to which this dissertation research directs attention and explains why these metrics are important.

First of all, this dissertation defines engagement with political content on an individual's Facebook News Feed as users interacting with political content in ways that can be seen by others. Those metrics include Reactions (separate emotional reactions of Like, Love, Haha, Wow, Sad and Angry), Shares, and Comments. This type of engagement with political content is closely related to posts on an individual user's News Feed. According to the News Feed algorithm on Facebook, posts that are likely to appear first are prioritized by the user's connections (friends or family members), activity (including both the user's own and their friends' postings, updates, and reactions to a post), and multiple people's reactions to a post they watched or read in the News Feed (Facebook, n.d.). Engagement can reveal a user's attention and information transmission practices (Garz et al., 2020; Schmidt et al., 2017), and can be used by algorithms to prioritize future content.

The general functions and meanings of Reactions, Shares, and Comments on a Facebook post are generally agreed upon among researchers (Bessi et al., 2016; Schmidt et al., 2017). "Likes" (which is a sub-type of Reactions) function as positive feedback to a post or an endorsement of the post. Liking political posts is usually associated with content that agrees with users' beliefs (Stroud et al., 2017). Earlier, Pariser (2011) critiqued the Liking function as problematic because of its limited applicability to different types of posts, such as those that are important as well as negative (e.g., such as natural disasters or bad news). Moreover, Likes involve minimal attention (Heiss et al., 2019) and are used for signaling purposes such as expressing one's feelings (Garz et al., 2020).

Since February 2016, Facebook has expanded its emotional responses to include more than Like by allowing users to engage with a total of five additional reactions (i.e., Love, Haha, Wow, Sad and Angry; Krug, 2016).[1] Studies have looked at the effects of diverse Reaction features available for interacting with news content (e.g., Larsson, 2018) and that use Facebook Reactions as proxies for predicting characteristics of the news content (e.g., causing controversies and debates) (e.g., Basile et al., 2017). This study considers Reactions, including Like, as expressions of emotional engagement that can be a mixture of approval (e.g., Love), disapproval (e.g., Angry), and less clear indicators about positive or negative reactions (e.g., Sad, Haha, Wow) in reference to the content or the content producer. Certain emotional reactions are associated with certain types of in language. For example, Angry reactions are associated with statements expressing threats or danger while Love reactions are linked to statements showing optimistic or happy motives (Sturm Wilkerson et al., 2021). Although Facebook could deprioritize posts garnering lots of Angry reactions, news coverage suggests that this was not always the case (Merrill & Oremus, 2021). Reactions are used in my dissertation to capture six distinctive feelings toward the content or the account that produced the post.

[1]Facebook continues to test new features for Reactions, for example, new reacts such as "Care" reactions for both Facebook and Messenger as response options for COVID-19 related updates in March 2020 and a "Thankful" reaction in several markets during the lead up to Mother's Day in 2016 (Hutchinson, 2020). The Care and Thankful reactions are not a focus of my dissertation because of their temporary characteristics.

A second measure, “Shares” indicates a desire to disseminate information to friends in the individual’s network and/or the desire to increase the visibility of content by reposting content on a user’s Timeline. Sharing is a feature that shows how an individual’s interpersonal network influences his/her own flow of information on Facebook. On social media, sharing can be seen as a social activity based on the networked “imagined audience” (Marwick & boyd, 2011). Sharing pro-attitudinal messages is a stronger predictor of political polarization and participation than merely being exposed to such pro-attitudinal messages (Johnson et al., 2020; Shin & Thorson, 2017). Sharing in my dissertation refers to motivated activity to spread content and/or to increase visibility within the sender’s network, which requires relatively higher engagement effort than reacting to content.

Finally, “Comments” can produce collective debates around a topic generated by a post. Compared to Reactions and Shares, Comments are more likely to inspire further engagement by others, depending on the language used. For example, unlike Shares and Reactions (except for certain responses in specific circumstances such as Haha and Angry), Comments can contain various forms of incivility (Coe et al., 2014). Negative and uncivil comments can prompt others’ intentions to participate in the discussion (Wang & Silva, 2018). In addition, posts that contain negative tonality (e.g., political failure or crisis) or that express emotions (e.g., words like happy, sad, worst) can yield further comments (Heiss et al., 2019). For Comments, the cost of engagement and cognitive effort is larger than the other metrics because Comments can be used to express an individual’s emotions or thoughts. Facebook algorithms also use Comments as an indication of News Feed priority (Facebook, n.d.). In my dissertation, Commenting refers to reactions that require the highest effort to engage with a post and that may incorporate both negative and uncivil components relative to content posted by others.

As this discussion illustrates, each metric (Reactions, Shares, and Comments) has a different user dynamic that needs to be examined separately. For example, Garz et al. (2020) demonstrated that posts with politically congenial messages are more likely to receive Likes and Shares compared to uncongenial messages; however, that does not apply to Comments. Political language in a post that primes partisan identity may be a predictor of social metrics, and each social metric may work differently.

It is important to understand what sort of content earns more Reactions, Comments, and Shares because the responses trigger the algorithm to prioritize the content and increase its visibility by exposing the content to a network of other users. Thus, a critical question is: What are the linguistic features that link engagement with political content posted on social media? By analyzing the comment section of the *New York Times*, Muddiman and Stroud (2017) found that language predicts engagement (i.e., recommendations and abuse flags) when incivility is included or when partisan identity is involved. Wang and Silva (2018) found that the presence of incivility in a Facebook comment leads to greater intentions to engage with the comment (i.e., aggregated intention for sharing, commenting, and liking). Although this prior research is helpful, it has not examined how engagement on social media at an aggregate level relates to polarization. Overall, (1) the effects of partisan and affective language are associated with increased user engagement and (2) the user engagement on Facebook ultimately guides the algorithm to prioritize the content. Based on that understanding, my dissertation understands whether affective and partisan language use in posts on Facebook predict user engagement metrics. In the next section, the role of partisan and affective language in contributing to political polarization is discussed.

Language as a Possible Driver of Affective Polarization

Settle (2018) identified Facebook as a platform that drives polarization by helping people recognize partisan identity, reinforcing biases toward out-groups, and leading people to over-generalize such perceptions. Consuming and interacting with political content aligned with one's political social identity on social media, therefore, is observed to promote stereotypes toward political out-groups, which is known to increase affective polarization. However, not clear is how such processes develop and what drivers lead to polarization among the mass public. In this dissertation, I propose that language is a key reason that content on Facebook polarizes, and that Facebook algorithms amplify this kind of content to encourage greater engagement. In the next section, I review studies that have examined language from a social psychological perspective and discuss its connections to affective polarization.

Language as Intergroup Bias

Earlier research on the topic of prejudice saw language as a symptom of underlying bias. By identifying the degree of prejudice relative to groups, Allport (1954) argued that language (i.e., antilocution) is the first indicator of group prejudice. When negative views about outgroups are expressed willingly, he argued, they serve as a potential indicator of prejudice, paving the way for a transition to a more intense level. In other words, language is assumed to be a fundamental sign of social stereotypes and intergroup biases.

Later research in social psychology examined language as a cognitive process that maintains social stereotypes and intergroup biases. Notably, van Dijk (1987) approached prejudice as a cognitive and social process that is delivered through communicative discourse by pointing out how racial prejudices are reproduced in everyday conversations. That is, interpersonal communication that often contains negative evaluations of minority groups expressed with a shared rhetoric can work as a persuasive strategy that increases racial prejudice. Maass et al. (1989) found that within an intergroup context, people tend to describe undesirable out-group and desirable in-group behaviors in higher levels of abstraction, indicating that biased language is used in a stable, highly implicit manner. Taken together, intergroup discrimination is expressed through both explicit and implicit language use.

Critics of an emphasis on cognitive factors have observed that social prejudice studies often ignore the role of emotions. In that regard, Leyens et al. (2000) suggested the need to pay attention to emotions as essential in studies of intergroup discrimination. For example, anger in relation to competition and conflict produces automatic biased attitudes against out-groups (DeSteno et al., 2004). Although this dissertation does not specifically distinguish between primary and secondary emotions, attributes of emotions can indicate an inclination of group bias. Based on an adaption of the Implicit Association Task (IAT), developed by Greenwald et al. (1998), Leyens et al. (2000) found that in evaluations of their own in-group compared to out-groups, regardless of the emotion's valence, people preferred to associate secondary emotions with their own group (i.e., more human quality emotions such as affection, pride, nostalgia, remorse, and rancor) and primary emotions (e.g., joy, sadness, anger, fear, disgust, and surprise) in evaluations of out-groups. This indicates that emotion reflects the perceived superiority of "our" in-group. Similarly, Vaes et al., (2002) found that email that used secondary emotions to ask for help from a stranger received more prosocial reactions than using primary emotions, meaning that the sender could appear to be closer to their own group when the soliciting language was associated with certain emotions considered as more human. These findings indicate emotions are related to biases in intergroup relations. Against that background, my dissertation examines language that is associated with affect as a rationale for bias toward political in- and out-groups, presenting yet another factor that leads to polarized attitudes in the political context.

Examining Language in Political Polarization

Political language is thought to reflect political beliefs and thought (Geis, 1987). For example, Frank Luntz, a political consultant hired by the Republican National Committee, identified language that resonated with voters using focus group and polling techniques and recommended that Republican politicians use certain words and phrases to appeal to the majority of their voters on issues (Luntz, 2004). Some of his work on political terms has changed the political scenery, including the use of “climate change” instead of “global warming” because the policies should be discussed as conservationist perspectives rather than those of (extreme) environmentalist (Luntz, 2004). Use of such partisan language has been identified among congressional members and, subsequently, by news media in the U.S. (Gentzkow & Shapiro, 2010).

In the realm of political polarization, a growing body of research has focused on how partisan language relates to political polarization. Note that the studies in this section address language and issue-based political polarization rather than affective polarization unless I mention the latter specifically. As the earlier discussion of political polarization suggests, substantial evidence of elite polarization has led to recent studies that analyze how mass partisan polarization is influenced by elite discourse. This line of research focuses on the language used in media coverage, such as newspapers and cable television. As a compelling example, Gentzkow et al. (2016) found that over time, the two major political parties in the U.S. have developed different ways of speaking and that politically slanted news media in the U.S. apply similar ideologically biased language usage. Not only has partisan language usage increased in media discourse, but analysis of the *Congressional Record* reveals that politically polarized language about issues has reached levels that are unprecedented dating back to 1873. Another noteworthy observation developed from research on political language is that partisan language use by Congressional members does not dictate their political ideology (Gentzkow et al., 2016) nor party (Monroe et al., 2008); rather, their political ideology (either conservative or liberal) or party (either Republican or Democrat) is more likely to influence their language use. This indicates a causal direction of political ideology leading to language use.

Linguistic evidence of elite polarization is also found in other types of elite discourse. In language usage preserved in Google Books, the appearance of polarizing phrases from the *Congressional Record* remained relatively low until the late 1990s, then afterwards political discourse became more polarized (Jensen et al., 2012). Given that affective polarization increased during the same time period, there is at least circumstantial evidence that partisan language usage in publications may explain the dynamics of polarization. Other studies have demonstrated that politically biased media channels resonate with language that is in accordance with the political ideology of elite discourse in the *Congressional Record* (Gentzkow & Shapiro, 2010; Martin & Yurukoglu, 2017). Together, these studies suggest that partisan language as used in many different contexts may reflect elite partisan polarization. Although elite partisan discourse is not a main focus of this dissertation, examining elite discourse has merit in understanding partisan distinctions in language usage.

Partisan language is also used among the mass public in the digital domain. For instance, linguistic features of users' Twitter feeds, including hashtags, can be used to detect political leanings, such as whether someone is a Democrat or Republican (Barberá, 2015; Colleoni et al., 2014; Pennacchiotti & Popescu, 2011). In terms of political polarization, clusters of political homophily observed in users' Twitter networks have shown a very limited connection between right- and left-leaning users in their retweet networks, depicted as identified homogeneity of content among clusters (Conover et al., 2011). These studies provide additional evidence of the extent to which partisan language is closely associated with political polarization on the Internet and social media.

As more concrete examples, there are several linguistic manifestations of politically targeted words, especially toward out-groups. Hossain et al. (2018) organized political slang by morphological classes: abbreviations (e.g., repub for Republican), acronyms (maga for Make America Great Again), nicknaming (e.g., presbo for President Obama and drumpf for Donald Trump), portmanteau (e.g., killary for killer and Hillary, retardican for retard and Republican), prefixes (adding a prefix to a word, e.g., antiobama, uberwealthy), spelling (intentional misspelling, e.g., mooselim (Muslim) and obammo (Obama), and suffixes (adding a suffix to a word, e.g., clintonian and trumpism). Except for a few cases, most of those words have negative connotations and use insulting tones toward political out-groups. Usage of political language particularly targeted toward outgroups reinforces underlying biases and affect toward associated groups.

As shown, research on political language tends to examine ideological and issue differences in how language is used and many of these words have an affective component. Missing from these discussions is an understanding of how the affective facets of political language in general influence people's attitudes. This dissertation assumes that language used online by both conservative and liberal political pages primes partisan identity. In particular, this study examines posts from partisan pages for its affective aspects (i.e., negativity and incivility).

In the next section, I review recent research on affective polarization and propose that a linguistic approach can make significant contributions to that body of research.

Negativity and Incivility Related to Political Polarization

Many research efforts have documented high levels of polarization, showing that it comes with an in-group vs. out-group mindset. However, there is a missing link about what drives affective polarization and how it occurs. This dissertation argues that the dynamics can be explained as a two-part process. The first process relates to exposure to partisan and affective language that attracts more engagement, thus leading the platform algorithm to post the content at the top of one's News Feed and increasing chances for exposure. The second process, I propose, involves the direct effects of political language exposure that accelerate the development of affective polarization. The following section specifies the corresponding research questions and hypotheses.

POSTS FROM PARTISAN FACEBOOK PAGES, ENGAGEMENT, AND POLARIZATION: RESEARCH HYPOTHESES AND QUESTIONS

Throughout this dissertation, I argue that partisan affective language and the attributes of Facebook that promote such language contribute to political polarization. Although the number of people who attend to partisan Facebook pages may be relatively small, they are likely to be individuals with a high interest in politics and who actively engage with political content. Given the networked and social attributes of Facebook, these individuals play an active role in spreading such content within their networks (e.g., Weeks, Ardèvol-Abreu, et al., 2017), and this content may thus influence a much broader cross-section of Americans, especially more politically moderate users. On average, Facebook users tend to have more than 20% of friends who are from the opposing party or ideological affiliations (Bakshy et al., 2015; Goel et al., 2010), which suggests chances of exposure to cross-cutting viewpoints. On a content level, approximately 6% of content that users see on their Facebook News Feed is political (Facebook, 2021), but the majority of top pages and publisher domains in the U.S., as computed by user engagement on the content (e.g., liking or commenting) or reach (i.e., how many people actually see the content), are political (e.g., USA Patriots for Donald Trump and Occupy Democrats for top pages; cnn.com and foxnews.com for publisher domains) (Facebook, 2020).[2] This indicates that while political content takes up only small portions on News Feed, engagement with and exposure to the content are substantial. Although people can block or unfriend those who share such content, it is difficult for Facebook users to be completely free from exposure to such content because they can encounter that content by chance. When individuals have a large number of friends and/or a wide variety of friends with different interests, chance encounters are more likely.

Negativity

News outlets cover politics increasingly by using negativity, as has been observed in the handling of news coverage. Patterson (1993) showed that negative news coverage of U.S. presidential candidates increased from 1960 to 1992, whereas positive news coverage decreased. More recent research demonstrates comprehensive and aggregated effects of negativity biases in 17 countries, an indication of the predominance of negative news coverage internationally (Soroka et al., 2019).

One rationale behind the prevalence of negativity bias in news content is that negative news attracts attention, leads to larger audiences, increases the perception of newsworthiness, and increases revenue. Hamilton (2004) explains that in news industries where news and information

---

[2] Note that these data were collected leading up to the 2020 election, but there is evidence of engagement with political posts in other times according to CrowdTangle. See more details from https://twitter.com/FacebooksTop10.

are regarded as commodities, privately-owned media organizations are able to spotlight certain features of news (e.g., biased coverage) in order to attract attention and maximize their profits.

As a demand-side explanation, negative framed news is found to be the type of news that people prefer to select more often (Soroka, 2014; Trussler, & Soroka, 2014) although negative news does not always attract more attention than positive news (e.g., Berger, 2013; Soroka et al., 2019). This is because people tend to give greater weight to negative versus positive information (e.g., Ito et al., 1998). Accordingly, negativity bias as expressed by negative tone plays a role in selective attention and in maximizing communication effects (Jing-Schmidt, 2007). In a nutshell, trends in negative news coverage in politics align with individuals' preferences toward negativity. In this study, the concept of negativity in political content is associated with the valence of a given text (i.e., whether it is positive, negative or neutral) and emotional attitude toward a topic or target (e.g., expressions of anger, joy, or sadness), as encompassed by a general concept of sentiment analysis (Mohammad, 2016).

As suggested by discussions of the increasing trends of negativity in the political information environment and people's psychological tendency to attend to negativity, user engagement with negative posts may reflect such tendencies. Even though there are relation-based factors on Facebook that may prevent users from interacting with negative content such as face-work (saving face) (Treste & West, 2013) or violating social norms (McLaughlin & Vitak, 2012), people may still be more likely to react to negative political content on Facebook due to trends in media coverage and their psychological predispositions. However, not all engagement metrics work the same way. In a European context, previous studies demonstrate that negative tone on Facebook attracts more engagement with political content, particularly Comments and Shares in response to postings on political actors' pages in Austria (Heiss et al., 2019) as well as on postings from news organizations' pages in France, Italy, Portugal, and Spain (Salgado & Bobba, 2019). Adding to this work, my dissertation seeks to analyze the relationship between the types of posts on Facebook partisan pages and people's engagement. This is important because more engagement with negative content on political pages can ultimately lead the algorithm to prioritize such content. The relationship between negative language and increased user engagement is directly observable (i.e., each post has both linguistic features on the post and metrics of user engagement), leading to the following hypotheses:

> **H1:** More negative content posted on partisan Facebook pages will receive a greater number of Comments.
>
> **H2:** More negative content posted on partisan Facebook pages will yield a greater number of Shares.

It is less clear how the Reaction metric works relative to negative political content on Facebook partisan pages. Previous studies did not find evidence that people tend to react to negative content by Liking it (Heiss et al., 2019; Salgado & Bobba, 2019), while others find that people press Like as a way to signal a notice of others' posts without expressing any emotional reactions (Spottswood & Wohn, 2019) or to express a relationship-based (with who posts the content) action that is not necessarily based on the content meaning (Sumner et al., 2018). Yet Reactions, such as Angry and Love are associated with negative and positive tones on posts from political actors' pages (Eberl et al., 2020). Given that negative content on partisan pages is likely to be negative toward an out-party and positive content is likely to be positive toward the in-party, and that audiences for these partisan pages tend to be likeminded, I anticipate that Angry and Love will be used to indicate disapproval and approval, respectively, of partisan posts. Haha, Sad, and Wow can be seen as less clear indicators in terms of positive or negative reactions. Against that background, my dissertation examines Reactions as separate emotional metrics with regard to negative language in partisan content that, in turn, can promote visibility of the content and more exposure. Thus, the following hypotheses and research question are posed:

> **H3a:** More negative content posted on partisan Facebook pages will receive a greater number of Angry reactions.
>
> **H3b:** More negative content posted on partisan Facebook pages will receive fewer Love reactions.
>
> **RQ1:** Does more negativity in content posted on partisan Facebook pages correlate with the number of Like, Sad, Haha, or Wow reactions?

In addition to Facebook foregrounding negativity, I also propose that the negativity of political content on the platform prompts increased affective polarization. People who follow likeminded political pages are more likely to be exposed to agreeable content (out-group negativity and in-group positivity) than to disagreeable content (in-group negativity and out-group positivity). Previously reviewed research on selective exposure supports this pattern, and research on exposure to counter-attitudinal posts suggests that those, too, can polarize. The focus here is on the specific type of language used. Negative language, more characteristic of posts targeted toward out-group, may escalate polarizing effects.

> **H4:** Affective polarization will correlate with lagged measures of the negativity of content posted on partisan Facebook pages.

Partisan content often contains negativity toward the out-group, which can aim at an out-group only or can compare an out-group with an in-group (e.g., out-group negativity and in-group positivity). Negativity attracts more attention and is likely higher when directed toward out-groups on political posts on Facebook (Rathje et al., 2021). It is expected that the relationship between affective polarization and the negativity of partisan content will be associated when an outgroup is presented as a target of the negativity. Effects of negative posts mentioning an in-group alone or without mentioning any explicit political entities are considered for comparison. I propose the following:

> **H5a:** Negative content posted on partisan Facebook pages that is targeted toward an out-group will correlate with affective polarization.

> **H5b:** Negative content posted on partisan Facebook pages that is targeted toward both an out-group and in-group will correlate with affective polarization.

Incivility

In addition to increasing negativity in the online media environment, incivility in political discourse has been identified as a source of political polarization and, therefore, the consequences of incivility in the context of affect politics have also been addressed. Although the definition of incivility is disputed, two main concepts are impoliteness and violations of (social or political) norms. Impoliteness can be understood as a lack of etiquette and formality that typically guides conversations smoothly while norm violation refers to undemocratic behaviors that undermine democratic goals (Papacharissi, 2004). In this study, incivility generally encompasses both impoliteness and violation of norms of a democratic system (Mutz, 2015).

What makes incivility distinct from negativity is its feature of being unnecessary and provocative; as such, this study considers incivility as a subset of negativity. Brooks and Geer (2007)'s concept of incivility captures "claims that are inflammatory and superfluous (p. 5)." Other research has identified features of incivility in public discussions (Coe et al., 2014; Kenski et al., 2017; Rains et al., 2017) that include vulgarity, name-calling, pejoratives for speech, lying accusations, and aspersion (Kenski et al., 2017; Rains et al., 2017). Although incivility can be in "the eye of the beholder" (Massaro & Stryker, 2012), the use of language that includes the aspects and substances listed above is considered uncivil in this research.

Users may be more likely to engage with uncivil political content on Facebook. Similar to the findings of studies that examined negativity, analyses suggest that incivility contained within comments is more engaging and induces polarizing effects (e.g., Gervais, 2015; Kim, 2018; Muddiman & Stroud, 2017). Likewise, on Facebook, incivility related to political issues often leads to increased hostility in readers' comments as well as increased intentions to participate in the discussion (Wang & Silva, 2018). Based on the previous studies, therefore, the following hypothesis is proposed:

> **H6:** More uncivil content posted on partisan Facebook pages will result in a greater number of Comments.

Turning to Shares as an engagement metric, the sharing of uncivil content can be motivated by partisan identity. Incivility in political discourse induces emotional reactions and increases engagement in sharing (Wang & Silva, 2018). Sharing political information generated by anger toward the opposing party can be an expression to punish the opposing party or take an action for the perceived injustice (Hasell, & Weeks, 2016). In a similar way of negativity, sharing uncivil content on News Feed may be restrained for users because of social relationships on Facebook (e.g., McLaughlin & Vitak, 2012; Treste & West, 2013). However, several mechanisms could explain sharing of uncivil political content. First, people perceive incivility differently depending on where it comes from (Kim, 2018; Muddiman & Stroud, 2017). That is, in-party incivility is seen as less uncivil than identical incivility from the out-party. Second, when people share content, they think about who will see it. For that reason, they are more likely to share uncivil content that originated with likeminded partisans and they may tend to think of other likeminded partisans when they share. Third, partisan pages on Facebook are more likely to attract in-group members and include posts that are uncivil toward the out-group. In other words, people who follow their partisan side's political pages are more likely to be exposed to like-minded content (out-group incivility and in-group civility) than to counter-attitudinal content (in-group incivility and out-group civility). In a similar manner, the sharing of congenial content directed toward political out-group incivility is more likely to occur than non-congenial content directed toward in-group incivility, therefore:

> **H7:** More uncivil content posted on partisan Facebook pages will result in a greater number of Shares.

This dissertation treats incivility as a subset of negativity and, therefore, anticipates that the emotional reactions will be similar to those that result from negativity, especially Angry responses, the reaction that is most clearly connected to negative valence. Likewise, as proposed in the negativity section, Love may be seen as the opposite mechanism of negativity (i.e., positive) and, therefore, may be viewed as the opposite of incivility (i.e., civility), which is a subset of positivity. It is unclear how the other emotional Reactions to uncivil content work in the context of political content. Thus, the following research hypotheses/question are proposed:

> **H8a:** More uncivil content posted on partisan Facebook pages will result in a greater number of Angry reactions.

> **H8b:** More uncivil content posted on partisan Facebook pages will result in fewer Love reactions.

> **RQ2:** Does more uncivil content posted on partisan Facebook pages correlate with the number of Like, Sad, Haha, and Wow reactions?

Incivility, which is prevalent in political discourse, involves visceral reactions from the audience. For example, viewing incivility as "outrage," Sobieraj and Berry (2011) demonstrated that such outrage permeates mainstream and online media as expressed by insulting, name-calling, obscenity, or partisan language and emotional displays. Discourse from political talk shows tends to exhibit incivility and emotionally heated expressions that can lead to political polarization because of the tendency of human brains to pay more attention to conflict, according to evolutionary psychological perspectives (Mutz, 2006b). Incivility can also lead to decreased respect for the political positions of individuals with opposing views and intensify negative reactions to others based on television programs (Mutz, 2015). Moreover, exposure to incivility can lead to polarized views in public discourse (Hwang et al., 2014; Kim & Kim, 2019), increased affective reactions, and use of uncivil/critique comments in responses to an original post (Gervais, 2015) as well as greater tolerance toward in-group members and their comments (Kim, 2018).

Similar to negativity, it is assumed that partisan uncivil messages that are directed toward partisan out-groups are more likely to happen in political pages on Facebook and may induce polarizing effects. Both derogatory and category group labels provoke stereotypes (Carnaghi & Maass, 2007), and stereotyping the outgroup can lead to polarization. Also, the relationship between feelings toward the parties and uncivil partisan content on Facebook may be much clearer when an outgroup is displayed as a target compared to less explicit cases or when the ingroup only is a target. Therefore, the following hypotheses are proposed:

> **H9:** Aggregated affective polarization will correlate with measures of the incivility of content posted on partisan Facebook pages.

> **H10a:** Uncivil content posted on partisan Facebook pages that is targeted toward an out-group will correlate aggregated affective polarization.

When posts are directed to both partisan out-groups and in-groups in the context of incivility, the level of incivility is also potentially high, which may lead to polarization. These cases of out-group only mentions and in- and out-group mentions will be contrasted with cases where the in-group is the only group mentioned and when there are no partisan explicit targets. The following hypothesis is posed:

> **H10b:** Uncivil content posted on Facebook that is targeted toward both an out-group and in-group will correlate with aggregated affective polarization.

Negativity vs. Incivility

Given that incivility is seen as an unnecessary and derogatory form of negativity, the effects of engagement and affective reactions should be greater when people are exposed to uncivil content than to negative content in political discourse. Although it is the case that for politician accounts on Twitter, both negative and attacking messages are engaging, attacking messages have a greater impact on retweeting (Fine & Hunt, 2021). Therefore, I test the following hypotheses to compare negativity and incivility posted in Facebook partisan pages.

> **H11a:** Compared to the relationship between negative posts on partisan Facebook pages and the number of Comments, the relationship between uncivil posts on partisan Facebook pages and the number of Comments will be stronger in magnitude.

> **H11b:** Compared to the relationship between negative posts on partisan Facebook pages and the number of Shares, the relationship between uncivil posts on partisan Facebook pages and the number of Shares will be stronger in magnitude.

> **H11c:** Compared to the relationship between negative posts on partisan Facebook pages and a number of "Angry" Reactions on the content, the relationship between uncivil posts on partisan Facebook pages and the number of "Angry" Reactions will be stronger in magnitude.

It is not clear that how negativity and incivility will work with respect to the other emotional Reactions. This is examined through a research question:

> **RQ3:** Does the negativity or incivility of content posted on partisan Facebook pages correlate more strongly with the number of Like, Sad, Haha, Wow, and Love reactions?

Comparisons between incivility and negativity are also examined through political polarization (i.e., feelings toward both parties). Incivility in political content can cause intense and hyper-partisan reactions, which is from the inflammatory and harmful features. Due to stronger affective and partisan reactions for incivility than negativity, effects of uncivil content posted on partisan pages on affective polarization in the public will be greater than those of negative content. Therefore, the following hypothesis is proposed:

H12: Compared to the relationship between aggregated affective polarization and negativity of content posted on partisan Facebook pages, the relationship between affective polarization and incivility of content posted on partisan Facebook pages will be stronger in magnitude.

Partisan Differences

People may engage with and react to political content differently depending on their political leanings. Research demonstrates that conservatives have greater psychological and physiological tendencies to react to negative information and environments compared with liberals (Hibbing et al., 2014). For instance, conservatives are more likely to pay attention to negative stimuli of angry faces (Carraro et al., 2011) and to spend more time looking at negative images (Dodd et al., 2012) than liberals do. On the other hand, others argue that liberals and conservatives are similar rather than different in dealing with negativity and threats (Brandt et al., 2014). Based on suggestions by previous studies that there are differences according to political leanings with respect to how people engage with negative and uncivil content, the following research questions are proposed in the context of political content on Facebook.

RQ4: Are there partisan differences in the relationship between negative posts on partisan Facebook pages and user engagement metrics (Shares, Comments, Like, Angry, Sad, Haha, Wow, and Love reactions)?

RQ5: Are there partisan differences in the relationship between uncivil posts on partisan Facebook pages and user engagement metrics (Shares, Comments, Like, Angry, Sad, Haha, Wow, and Love reactions)?

It is also possible that the relationship between negative or uncivil posts and affective polarization differs by partisanship. Research finds that right-leaning posts on Facebook tend to include attacks on the opposing party more often than left-leaning posts (Sturm Wilkerson et al., 2021) and Republican candidates are more likely to use simple and direct speech than Democratic candidates (e.g., Jarvis, 2004; Tetlock, 1983), which possibly can lead to a greater influence of an out-group targeting content from Republican pages on affective polarization. On the other hand, Republican identifiers are less likely to use negative language in response to how they think about voting than Democrat identifiers (Jarvis & Jennings, 2017). The political entities mentioned in the posts also may yield different effects on polarization depending on one's partisanship. With less clear directions over partisan differences in the effects of political targets, the following research questions are explored:

> **RQ6:** Are there partisan differences in the relationship between negative posts on partisan Facebook pages and aggregated affective polarization by the four types of political targets (i.e., out-group only, in-group only, both in-/out-groups, and without explicit political entities)?

> **RQ7:** Are there partisan differences in the relationship between uncivil posts on partisan Facebook pages and aggregated affective polarization by the four types of political targets (i.e., out-group only, in-group only, both in-/out-groups, and without explicit political entities)?

THE CAUSAL EFFECTS OF NEGATIVITY AND INCIVILITY: RESEARCH HYPOTHESES AND QUESTIONS

I propose that negative and uncivil language in Facebook posts on partisan pages display partisan social identity and have effects at the individual level. This section focuses on partisan negativity and incivility as linguistic features that produce more engagement and have polarizing effects among individuals. This individual approach complements the aggregate data analysis strategy because it is better able to isolate causal effects.

Effects of Negativity and Incivility

To examine the polarizing effects of partisan language, two features of political language, its attributes (negative, uncivil, or neutral) and its target (in-group, out-group, or no target), are tested. The idea that partisan language use contributes to increased political polarization – prompting users to engage with the content and leading to greater polarization – is examined. In this individual level of study, I propose that targeting ingroups or outgroups in posts will garner more attention because they implicate people's sense of identity.

As noted previously, negativity and incivility attract more attention. For social metrics, exposure to negativity and incivility of political content invites engagement and generates affective reactions compared to exposure to neutral content, which increases chances of interactions with Comments, Shares, and Angry reactions as an expression of punishment against the opposing party or disapproval against the content. However, posts with presence of targets indicate pro-attitudinal (i.e., those targeting an outgroup with negativity or incivility) or counter-attitudinal posts (i.e., those targeting an ingroup with negativity or incivility), which will be controlled for to see effects of negativity and incivility on social engagement. People are more likely to interact with such content, which leads me to propose the following:

> **H13:** Holding the presence of political targets (i.e., no explicit, in-group, or out-group targets) constant, compared to political posts with neutral terms, political posts with negative or uncivil terms will receive more user engagement of (a) Comments, (b) Shares, and (c) Angry reactions.

Not only do I predict that these posts will influence engagement, I also propose that they will affect polarization. Based on the earlier discussion that selective exposure and incidental exposure both potentially leading to polarization, I propose the following hypothesis:

> **H14:** Holding the presence of political targets (i.e., no explicit, in-group, or out-group targets) constant, compared to neutral political posts, political posts with negative or uncivil terms will yield more affective polarization.

Effects of Political Targets

Also, the presence of a partisan target will motivate more engagement and maximize its engaging effects with partisan identity. Considering that ingroup targeting content is counter-attitudinal, outgroup targeting content is pro-attitudinal and non-explicit targeting content is a reference, social engagement metrics can show different patterns by the presence of political targets. For Commenting, both presence of ingroup (i.e., counter-attitudinal content) and outgroup targets (i.e., pro-attitudinal content) can attract engagement. Users on Facebook are more likely to interact with pro-attitudinal content (Pedersen et al., 2021) while exposure to cross-cutting content on Facebook motivates people to participate in the discussion due to the mediated anger (Lu & Myrick, 2016). Still, there are some chances that many users decide not to interact with counter-attitudinal content (Gearhart & Zhang, 2015). Therefore, I expect that compared to those with presence of non-explicit targets, posts with presence of outgroup or ingroup targets will invite more engagement. For Sharing, it is much clearer that posts with outgroup target presence will receive more engagement. Research finds that pro-attitudinal content is more engaging in sharing (An et al., 2014; Johnson et al., 2020; Shin & Thorson, 2017) than counter-attitudinal content. Furthermore, incidental exposure to counter-attitudinal content motivates people to seek and share like-minded content (Weeks, Lane, et al., 2017). Exposure to outgroup target posts will increase the likelihood of sharing than exposure to ingroup target posts. For Angry reactions, political posts with the presence of ingroup targets is engaging. Exposure to cross-cutting content generates emotions of anger (Lu & Myrick, 2016) and anxiety (Lu, 2019), which may lead to reaction of disapproval. Given the findings in research, I argue the following hypothesis:

> **H15:** Holding the language attributes of political posts (i.e., neutral, negative, uncivil content) constant, compared to non-target political posts, (a) political posts with ingroup or outgroup target will receive more user engagement of Comments, (b) political posts with outgroup target will garner more Shares, and (c) political posts with ingroup target will yield more Angry reactions.

Building on the literature showing polarizing effects from exposure to both pro- and counter-attitudinal information (for pro-attitudinal content, e.g., Beam et al., 2018; Conover et al., 2011, for counter-attitudinal content, e.g., Bail et al., 2018; Weeks, Lane, et al., 2017), I predict:

> **H16:** Holding the language attributes of political posts (i.e., neutral, negative, uncivil content) constant, compared to non-target political posts, political posts with ingroup or outgroup targets will yield more affective polarization.

Negativity vs. Incivility

As incivility may arouse greater affective reactions and generate more engagement, I predict that the relationships will be stronger for incivility than for negativity. This relationship is examined on engagement of Comments, Shares, and Angry reactions and affective polarization after controlling for the effects of political in-/out-group targets:

> **H17:** Controlling for the presence of political targets (i.e., in-group or out-group targets), the relationship between political posts and user engagement of (a) Comments, (b) Shares, and (c) Angry reactions will be stronger for incivility than for negativity.

> **H18:** Controlling for the presence of political targets (i.e., in-group or out-group targets), the relationship between political posts and affective polarization will be stronger for incivility than for negativity.

In-group vs. Out-group

Given previous discussions that both selective and incidental exposure lead to political polarization, it is not clear whether posts targeting a co- or opposing group will yield more engagement (i.e., Reacting, Sharing, and Commenting) or more polarization. An experiment will enable me to examine this relationship. One relevant study shows that those who are politically prejudiced are more likely to suppress hostile rhetoric toward their own party than to promote hostile rhetoric directed toward their opposing party (Lelkes & Westwood, 2017). This may indicate asymmetric preferences with regard to partisan language by which people tend to dislike negative and uncivil language toward an in-group more than they dislike identical language toward an out-group. Although exposure to negative and uncivil political content about out-groups is more likely to occur, there are chances of encountering in-group targeted negative and uncivil content because of diversity within an individual's social network. The language attributes of negativity and incivility will be held constant. Given the lack of clear theoretical guidance, I pose the following research questions:

RQ8: Are there differences in engagement with political posts targeted toward an in-group and those targeted toward out-group after controlling for the language attributes of political posts (i.e., negative or uncivil content)?

RQ9: Are there differences in affective polarization with political posts targeted toward in-group and those targeted toward out-group after controlling for the language attributes of political posts (i.e., negative or uncivil content)?

Chapter 3: Methods for Identifying the Effects of Language on Polarization

Computational social scientific methods open new angles on research questions, yield innovative ways of finding answers, and provide opportunities for inter-disciplinary perspectives. Their potential pays off when computational methods are guided by strong theoretical foundations and followed by careful validations. Computational methods can use pre-labeled data by millions of crowdsource workers who perform tasks paid or voluntarily, classification and prediction models that have outperformed human coders in various contexts, the aggregation of text data to understand the general tone of a text, and the extraction of meaningful information from complex and large datasets. Taking advantage of this method, I use computational methods combined with human guidance and validation to investigate the relationship between language and polarization.

This chapter introduces multiple methods to examine language as a driver of polarization at an aggregated and individual level. To do an aggregated level analysis, I use various approaches including content analysis by human coders, automated programs to identify millions of political posts' language attributes, and time series statistical models. For the individual level, I use an experimental setting which resembles a Facebook News Feed to examine the direct effect of language attributes from political posts on affective polarization. Details of the experimental design, research procedures, stimuli creation, and dependent variables are included in this chapter.

Identifying Negative and Uncivil Language on Partisan Facebook Pages Using Computational Social Scientific Methods

Several methods have been used to examine affective polarization, including large-scale surveys with representative samples (e.g., Haidt & Hetherington, 2012; Iyengar et al., 2012; Mason, 2015, 2016, 2018), implicit measures (e.g., Iyengar & Westwood, 2015), and experiments (e.g., Mason, 2016; Suhay et al., 2018). Although this dissertation does not examine each methodology's advantages and limitations in detail, I briefly review the methods on which my dissertation is primarily focused.

Data-driven linguistic features can provide strong evidence of political biases associated with intergroup relations. Partisan language has been a rapidly growing area of research for the past decade through the use of various techniques, including computational approaches (e.g., Grimmer & Stewart, 2013; Laver et al., 2003; Monroe et al., 2008; Yu et al., 2008). Although not specifically focused on political texts, language posted on Facebook can be used to distinguish personality, gender, and age (Schwartz et al., 2013). Accordingly, my dissertation applies several approaches for the purpose of investigating partisan and affective language use in political texts on Facebook. Examining the effects of political language extends findings from previous studies using surveys or implicit measurement because language addresses both explicit and implicit facets of affective intergroup attitudes. Political language involves not only what is stated (e.g., issue positions, ideology, and partisan identity), but also what is conveyed by the choice of words and how underlying values and beliefs are framed and rhetorically devised.

I use both time series analysis and an experiment assisted by computational methods to study how language influences social media engagement and affective polarization. With aggregate data, this research examines political language with machine learning and large text data, which allows me to analyze the underlying dynamics of polarization. Also, time series has merit for examining changes over time. To understand causality and the psychological process, I use an experiment. Ultimately, as mentioned in the introduction to this dissertation, the goal of this research is to examine whether language as expressed by individuals' social media use, here on Facebook, can influence people's affective evaluations of political groups.

Aggregate Level of Analysis: Methods

Identifying partisan pages on Facebook is the focal area of my dissertation. Content analysis and human assisted coding were applied to collect a substantial number of partisan pages on Facebook. Then, public opinion data on affective polarization was prepared for use in time series analysis. Validation for automated sentiment and incivility analysis in the context of partisan content on Facebook was applied through content analysis with human coders. Later, model selection and a strategy for social engagement and aggregated affective polarization analyses are discussed as well as introducing a control variable.

Identifying Political Pages on Facebook

To collect political pages on Facebook, I began with several seed sites that were located at the top of the search page on Google. Searching for the keywords "conservative pages on Facebook" and "liberal pages on Facebook" on Google, several sites were identified, such as "The Comical Conservative" (2.1M Likes, as of June 2020) for right-leaning pages and "Being Liberal" (1.6M Likes, as of June 2020) for left-leaning pages. These pages actively engaged their followers by posting political content, such as news, photographs, videos, and links. Next, similar sites that traffic in political content using the "Related Pages" and "Pages Liked by This Page," both featured on Facebook, were added. This search was repeated until a saturation point was reached (i.e., a substantial overlap of political pages was achieved from the separate lists of right- and left-leaning pages).[3] Among the collected pages, this list was narrowed to those pages that had more than three

[3] Pages that represent a verified political figure account (e.g., U.S. Senator Bernie Sanders) were removed from the list.

political posts within six months. This list consists of 64 pages from liberal sites (CrowdTangle Team, 2020a) and 64 pages from conservative sites (CrowdTangle Team, 2020b).[4]

Later, this list was supplemented with similar pages using the Search URL featured on CrowdTangle,[5] a content discovery tool and social analytics platform that is owned by Facebook. CrowdTangle surfaces public posts on Facebook that are created by public accounts with more than 100K Likes or verified Facebook profiles and public groups.[6] Using the list of political pages that I collected above, I randomly extracted 1,000 links among the link posts that were shared more than 10,000 times from the liberal and conservative pages. With those two sets of 1,000 links, I searched the URLs for additional partisan pages that shared the same links in the search feature on CrowdTangle. Through this process, I identified 9,557 pages with 1,000 links from conservative sites and 26,103 pages with 1,000 links from liberal sites. Among the total number of pages, those that were selected had the following characteristics: currently active (not broken pages), had at least three political posts for the past six months (between August 5, 2020 and February 5, 2021), had more than 10K Likes, and posted content in English. After removing duplicated pages, 6,831 unique pages shared the same links as the previously identified liberal sites, 2,010 unique pages shared the same links as the previously identified conservative sites, and 716 pages shared links from both liberal and conservative sites. I include the sites that share links from both liberal and

---

[4] At the end of November 2020, the list had 64 conservative and 64 liberal pages. However, as of December 28, 2020, one page (The New Revolution II, https://www.facebook.com/TheNewRevolutionII) from the conservative sites and one page (Political Humor, https://www.facebook.com/politicalhumor) from the liberal sites were deleted from the lists on Crowdtangle. The posts were saved before removal and they are included in the analysis.

[5] With support of Social Science One and a Data Use Agreement between Facebook and the University of Texas at Austin, access to CrowdTangle's data was made available for this current project. For more information, see https://www.facebook.com/facebookmedia/blog/crowdtangle-for-academics-and-researchers.

[6] CrowdTangle does not track private posts and profiles, paid or boosted posts, and several other metrics, such as page reach, traffic, or clicks on the posts and pages.

conservative sites to determine their political leanings by human coders. Details are discussed in a subsequent paragraph.

Because my dissertation is interested in partisan pages as political sources that create and spread hyper-partisan content, I decided not to include several types of pages: pages including personal information posted by verified individuals (e.g., pages using personal photo and/or name on their profile with a verified blue badge), verified public figures, government officials, and verified official news organizations. These types of pages can be highly political and/or partisan, but their content and impact on affective polarization may differ from that of partisan pages. My dissertation examines partisan pages that play a role of generating and providing widely shared political content that could contribute to political polarization.

To reach agreement regarding the pages included in my dissertation, three coders were trained in coding three categories with a subset of the codes (n = 300): page administrators' account locations and language of the posts, verified public figures/personal pages, and verified official news organizations. We looked at the Page Transparency section and three recent posts to determine whether the page's admin location was the United States (including mixed cases with other country locations, for instance, the United States and Singapore) and their posts were written in English (Krippendorff's alpha = 0.97). We then separately coded pages of verified public figures, individuals, and government officials (i.e., pages with a personal profile with a verified blue badge) (Krippendorff's alpha = 0.95) and pages of verified official news organizations (i.e., the page's categorization as a broadcasting media production, magazine, media news company, radio station, TV shows/channels, etc.) (Krippendorff's alpha = 0.84). When the page did not apply to cases of verified public figures, individuals, government officials, and official news organizations, we proceeded to code whether the page was political and the partisan leaning.

For independent coding to identify political pages and determine their political leanings, I took a random sample of 1,015 pages (692 pages that were identified by liberal links, 205 pages that were identified by conservative links, and 118 pages from both sides, a partisan imbalance proportional to the pages that were identified using the previously described methods). The three coders (two undergraduate research assistants and the author) determined whether a majority of posts on those pages were about politics (7 or more) from the most recent 10 posts. Political content could be shared news stories, images, or messages about political figures, parties, issues, and political actions. Content had to be explicitly tied to politics in order to be considered as political. Some pages mostly posting non-political content were excluded, although their topical categorization can be seen as political. For example, religious pages posting words from the bible or promoting local church, health related pages but mainly about meditation or yoga, and pages for announcement or updates from local police department including police members photos were among the excluded cases. The inter-coder reliability score for whether the page was political was strong (Krippendorff's alpha = 0.82). If the page was political, we coded the page's partisan leaning by looking at the About section (biographic information) and the 10 most recent posts. Categories were Republican or conservative (R), Democratic or liberal (D), and other or not sure (coded as 0). Content had to clearly support and/or lean toward one political party, or content had to clearly be against the opposing party. Inter-coder reliability for partisan leanings was acceptable (Krippendorff's alpha (nominal) = 0.80). Since we reached a reliable inter-coder agreement, the remaining pages (n = 8,542) were coded individually. A full codebook of identifying partisan pages on Facebook is in Appendix A.

It is worth noting that in the coding process and creation of the list of partisan pages in late March 2021, some pages were deleted or turned into private pages, temporarily or permanently. Those pages were not included in the final dataset (n = 260).

The final number of the political pages was 811 right-leaning and 682 left-leaning pages. From these pages, posts were scraped between January 1, 2008 and November 17, 2020 (right-leaning posts; n = 7,491,622, left-leaning posts; n = 6,978,592). Information about the 10 most popular sites with the largest number of page Likes is summarized in Table 3.1.

Table 3.1: 10 Most Popular Partisan Pages on Facebook (November 2020).

Leanings	Page Likes	Name	User ID	Page Created
D	8,773K	Occupy Democrats	/OccupyDemocrats	2012-09-21
D	6,700K	The Other 98%	/TheOther98	2010-04-03
D	2,288K	The Young Turks	/TheYoungTurks	2009-12-20
D	2,150K	Madam President	/MadamPresidentProject	2012-11-10
D	2,094K	Democrats	/electdemocrats	2008-02-28
D	1,805K	Being Liberal	/beingliberal.org	2009-11-05
D	1,708K	Women's Rights News	/WOMENSRIGHTSNEWS	2011-03-01
D	1,684K	Impeach Trump	/impeachtrumpasap	2017-01-11
D	1,642K	Democratic Party	/democrats	2008-04-17
D	1,607K	CAP Action	/AmericanProgressAction	2009-04-13
R	8,975K	IJR	/theijr	2009-12-05
R	7,896K	For America	/ForAmerica	2010-09-14
R	5,116K	The Western Journal	/WesternJournal	2009-08-25
R	5,060K	Freedom Works	/FreedomWorks	2007-11-08
R	5,027K	NRA Institute for Legislative Action	/NationalRifleAssociation	2008-01-06
R	4,393K	Donald Trump For President	/DonaldTrump4President	2015-08-07
R	4,367K	National Association for Gun Rights	/nagrfb	2009-07-30
R	4,182K	Conservative Tribune by WJ	/theconservativetribune	2013-10-08
R	3,273K	Donald Trump Is Our President	/DTIOUR45P	2016-08-13
R	3,151K	Conservative News Today	/ConservativeNewsToday	2011-01-08

Among the items that CrowdTangle provided, the most important fields for this dissertation are Message, Link Text, Image Text, and Description of postings on political pages. The combined texts were used in order to examine language usage. According to the CrowdTangle codebook (Garmur et al., 2019), Message captures “the blurb of the post, written when the post is uploaded” and Link Text presents “the headline of a link URL or the title of a native video” (e.g., a news article title). Image Text includes the OCR (optical character recognition) image text extraction of the post, as available on CrowdTangle. Description presents a few lead sentences of articles shared from a link. Posts that did not have any text (e.g., shared photo or video links, only emojis without any text messages) were not included in the analysis (right-leaning posts n = 140,984; left-leaning posts n = 168,568).

The final number of posts used for analysis is 7,350,638 from the conservative pages and 6,810,024 from the liberal pages.

Dependent Variables

Engagement Metrics. Reactions (e.g., Like), Comments, and Shares are metrics associated with each post collected between January 1, 2008 and November 17, 2020 from CrowdTangle. Analysis of the other Reactions beyond Like (Angry, Love, Haha, Wow, and Sad) is limited to after February 23, 2016 when they were introduced (Krug, 2016), which is 6,105,709 from the conservative pages and 5,046,175 from the liberal pages. The final number of Reactions after removing posts that were not able to measure for sentiment or incivility analysis[7] was 10,991,537 (6,079,177 from the conservative pages and 4,912,360 from the liberal pages). For Comments and

[7] Posts with empty messages or non-textual data only (e.g., emojis) and in non-English languages are not included (from the conservative pages, n = 26,532; from the liberal pages, n = 133,815).

Shares, after removing posts that were not able to measure for sentiment or incivility analysis[8], 13,955,869 posts were used for analysis (7,317,844 from the conservative pages and 6,638,025 from the liberal pages).

Public Opinion Data for Affective Polarization. My dissertation research used two surveys to capture aggregated affective polarization: (1) the NBC News/Wall Street Journal Poll (NWSJ, hereafter) examined how people felt toward the Republican and the Democratic Parties with response options very positive, somewhat positive, neutral, somewhat negative, or very negative (positive or negative feeling, hereafter) as initiated since October 26, 1990, and (2) the Democracy Corps Survey (DCS, hereafter), that examined how people felt toward the Republican and the Democratic Parties, with a score of 100 meaning a very warm, favorable feeling, a score of 0 meaning a very cold unfavorable feeling, and a score of 50 meaning not particularly warm or cold (feeling thermometers, hereafter) initiated since October 1, 1999. The survey data were collected from searches of the iPOLL Databank and provided by the Roper Center for Public Opinion Research. Using two separate datasets allows for an understanding of whether observed relationships between language and affective polarization are conditional on the survey used.

Polls from NWSJ between January 2008 and October 2020 (k = 97 studies) were examined corresponding to the Facebook post data that I collected. The NWSJ's survey response items were aggregated as percentages: positive including both very positive and somewhat positive (toward Democratic Party, M = 38.6; toward Republican Party, M = 39.8), negative including both very negative and somewhat negative (toward Democratic Party, M = 39.4; toward Republican Party, M = 44.8), and neutral (toward Democratic Party, M = 20.7; toward Republican Party, M = 23.0).

[8] Likewise, those unmeasurable posts are removed (from the conservative pages, n = 32,794; from the liberal pages, n = 171,999).

The neutral category was not included for analysis. The trends of feelings toward the Democratic and Republican Parties are displayed in Figure 3.1.1.

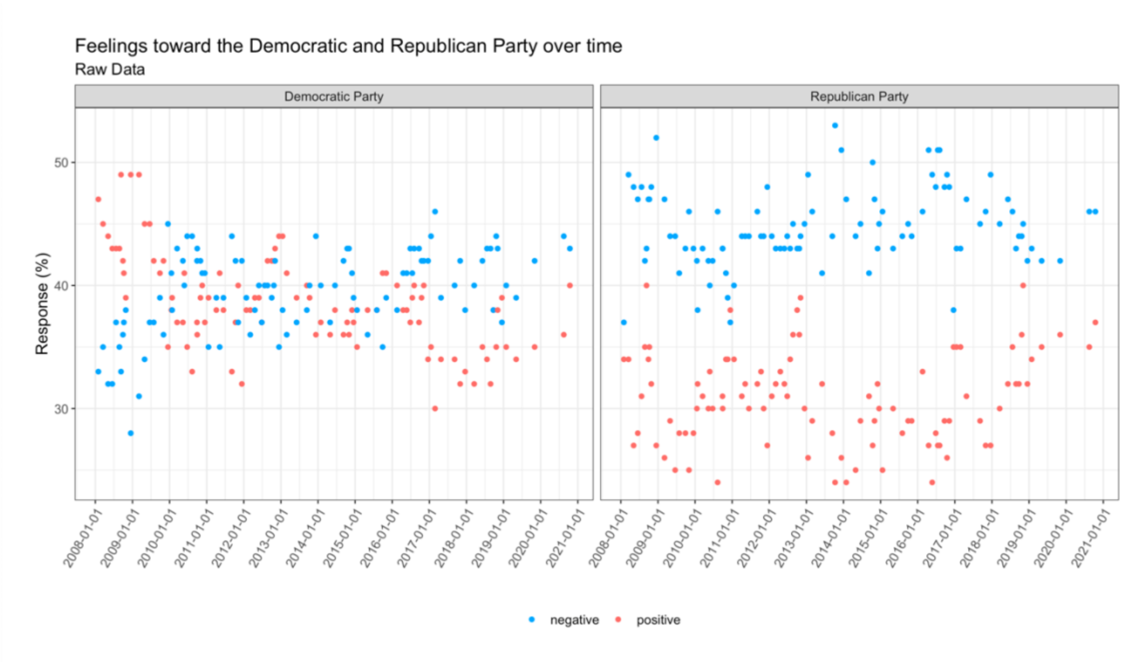


Figure 3.1.1: Feelings of the Democratic and Republican Party Over Time (Raw NWS Data).

Polls from DCS between January 2008 and September 2019 were investigated (*k* = 96 studies).[9] The DCS responses consist of four aggregated categories with percentages of 51-100 warm (toward Democratic Party, *M* = 40.7; toward Republican Party, *M* = 33.7) , 0-49 cool (toward Democratic Party, *M* = 42.6; toward Republican Party, *M* = 46.1), 50 not particularly warm or cold (toward Democratic Party, *M* = 14.6; toward Republican Party, *M* = 18.2), and never heard of/don't know/refused (deleted due to missing cases). The categories of the score of 50, not particularly

[9] I excluded one study in February 2016 as an outlier because time series analysis is highly vulnerable to outliers and response to the Democratic Party was unusually different from other studies. In February, 2016, the recorded responses of how people rate their feelings toward the Democratic Party were '0-49 Cold' as 90 %,'51-100 warm' as 1 %, '50 Not particularly warm or cold' as 8%, and Never heard of/Don't know/Refused as 1% (Mean 15.3) (although response of how they rate feelings toward the Republican Party were '0-49 Cold' as 17%, '51-100 warm' as 63%, '50 Not particularly warm or cold' as 19%, and Never heard of/Don't know/Refused as 1% (Mean 62.9) which does not appear to be an outlier). Compared to other studies with the same response categories ranging 24% to 59%, I took out the study in February 2016 as an outlier.

warm or cold, were excluded for analysis. The trends of feeling thermometers for the Democratic and Republican party are illustrated in Figure 3.1.2.

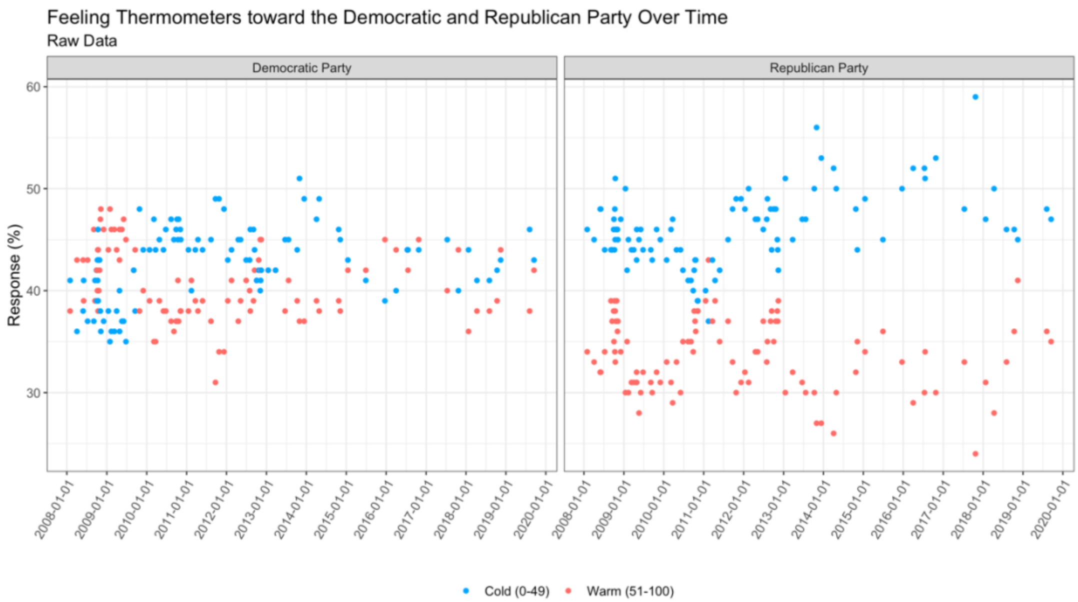


Figure 3.1.2: Feeling Thermometers of the Democratic and Republican Party Over Time (Raw DCS Data).

Because public opinion data from NWSJ and DCS are time series data with irregular time intervals, I used several methods to make the intervals regular. I aggregated the public opinion data by each quarter within a year on average. Although not a perfectly regular time series, as I describe shortly, it enabled me to apply a time series analysis in a regular format by considering its characteristics, such as autocorrelation and trend.

After quarterly aggregation, there were some missing data from the NWSJ (5%) and DCS (17%) surveys. To find a proper way to interpolate missing data, I created interpolation values using several methods from the imputeTS library in R (Moritz & Bartz-Beielstein, 2017) as follows: linear interpolation, Kalman Smoothing on ARIMA models, Last Observation Carried Forward (LOCF), and moving average. As reference data to compare with the four interpolation methods, I also created values that are predicted by a univariate forecasting model using the ETS function (exponential smoothing state space model) from the fable library in R with available time series data. For instance, when a study in Q1 2014 was missing, I used data from the beginning of Q4 2013 to predict the missing value in Q1 2014. Later, this predicted value was included in the dataset to predict the next missing value. I repeated this process until all the missing values were determined. Although this is not a perfect way to deal with missing values in a time series data, I implemented the four interpolation methods and one predicted value by the ETS (exponential smoothing) model and chose the values based on the performance of interpolation methods in comparison with the ETS model. ETS model is also commonly used algorithms for forecasting time series same as ARIMA models (Hyndman et al., 2002). In other words, the ETS forecasting predicted values were used for validation by comparing errors with the interpolated values. I chose the best interpolation method for each variable. To be specific, LOCF performed best for two DCS feeling thermometer variables (warm toward Democratic Party and Republican Party) and two NWSJ sentiment variables (negative/positive feeling to Democratic Party). Moving averages outperformed the others for two DCS feeling thermometer variables (cold feeling thermometer toward Democratic Party and Republican Party) and Kalman smoothing showed the best performance for two NWSJ variables (negative/positive feeling to Republican Party). A summary of validation metrics is presented in Table 3.2. Figure 3.2.1 and Figure 3.2.2 depict the interpolated data.

Table 3.2: Validation Metrics between Interpolated and Predicted Values.

Study	Variable	RMSE	R^2	Best Model
DCS	Feeling thermometers Democratic Party (cold, 0-49)	0.779	0.930	MA
DCS	Feeling thermometers Democratic Party (warm, 51-100)	0.662	0.957	LOCF
DCS	Feeling thermometers Republican Party (cold, 0-49)	0.339	0.992	MA
DCS	Feeling thermometers Republican Party (warm, 51-100)	0.367	0.989	LOCF

NWSJ	Negative feeling toward Democratic Party	0.522	0.970	LOCF
NWSJ	Positive feeling toward Democratic Party	0.048	1.000	LOCF
NWSJ	Negative feeling toward Republican Party	0.134	0.997	Kalman
NWSJ	Positive feeling toward Republican Party	0.251	0.995	Kalman

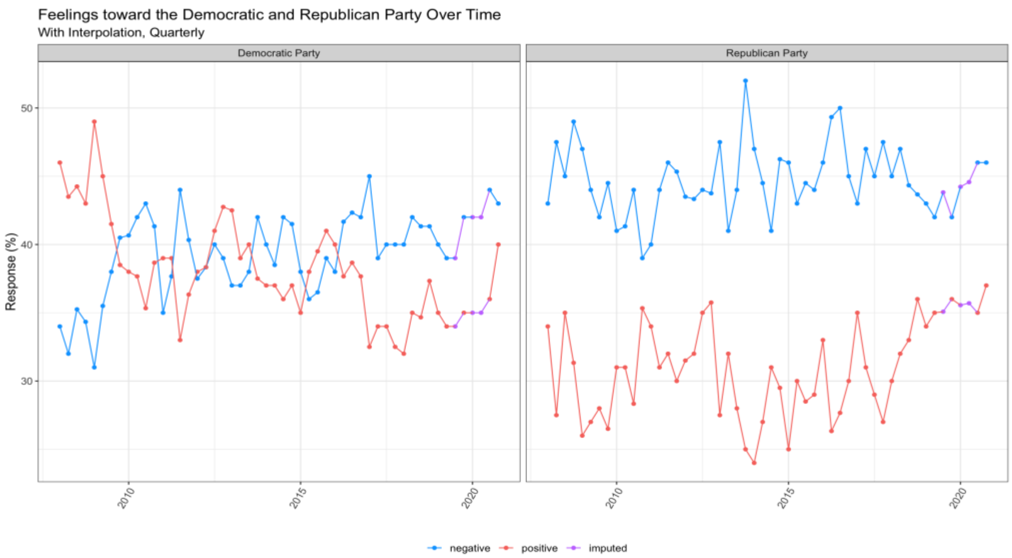


Figure 3.2.1: Feelings toward the Democratic and Republican Party Over Time (DCS), Quarterly with Interpolation.

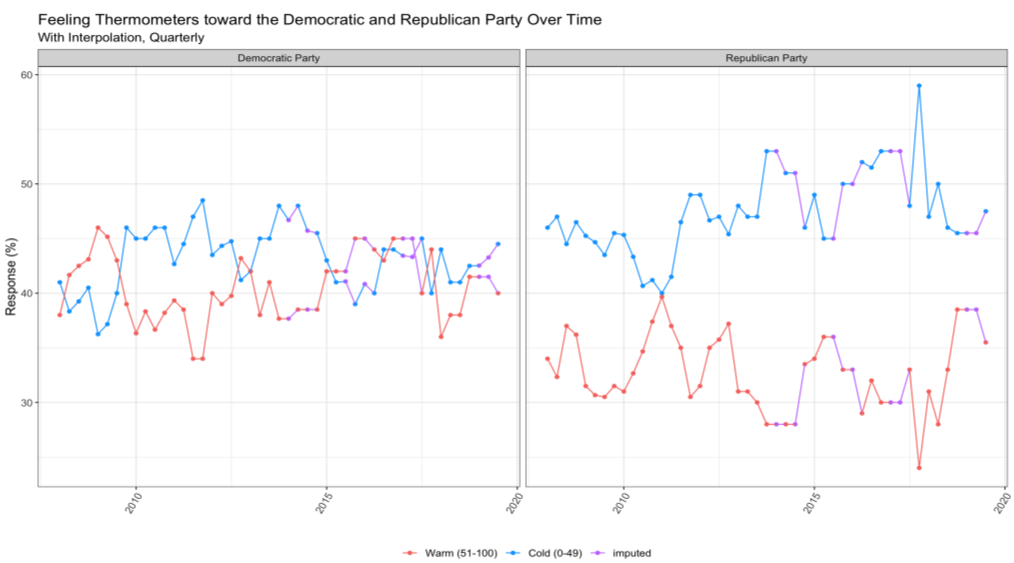


Figure 3.2.2: Feeling Thermometers of the Democratic and Republican Party Over Time (NWSJ), Quarterly with Interpolation.

Finally, I computed two measures of affective polarization: relative and absolute. For relative sentiment, I examined the difference between net sentiment toward the Democratic Party and net sentiment toward the Republican party (the net sentiment obtained by subtracting negative sentiment from positive sentiment for the Democratic [Republican] Party). For relative feeling thermometer, I used the difference between the net feeling thermometer toward the Democratic Party and the net feeling thermometer toward the Republican Party (the net feeling thermometer obtained by subtracting the cold thermometer from the warm thermometer for the Democratic [Republican] Party). The relative distance between the net feelings toward the two parties for affective polarization indicates less polarization if the score is close to 0 while more polarization exists if the score is away from 0. If the score is positive, it is more favorable toward the Republican Party in general and if the score is negative, it is more favorable toward the Democratic Party overall. Those relative measures are to see the direction of affective polarization by partisanship.

For instance, net sentiment toward the Democratic Party is 12 in Q1 of 2008, obtained by subtracting negative sentiment toward the Democratic Party (34% of the public expressing negative sentiment) from the positive sentiment toward the Democratic Party (46% of the public expressing positive sentiment) and the net sentiment toward the Republican Party is -9 by subtracting the negative sentiment toward the Republican Party (43%) from the positive sentiment toward the Republican Party (34%). The relative sentiment in 2008 Q1 is -21 by subtracting the net sentiment toward the Democratic Party (12) from the net sentiment toward the Republican Party (-9).

Likewise, I computed two measures of the absolute value of affective polarization: 1) partisan absolute sentiment measure, or the absolute value of the difference between the two parties' net sentiment (i.e., an absolute value of the relative partisan sentiment) and 2) partisan absolute feeling thermometer measure, or an absolute value of the difference between the two parties' net feeling thermometer (i.e., an absolute value of the relative partisan feeling thermometer). For some of the aggregate analyses, I chose the absolute value of affective polarization to see if polarization exists regardless of its directions by partisanship, with higher values indicating greater polarization.

Independent Variables

Negativity of Posts on Partisan Facebook Pages. Sentiment analysis (i.e., identifying the polarity – positive or negative – of a given text) has become a widely used tool that examines text, particularly online and social media text. Ribeiro et al. (2016) presented a benchmark comparison of twenty-four sentiment analysis methods, which are popularly used as "off-the-shelf" tools in practice, with eighteen labeled datasets that cover movie/product reviews, messages posted on social media, and comments in news articles. They found that the performance of each tool varies across datasets, for example, by the source of the dataset (e.g., whether it is from Twitter or comments from news articles), the number of label classifications (e.g., two label classifications as positive and negative or three label classifications, including neutrality), and the appearance of emoticons. Although not universal, several methods perform better, such as VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto & Gilbert, 2014), LIWC (Linguistic Inquiry and Word Count) (Pennebaker et al., 2015), and SentiStrength (Thelwall, 2017). Those methods have been applied and validated in various political contexts, such as predicting election results by using the sentiment of tweets associated with a politician as analyzed by LIWC (Tumasjan et al., 2011), measuring the sentiment of political conversations about President Trump on Twitter using VADER (Shugars & Beauchamp, 2019), and capturing the sentiment of tweets about the 2016 presidential candidates generated by bots and humans using SentiStrength (Bessi & Ferrara, 2016).

For my dissertation, I analyze these three (LIWC, VADER, and SentiStrength) in addition to LSD (Lexicoder Sentiment Dictionary) (Young & Soroka, 2012), which has been validated in political discourse such as news articles from the *New York Times* and the *Washington Post* (Soroka et al., 2015) and tweets about elections across the U.K., U.S., and Italy (Murthy, 2015).

Using these methods, this study compared the performance of a random subset of human-coded data (1,000 Facebook posts from the partisan pages identified for this project) and selected the method that most correctly evaluated negativity on Facebook posts. As there is no single method that is applicable for all studies (Wolpert & Macready, 1997), this validation process was to gain better interpretations of the results given the particular Facebook posts used in this dissertation.

To produce manually coded negativity in a political text, three coders (two undergraduate research assistants and the author) evaluated a random sample of 1,000 political posts (500 each from both liberal and conservative pages). We evaluated the posts' valence (positive or negative affect) and the emotions contained within the text (positive – e.g., joy, affection, pride, nostalgia, funny; negative – e.g., sadness, anger, fear, disgust, remorse, anxiety) in order to categorize the post as positive, negative, or neutral (see the codebook of negativity in the Appendix B for details). The inter-coder reliability was strong (as a three-value classification, Krippendorff's alpha (nominal) = 0.81; Krippendorff's alpha (ordinal) = 0.86; as a binary classification, for positive sentiment, Krippendorff's alpha = 0.83; for neutral sentiment, Krippendorff's alpha = 0.74; for negative sentiment, Krippendorff's alpha = 0.84). In creating a dataset for comparison with the automated methods, I excluded cases in which the three coders were not able to reach agreement after discussion of the results (e.g., posts that need further context to interpret or can be interpreted differently, n = 7) and posts that contained duplicated text (e.g., repeated only the page name or

"Timeline Photo," n = 16). This left 977 posts for comparison with the automated measures (positive codes n = 202, neutral codes n = 154, and negative codes n = 621).

I computed the performance scores Precision, Recall, and *F*1 for the automated classification as presented Table 3.3. Because the four different measures of automated sentiment analysis in my dissertation generated a continuous score as a unit of analysis (LIWC, min = -50, max = 50; LSD, min = -0.5, max = 0.5; VADER, min = -1, max = 1; and SentiStrength, min = -4, max = 4), and my focus in this section is to identify negativity of the text, I created a binary variable that captures 1 if the score is lower than 0 (negativity as a target classification) and 0 if the score is higher than 0 (positivity as a reference classification). For comparison, Table 3.3 summarizes the automated classification and its performance based on the human coding in three cases: (a) using all human codes including neutral outputs as positive cases (n = 977), (b) using all human codes excluding neutral cases (n = 823), and (c) using codes only with agreement among the three coders and excluding neutral cases (n = 720).[10] Looking across, SentiStrength showed the best classification performance for negativity. SentiStrength is a continuous measure ranging from -4 (strongly negative) to 4 (strongly positive). The SentiStrength score was centered and standardized for analysis ($M = 0$, $SD = 1$, min = -2.64, max = 3.30).

Table 3.3: Comparison of Performance by Automated Sentiment Analysis with Human Codes.

Human Code Classification	Measure	Accuracy	Precision	Recall	*F*1	Coverage[11]
(a) All codes, n = 977	LIWC	0.69	0.87	0.64	0.74	0.69

[10]Adding neutral codes from human coders as positivity shows consistent findings shown as Table 3.3.
[11] Because negativity is a binary classification (i.e., negativity or not), I added a metric of coverage that measures the percentage of posts that the automated sentiment analysis computes as either positive or negative excluding neutrality (scored as 0) (Ribeiro et al., 2016). The coverage score is affected by the original scales of automated sentiment analysis methods. The original scales of LIWC and SentiStrength are discrete integers (e.g., -50 or 4) that contains 0, which makes a coverage score low because it excludes more posts scored by 0. On the other hand, those of LSD and VADER are continuous decimals (e.g., -0.341 or 0.142) that are less likely to have an exact zero scored post, which makes a coverage score high because it excludes a small number of zeros.

- agreed by 2 or more	LSD	0.71	0.81	0.74	0.77	0.81
- neutral as positive	VADER	0.70	0.84	0.68	0.75	0.89
- 356 as positive, 621 as negative cases	SentiStrength	0.77	0.83	0.83	0.83	0.67
(b) Non-neutral codes,	LIWC	0.69	0.92	0.64	0.76	0.69
n = 823	LSD	0.73	0.88	0.74	0.81	0.81
- agreed by 2 or more - neutral excluded	VADER	0.72	0.93	0.68	0.78	0.89
- 202 as positive, 621 as negative cases	SentiStrength	0.81	0.91	0.83	0.87	0.67
(c) Non-neutral codes	LIWC	0.70	0.93	0.66	0.77	0.71
& coder agreement,	LSD	0.75	0.91	0.75	0.82	0.81
n = 720 - agreed by 3 coders	VADER	0.72	0.94	0.69	0.80	0.89
- neutral excluded - 161 as positive, 559 as negative cases	SentiStrength	0.83	0.93	0.84	0.88	0.68

Aggregated sentiment, which is the arithmetic mean of all posts' tone by each quarter within a year, was assessed and is presented in Figure 3.3.1. As shown, Facebook posts became increasingly negative over time.

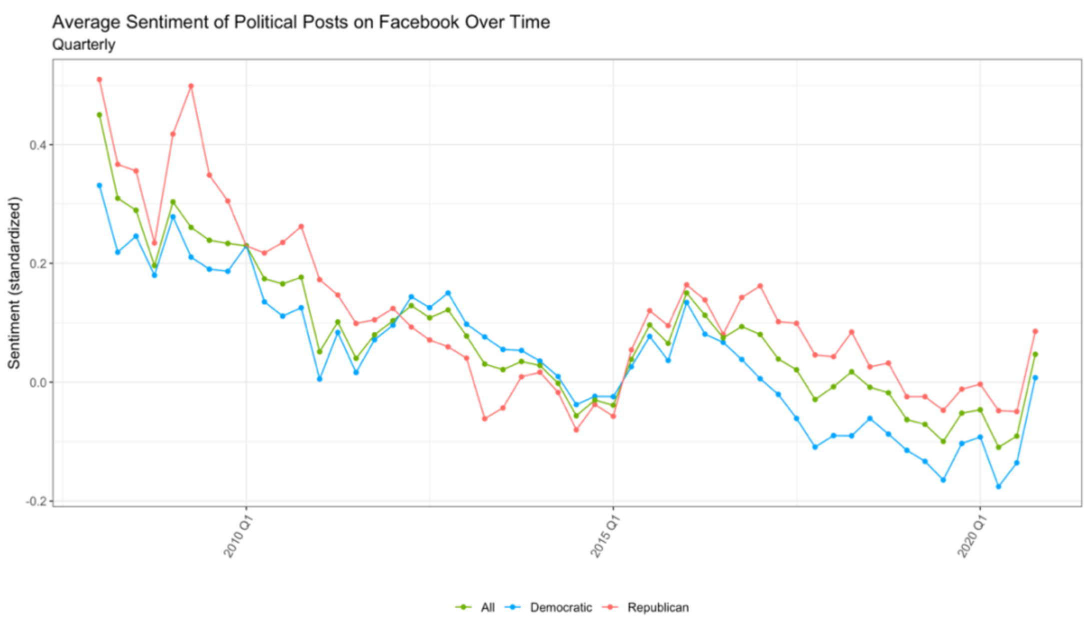


Figure 3.3.1: Average Sentiment (Negativity below than the Zero) of Political Posts on Facebook Over Time (Aggregated Quarterly).

Incivility of Posts on Partisan Facebook Pages. Using the Google Perspective API (which is based on a supervised machine learning model with various online text datasets),[12] this study tested Toxicity, Identity Attack, Insult, and Likely to Reject, each of which capture a post's level of incivility, continuously ranging from 0 to 1. According to the description of the Perspective API (Conversation AI, 2017), Toxicity means a "rude, disrespectful, or unreasonable comment that is likely to make you leave a discussion," Identity Attack refers to "negative or hateful comments targeting someone because of their identity," Insult indicates an "insulting, inflammatory, or negative comment towards a person or a group of people," and Reject captures the likelihood of a comment being rejected based on the standards of *New York Times* journalists. Similar to the assessment of the negativity measures, this study compared the four Perspective API variables with a subset of human-coded samples. Because the incivility measures are continuous, this process was also for determining a threshold for incivility compared to human codes, which used a binary scale. Again, this study is not intended to validate the algorithms, rather its purpose is achieving a better understanding of posts from partisan political pages on Facebook. In previous studies, the toxicity score showed comparable performance when evaluated against validated organic dictionary methods (Muddiman et al., 2019) and performed well in a social media context, such as YouTube comments (e.g., Obadimu et al., 2019) and Twitter (e.g., Georgakopoulos et al., 2019).

Because this dissertation considers incivility as a subset of negativity, I used 621 posts that were coded as negative from the manual coding procedure and assessed incivility (276 posts from liberal sites; 283 posts from conservative sites). With two coders (an undergraduate researcher and

[12] The Perspective was created by Jigsaw and Google's Counter Abuse Technology team for a research project Conversation-AI. The Perspective API is trained by public online text datasets, such as Wikipedia, using machine learning models and crowd human coders. For details, see https://www.perspectiveapi.com and https://conversationai.github.io.

myself), the incivility code was reliable (Krippendorff's alpha = 0.81). For better comparisons with the automated incivility measures, I only included codes that achieved agreement between two coders for validation (n = 514). The codebook for incivility is in Appendix B.

Because the four measures of incivility (Toxicity, Identity Attack, Insult, and Likely to Reject) are continuous scores between 0 and 1, thresholds of classification were applied (e.g., if the incivility score was higher than a threshold of 0.1, posts above 0.1 were classified as uncivil, and posts below 0.1 as uncivil). The performance of the measures at various thresholds are shown in Figure 3.3.2.[13] Insult demonstrated the best performance at a threshold of 0.2 (accuracy: 0.73; precision: 0.78; recall: 0.82; *F*1: 0.80) as displayed in Table 3.4. The Insult score is centered and standardized for analysis (*M* = 0, *SD* = 1, min = -1.08, max = 3.50).

Table 3.4: Performance of Incivility Measures with Human Codes.

Measures	Accuracy	Precision	Recall	*F*1
Insult	0.73	0.78	0.82	0.80
Toxicity	0.68	0.70	0.87	0.78
Reject	0.64	0.64	0.96	0.77
Identity Attack	0.59	0.78	0.49	0.60

[13] Results of accuracy and *F*1 are displayed only because *F*1 combines both recall and precision results.

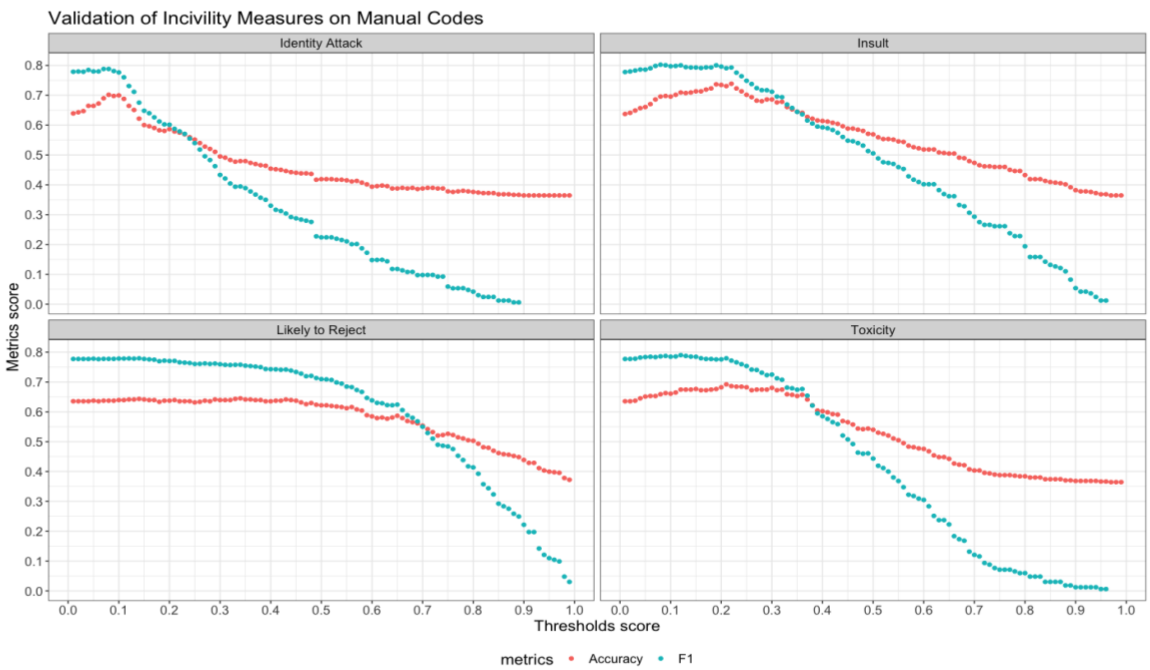


Figure 3.3.2: Comparison of Incivility Measures on Human Coding.

The aggregated tone of incivility, the arithmetic mean of the posts' toxicity for each quarter in a year, is shown in Figure 3.3.3.

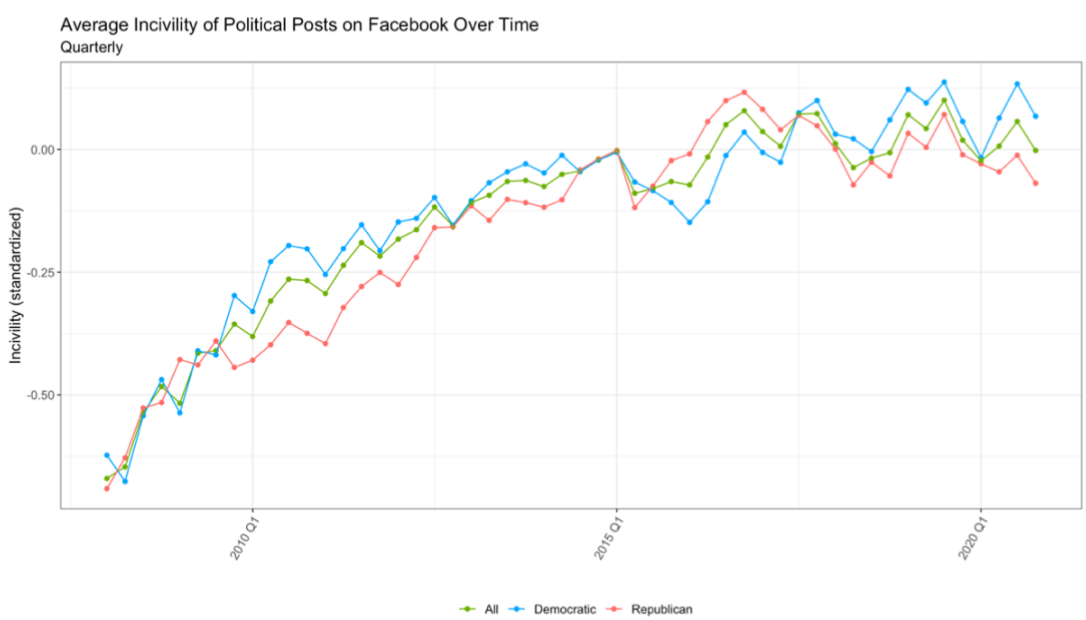


Figure 3.3.3: Average Incivility of Political Posts on Facebook Over Time (Aggregated Quarterly).

Partisan In- and Out-group Targets. By applying named entity recognition (NER) and part-of-speech (POS) tagging for posts in partisan pages, this dissertation identified political figures or organizations mentioned in the posts. After gathering entity information from all partisan posts through NER from the spaCy library on Python, I chose three labels related to political figures and organizations (PERSON, People, including fictional; ORG, Companies, agencies, institutions, etc.; NORP, Nationalities or religious or political groups). Additionally, I applied POS tagging for political figures and organizations that might not be identified from the NER process. I selected bigrams (a pair of consecutive word units) consisting of possible combinations of three part-of-speech tagging (ADJ: adjective; NOUN: noun; PROPN: proper noun). For example, Donald Trump is a bigram tagged as a proper noun and a proper noun, right-wing media as a noun and a noun, and 45th president as an adjective and a noun. Note that I set hyphenated words as an adjective and some special characters such as hashtags were considered as a noun. List of words and phrases from NER and POS procedures were created to identify additional political entities and their names from the partisan pages.

To validate the process, two human coders (an undergraduate research assistant and myself) coded several sample entities and evaluated reliability. I randomly selected 1,000 posts and extracted political figures and organizations by applying NER. After removing some non-related words (e.g., emojis only or URLs) and duplicated entities (because some entities are assigned more than one label), 1,100 identified entities were manually examined. With a selected and modified codebook of people and organizations (Muddiman et al., 2020) that guides categorizing political entities from Google or Wikipedia search, the coders had reliable categorizations for whether the entities were politicians/government officials (Krippendorff's alpha = 0.92), Democratic Party/liberals (Krippendorff's alpha = 0.81), Republican

Party/conservatives (Krippendorff's alpha = 0.85) or news/media (including related people) (Krippendorff's alpha = 0.84).

For news/media entities that did not show clear political leanings from Google or Wikipedia searches, I added political leanings (i.e., categorizing media outlets as conservative or liberal leanings) from previous research that identifies political leanings/slant of news and media organizations (Budak et al., 2016; Gentzkow & Shapiro, 2010) and popular sources that partisans share (Bakshy et al, 2015). Related to this, I applied journalists' political leanings from the news/media outlet's political leaning where they work. For instance, those who work for Fox News were considered conservatives and those who work for CNN or MSNBC were seen as liberals (e.g., journalists, news hosts, or political commentators).

It is worth noting that there were some exceptions for political entities. I did not include "Robert Mueller" as a Republican in the analysis (although Mueller is a registered Republican from the Google and Wikipedia searches) because his name appeared in the context of the investigation of Russian interference into the 2016 presidential election rather than as a political target. Also, Michael Flynn was coded as a Republican although he is a registered Democrat from the Google and Wikipedia searches.

NER for all political posts with the three classifications (person, organizations, and political groups etc.) identified 17,581 entities from both political pages after removing some text that was unable to be categorized (e.g., emojis only or unidentifiable entities without context such as common first names). Some named entities had several names (e.g., Donald Trump have Donald J. Trump, @realDonaldTrump, #Trump2020, and Mr. Trump, etc.) and different formats from the original text (e.g., Donald TRUMP, DONALD J. TRUMP, and tRump). For coding, I kept all names and different formats to make identification of entities with clarity. Later, I removed special characters such as punctuation and quotation marks to match the content in political posts that is also pro-processed with the same procedure. The number of identified Republican and conservative entities from NER is 1,659 and the number of identified Democratic and liberal entities is 1,465.

Also, to identify cases where NER did not recognize words and phrases as political groups or entities (e.g., right-wing and left-wing), words with POS tagging as adjective, noun, and proper noun and phrases with combinations of POS tagging as an adjective, noun, and proper noun were identified. Then, after removing special characters of the identified words and phrases, I chose 80,671 words and phrases with more than 300 occurrences across all partisan pages. After removing most of common noun terms with which clearly are not politically associated, names that are not necessarily associated with public figures (e.g., common first names), and terms that were already identified from the NER process, the rest of words and phrases (n = 30,313) were reviewed with the Google and Wikipedia search to determine its political leanings. Note that this is not a unique number of political entities as some entities have multiple names or expressions. From this POS tagged phrases, 2,839 Democratic and liberal terms and 2,600 Republican and conservative terms were identified. After removing special characters and excluding duplicates from the political entities from the NER processing, 3,203 terms indicating Democratic and liberal leanings and 2,784 terms indicating Republican and conservative terms were identified to match the content on political pages.

After matching identified political entities with the posts by posts' political leanings, I created four categories of aggregation: in-group entities only (i.e., Democratic/liberal [Republican/conservative] entities that appeared in Democratic/liberal [Republican/conservative] posts), out-group entities only (i.e., Democratic/liberal [Republican/conservative] entities that appeared in Republican/conservative [Democratic/liberal entities] posts), no explicit political entities (neither in-group nor out-group entities from NER or POS procedures mentioned in the same post), and both in-group and out-group entities (both in- and out-group entities identified from NER and POS procedures are mentioned in the same post). A summary of the percentage of posts with political entities is presented in Table 3.5. More Democratic/liberal posts named out-group entities only (27.3%) than did Republican/conservative posts (17.2%) while more Republican/conservative posts included in-group entities only (21.8%) than did Democratic/liberal posts (16.5%). For modeling, negative and uncivil language with partisan targets varying out-group entities only, in-group entities only, no explicit entities, and both in- and out-group entities were aggregated by all political posts, Democratic/liberal posts only, and Republican/conservative posts only.

Table 3.5: Percentage of Posts by Presence of Political Entities.

Political Entities	Leanings	%
Out-group entities only	Democratic/Liberal	27.3
Out-group entities only	Republican/Conservative	17.2
In-group entities only	Democratic/Liberal	16.5
In-group entities only	Republican/Conservative	21.8
No explicitly partisan entities	Democratic/Liberal	37.1
No explicitly partisan entities	Republican/Conservative	42.0
Both in- and out-group entities	Democratic/Liberal	19.0
Both in- and out-group entities	Republican/Conservative	19.0

Analysis

Model Selection for Engagement Metrics. In order to test the relationship between the level of user engagement on Facebook and posts' negativity or incivility, I applied zero-inflated negative binomial mixed regression modeling for user engagement metrics because the engagement metrics in the dataset were over-dispersed (i.e., the variance was greater than the mean) and had a large portion of zeros (i.e., posts that did not receive any engagement) as shown in Table 3.6. Thus, the zero-inflated model has an advantageous fit. The zero-inflated model provides a separate estimation of fixed effects (explaining "sampling" zeros) and zero-inflated part (explaining "structural" zeros) (Hu et al., 2011). For instance, although the fixed effects estimation model predicts observations of Angry counts including some zero reactions because some posts could have received only a few Angry reactions when, in fact, the Angry reactions happened to have zero reactions (i.e., sampling zeros). In the zero-inflated part, the zero-inflation binary model fits the probability that the Angry reaction is not zero because some posts receive only zero reactions due to low visibility or low popularity of the page (i.e., structural zeros).

Table 3.6: Rates of Zeros in Reaction Metrics.

Reactions	Leanings	Zero Rate (%)	Date start	Date end
Like	Democratic/Liberal	5.5	1/5/08	11/17/20
Like	Republican/Conservative	7.8	1/9/08	11/17/20
Comments	Democratic/Liberal	24.0	1/5/08	11/17/20
Comments	Republican/Conservative	23.7	1/9/08	11/17/20
Shares	Democratic/Liberal	19.5	1/5/08	11/17/20
Shares	Republican/Conservative	19.2	1/9/08	11/17/20
Angry	Democratic/Liberal	42.0	2/23/16	11/17/20
Angry	Republican/Conservative	49.1	2/23/16	11/17/20
Love	Democratic/Liberal	39.3	2/23/16	11/17/20
Love	Republican/Conservative	46.4	2/23/16	11/17/20
Sad	Democratic/Liberal	49.8	2/23/16	11/17/20
Sad	Republican/Conservative	60.9	2/23/16	11/17/20
Wow	Democratic/Liberal	43.4	2/23/16	11/17/20
Wow	Republican/Conservative	47.7	2/23/16	11/17/20
Haha	Democratic/Liberal	45.0	2/23/16	11/17/20
Haha	Republican/Conservative	44.8	2/23/16	11/17/20

To implement the zero-inflated negative binominal mixed modeling, I used two R libraries: the GLMMadaptive library (Generalized Linear Mixed Model using Adaptive Gaussian Quadrature) (Rizopoulos, 2021) for Shares and reactions of Angry, Love, Haha, and Sad and the

NBZIMM library (Zhang & Yi, 2020) for Likes, Comments and Wow reactions.[14] Those two libraries can fit a model relatively quickly despite the very large data size (about 11 million posts for both Republican/conservative and Democratic/Liberal pages). Model evaluation metrics such as AIC and BIC were computed from the GLMMadaptive library because those metrics were not available from the NBZIMM library. Comparing those two libraries to other libraries that provide similar features available in R (e.g., glmmADMB, glmmTMB) with a subset of data generated similar fixed and random effects estimates.

For parameters, I specified an intercept only in the zero-part assuming that the probability of the zero reactions was constant across the posts. For a random effect, I used political page names as a grouping variable for a random intercept to control the page's information (such as the page's popularity or size).

Strategy for Public Opinion and Language on Facebook. Using time-series analysis, I modeled lagged negativity and incivility in the posts at time *t-1* on affective polarization at time *t*, indicating that negativity and incivility in the language of the postings led to affective polarization. Instead of constructing a univariate time series analysis with exogenous variables, I ran a standard regression model with the residuals of a univariate time series of affective polarization as a dependent variable and the key independent variables (e.g., the lagged or contemporaneous negativity and incivility). I examine two facets of affective polarization: (1) the *relative* distance of public opinion between the two parties, which captures how polarization is shaped by partisan leanings, and(2) the *absolute* distance of the public opinion between the two parties, which

[14] Due to the large size of the dataset and/or singularity issues, Likes, Comments, and Wow reactions did not converge in the GLMMadaptive library. Findings from the NBZIMM and GLMMadaptive libraries were consistent for the other measures.

indicates how much polarization there is, regardless of which party people feel more favorably toward.

To be specific in modeling process, I tested whether the affective polarization measures (both relative and absolute measures of partisan sentiment and feeling thermometer) had a unit root using Dickey-Fuller (Fuller, 1996), Phillips-Perron (Phillips & Perron, 1988), and KPSS (Kwiatkowski et al., 1992) tests. A unit root test determines whether differencing (i.e., a process of making a time series stationary so that properties of the time series are independent on the time when the series is observed) is required (Hyndman & Athanasopoulos, 2018). In addition, I used a test for fractional integration (FI) because the time series were not fully integrated but potential fractional integration was identified. Then, I chose the best model properties of AR (autoregression) and MA (moving average) components according to AIC (Akaike's Information Criteria) and BIC (Bayesian Information Criteria). Finally, I applied a similar approach of a regression with ARIMA errors as Hyndman and Khandakar (2008) suggested, but chose a regression with ARFIMA errors, instead ARIMA. In summary, I constructed a univariate time series model, took the residuals that AR (auto regressive), FI (fractional integration), and MA (moving average) properties produced (i.e., pre-whitening), and built a standard regression model with the exogenous variables in which I am interested, namely negativity and incivility. Note that when constructing a regression model with multiple independent variables, the dependent variable is (1) the residuals of a univariate ARFIMA model of either the relative value of partisan sentiment or feeling thermometer or (2) the residuals of a univariate ARFIMA model of either the absolute value of partisan sentiment or feeling thermometer, rather than directly fitting a univariable time series of the actual values of sentiment or feeling thermometer with exogenous variables.

For partisan sentiment, tests for stationarity show that the null hypothesis that a time series has a unit root is rejected (Dickey-Fuller test statistic = -4.01, $p < .05$; Phillips-Perron test statistic = -26.01, $p < .01$; KPSS test statistic = 0.11, $p > .1$). The best model fit (using auto.arima in the R forecast library) was ARIMA (0, 1, 1) (AIC=340.32, BIC=344.19). Fractional integration was identified (d = 0.45) from the R fracdiff library (Haslett & Raftery, 1989). The final model of ARFIMA is (0, 0.45, 1) (AIC = 203.18; BIC = 210.98). The residuals were taken for a regression model with the exogeneous variables. Figure 3.4.1 shows an overview of the relative partisan sentiment. ACF (an autocorrelation function) measures the relationship between the current and previous lags and PACF (a partial autocorrelation function) measures the relationship between the current and previous lags after removing the effects of earlier lags (Hyndman & Athanasopoulos, 2018). Both ACF and PACF helps us analyze the temporal dynamics of a time series. Figure 3.4.2 demonstrates the residuals of a univariate ARFIMA (0, 0.45, 1) on the relative sentiment measure. Residuals display a normal distribution and there is no significant correlation in the residual series.

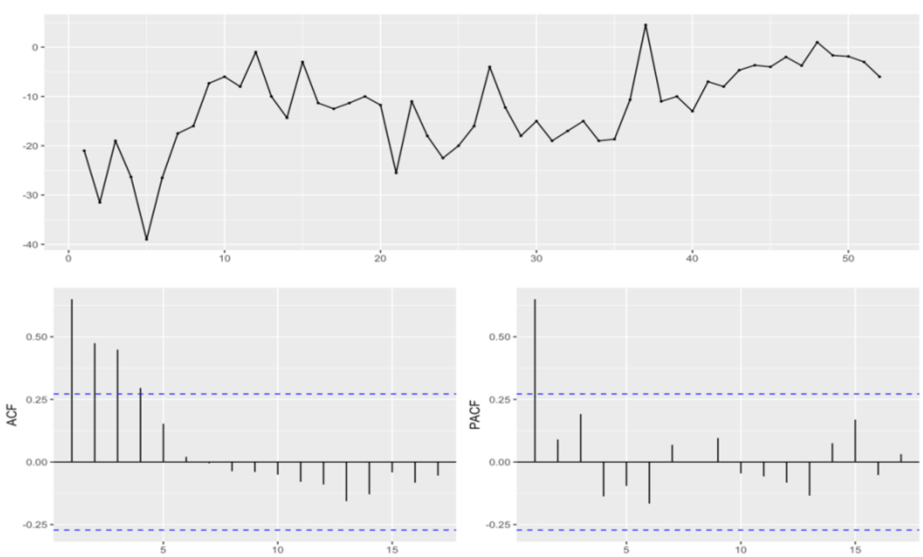


Figure 3.4.1: Time Series of Partisan Relative Sentiment with ACF and PACF.

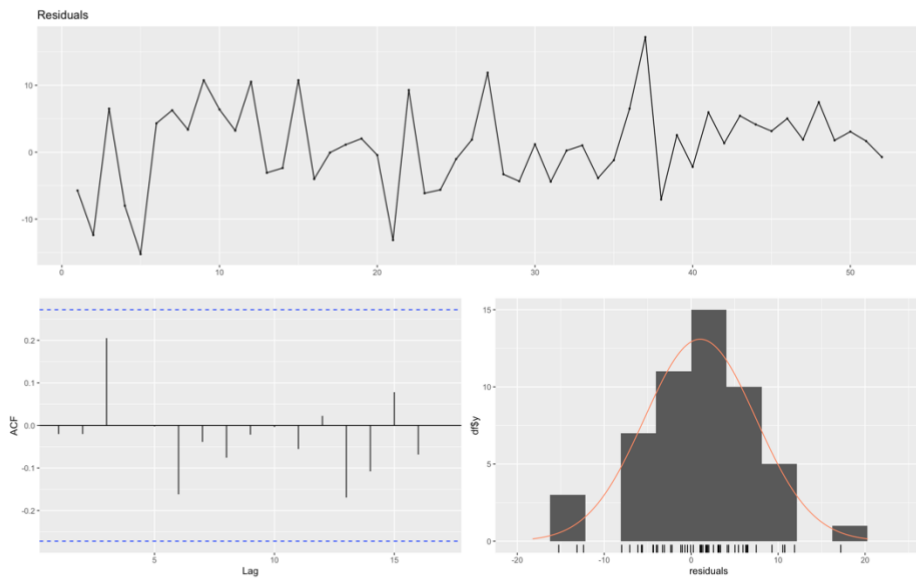


Figure 3.4.2: Residuals of ARFIMA on Partisan Relative Sentiment.

For the absolute value of partisan sentiment, the same tests were applied (Dickey-Fuller test statistic = -3.76, $p < .05$; Phillips-Perron test statistic = -23.6, $p < .01$; KPSS test statistic = 0.11, $p > .1$). The best model fit was ARIMA (0, 1, 1) (AIC=332.21, BIC=336.07). Fractional integration was identified ($d = 0.46$) as well. The final model of ARFIMA is (0, 0.46, 1) (AIC = 195.66; BIC = 203.46). The residuals were taken for a regression model with the exogeneous variables. Figure 3.5.1 shows an overview of the absolute value for partisan sentiment measure. Figure 3.5.2 demonstrates the residuals of a univariate ARFIMA (0, 0.46, 1) on the partisan sentiment measure, which is close to a normal distribution and do not present clear correlation patterns.

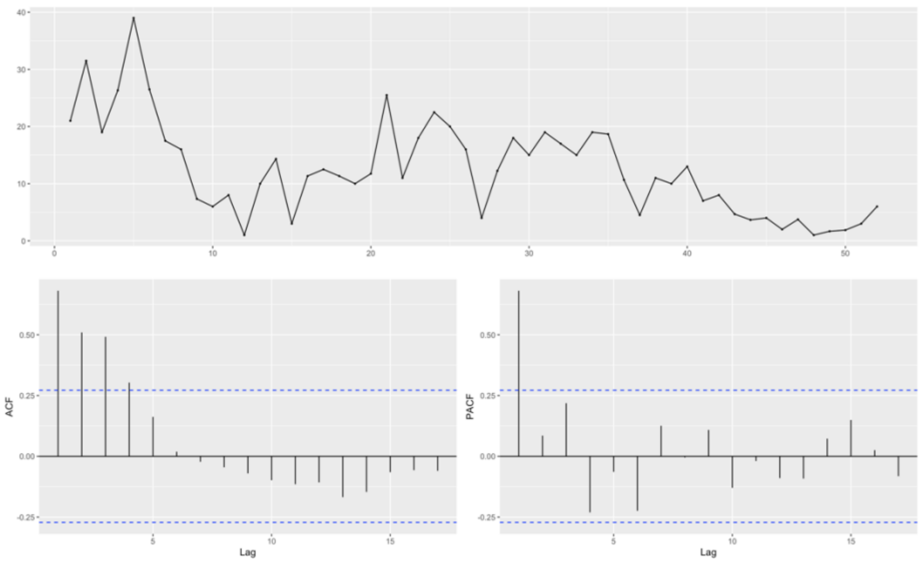


Figure 3.5.1: Time Series of Absolute Value of Partisan Sentiment with ACF and PACF.

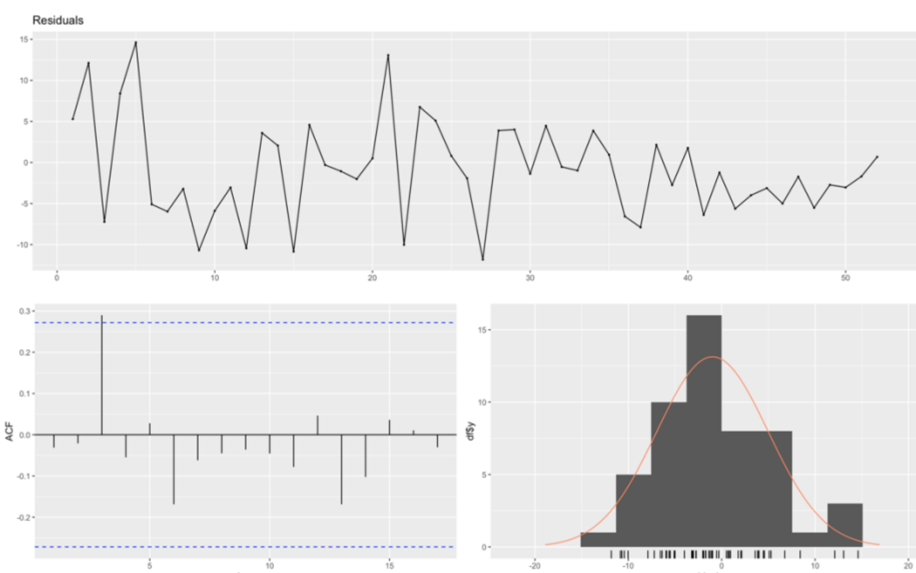


Figure 3.5.2: Residuals of ARFIMA on Absolute Value of Partisan Sentiment.

Likewise, the time series of the partisan relative feeling thermometer is stationary (Dickey-Fuller test statistic = -3.82, $p < .05$; Phillips-Perron test statistic = -25.67, $p < .01$; KPSS test statistic = 0.11, $p > 0.1$), and the best model fit of ARIMA was (1, 1, 0) (AIC = 314.91, BIC = 318.28). Given that fractional integration was found ($d = 0.38$), the final model ARFIMA is (1, 0.38, 0) (AIC = 196.39; BIC = 203.79). The residuals were taken for a regression model with the exogeneous variables. Figure 3.6.1 shows an overview of the relative value for partisan feeling thermometer. Figure 3.6.2 displays the residuals of a univariate ARFIMA (1, 0.38, 0) on partisan feeling thermometer, which show a normal distribution and not specific correlation patterns.

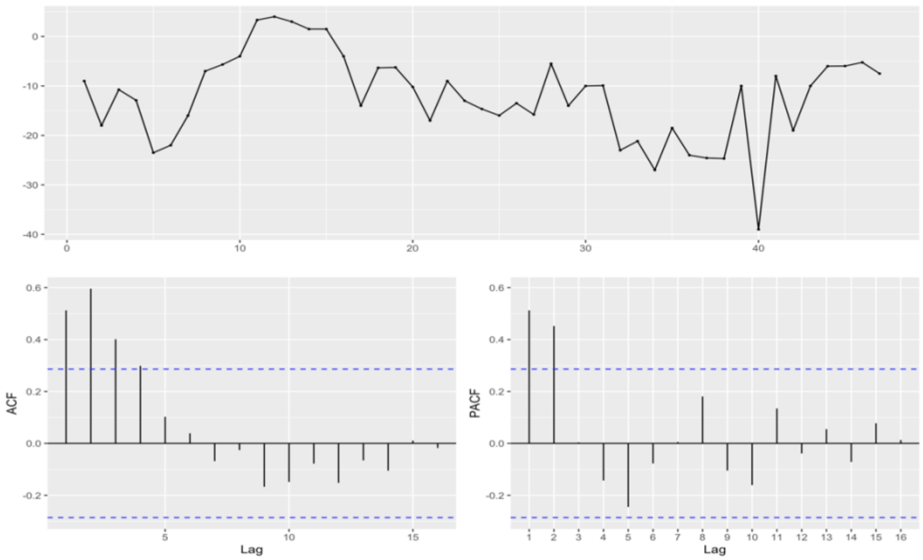


Figure 3.6.1: Time Series of Partisan Relative Feeling Thermometer with ACF and PACF.

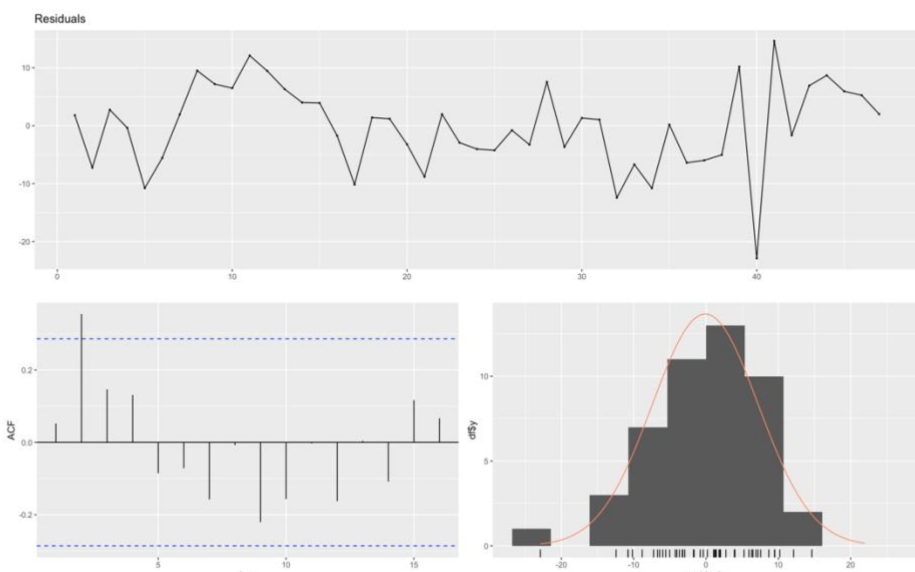


Figure 3.6.2: Residuals of ARFIMA on Partisan Relative Feeling Thermometer.

For the absolute value of partisan feeling thermometer, the same approaches were tested (Dickey-Fuller test statistic = -4.28, $p < .05$; Phillips-Perron test statistic = -31.61, $p < .01$; KPSS test statistic = 0.11, $p > .1$). The best model fit of ARIMA was (1, 0, 2) (AIC=317.54, BIC=326.79). Fractional integration was identified ($d = 0.33$) and the final model ARFIMA is (1, 0.33, 2) (AIC = 185.64; BIC = 196.74). The residuals were taken for a regression model with the exogeneous variables. Figure 3.7.1 shows an overview of the absolute value for partisan feeling thermometer measure. Figure 3.7.2 displays the residuals of a univariate ARFIMA (1, 0.33, 2) on the partisan feeling thermometer measure, which is close to a normal distribution and not much evidence of correlation patters.

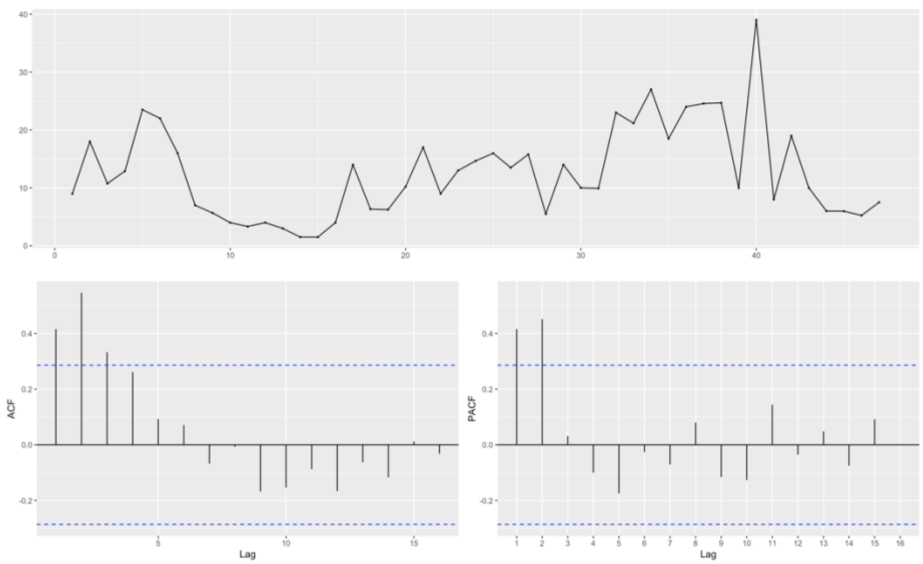


Figure 3.7.1: Time Series of Absolute Value of Partisan Feeling Thermometer with ACF and PACF.

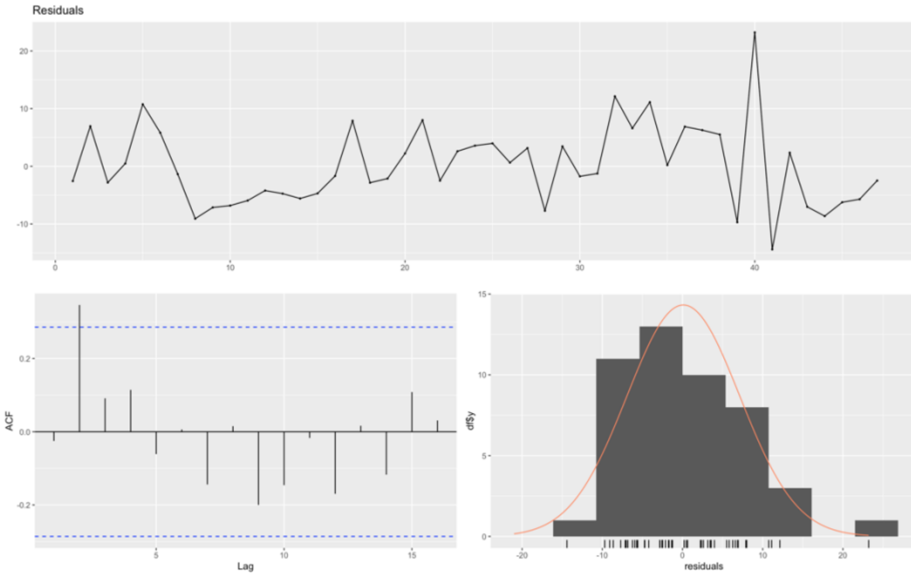


Figure 3.7.2: Residuals of ARFIMA on Absolute Value of Partisan Feeling Thermometer.

Control Variable

To rule out a possible rival relationship between language and aggregate polarization, an economic variable was used as a control variable. Economic conditions could influence the aggregate level of polarization among the public and could be related to the use of negative and uncivil language on social media. Studies have found that economic factors in terms of income inequality and policy changes may drive political polarization (Feddersen & Gul, 2014; Winkler, 2019). These studies find that an aggregate level of rising political polarization is closely associated with greater economic policy uncertainty.

It is also possible that economic conditions relate to negative and uncivil language use. On social media, the tone of Twitter data corresponds to perceptions of societal and economic changes, such as consumer confidence (Pasek et al., 2018; Soroka et al., 2018). Further, research has shown that affect is related to how people process political information about economic conditions (e.g., Conover & Feldman, 1986; Hester & Gibson, 2003; Soroka et al., 2015). Taken together, there may be a relationship between partisan language use on Facebook and the public's economic perceptions.

Economic Condition. To control for economic conditions, the Consumer Confidence Index (CCI) was aggregated by each quarter within a year (i.e., averaging monthly scores). The CCI refers to "an indication of future developments of households' consumption and savings, based on answers regarding their expected financial situation, their sentiment about the general economic situation, unemployment, and capability of savings" (OECD, 2021). This index is the aggregated measure of four questions with seasonal adjustment (e.g., weather changes, holidays) and response weights (a more extreme response receives more weight) given five response options (a lot better, a little better, the same, a little worse, and a lot worse). Following are the four items: (a) assessment of financial situation over the last 12 months, (2) expected financial position over the next 12 months, (3) expected general economic situation over the next 12 months, and (4) expected major purchases over the next 12 months. A score of more than 100 indicates an increase in consumer confidence toward the future economic situation, resulting from respondents' tendency to save less and spend more over the next 12 months, while a score of less than 100 displays a pessimistic attitude towards future developments in the economy, possibly resulting in an inclination toward saving more and spending less (OECD, 2021). The concept of consumer confidence has been examined in the context of politics and media by other works (e.g., De Boef & Kellstedt, 2004; Pasek et al., 2018).

Individual Level of Analysis: Methods

Using a Facebook setting, a survey experiment was conducted. Details of the research design, participants, stimuli creation, variable operationalization, and model selection for analysis are discussed in turn.

Research Design

A 3 (language target) x 3 (language use) between-subjects experiment was used to determine the effects of partisan and affective language: whether the language explicitly targeted Democrats, Republicans, or neither and whether the language use was negative, uncivil or neutral.

Participants completed the consent form and four screening questions about age, residency, whether they have a Facebook account, and how often they use Facebook (those who answered less than 18, not U.S. residents, did not have a Facebook account, or "never" used Facebook did not participate in the study). Then, respondents were randomly assigned to one of 9 conditions, each of which contained 6 political and 10 non-political posts in a random order displayed as part of a mock Facebook News Feed. They were asked to interact with posts as they would posts appearing in their News Feed. They could click to read news stories embedded in a post, comment on, react to (Like, Love, Haha, Wow, Sad, and Angry), or share each post if they wanted. After exposure to the experimental treatment, a question asking if they were able to see the Facebook posts followed. Afterward, they were asked to write a brief paragraph about their thoughts and feelings regarding the posts that they saw. Questions that measure political polarization, emotions (e.g., angry, anxious, hopeful), and partisan stereotypes (e.g., honest, selfish) followed. The study concluded with political background and demographic questions.

Participants

Data from 1,035 respondents who were at least 18 years old and U.S. residents with an active Facebook account that they accessed more than "never" were collected through Amazon Mechanical Turk (MTurk). Past research has found that experimental findings on Mturk replicate findings from other settings (Berinsky et al., 2012; Coppock, 2019). The participants were paid $1.80 for their participation. Respondents that answered "No" to if they were able to see the

Facebook posts, did not complete, and with duplicated IP addresses are not included in the data (n = 75). In the analysis, it is limited to the sample of 920 participants who identified as strong or leaning partisans as related work suggests (e.g., Levendusky, 2013). 36.2% identified as female, 65.4% as white non-Hispanic, 3.6% as Black non-Hispanic, 26.0% as Hispanic, and 2.3% as Asian. The average age was 37.83 (*SD* = 10.83) and 81.5% identified as having a college degree or more. 63.6 percent identified as Democrats and 32.2 percent as Republicans. Chi-square tests supported that there were no differences among the 9 experimental conditions on those demographic variables.

Stimuli

Political Posts with Negative and Uncivil Language. I used two steps in creating stimuli: 1) using a feature selection and clustering process, posts containing higher negativity or incivility across political topics were identified from the Facebook posts collected for the aggregate part of this dissertation and 2) inspired by those posts, comparable messages were crafted as stimuli for the experimental conditions. Although the stimuli are not perfect in terms of external validity, meaning that I did not use political content on the actual partisan pages, I chose internal validity to take advantage of the benefits of an experiment setting in order to examine cause-and-effect relationships between attributes of language and outcomes (social engagement and affective polarization).

For (1), I took the following steps: posts that were negative (i.e., SentiStrength sentiment score is less than 0) or uncivil (i.e., Insult score from Google Perspective API is more than 0.2) were selected while duplicated posts were removed. I then took the top 10 percent of the negative posts (lower than or equal to -3 of SentiStrength score, n = 671,110) and the top 10 percent of the uncivil posts (higher than or equal to 0.7739 of Insult score, n = 392,657) for language analysis.

Using the spaCy library on Python, words in those posts were then tokenized with lemmatization. Then, stopwords (i.e., not informational words, such as prepositions, conjunctions, etc.), pronouns, and non-alphabetical characters were removed. Only lemmatized words with more than 100 occurrences including both unigrams (i.e., a word) and bigrams (i.e., a pair of consecutive word units) were included.

To extract negative and uncivil language, this study applied Monroe et al.'s (2009) feature selection approach that examines language usage rates computed by each word or n-gram (i.e., weighted log-odds-ratio) with informative Dirichlet prior (i.e., an expected distribution of words in a random or sample text). Monroe et al.'s (2009) study examined various techniques that capture partisan differences in political texts and suggested several approaches for feature selections, such as shrinkage and regularization. In this dissertation, I compute the Z-score for each word and phrase using this process. Table 3.7.1 and Table 3.7.2 show the top 30 words and phrases that occur in posts that have high levels of negativity and incivility. Although most of these words are not necessarily, by themselves, negative or uncivil, their repetition in posts that contained negativity and incivility are instructive for the types and topics of posts to create for the experimental stimuli.

Table 3.7.1: Phrases that Occur in Highly Negative Posts listed by Z-score (Top 30).

Posts from Democratic pages (z-score)		Posts from Republican pages (z-score)	
trump (-81.37)	right wing (-25.75)	obama (75.15)	leftist (42.4)
republican (-43.21)	raw story (-24.85)	hillary (62.93)	terrorist (42.24)
donald (-42.76)	pandemic (-24.05)	liberal (59.3)	muslim (40.38)
donald trump (-42.07)	puerto (-23.83)	illegal (55.4)	breitbart (38.24)
republicans (-40.62)	raw (-23.49)	party news (51.7)	breaking (37.8)
gop (-40.05)	president donald (-23.3)	democrat (50.59)	illegal immigrant (36.03)
trump administration (-32.99)	gun violence (-22.79)	left (48.1)	video (36.02)
bernie (-31.23)	puerto rico (-22.16)	isis (47.61)	medium (35.82)

climate (-29.46)	rico (-21.97)	tea party (47.54)	gateway pundit (35.07)
occupy (-28.45)	administration (-21.69)	news (46.19)	gateway (34.23)
sanders (-28.15)	nowthis (-21.38)	clinton (45.82)	liberty (33.41)
bernie sanders (-27.01)	white supremacist (-21.36)	tea (45.05)	antifa (32.73)
health (-26.8)	indigenous (-21.28)	illegal alien (44.39)	islam (32.67)
president (-26.1)	democracy (-21.27)	alien (43.86)	redstate (32.38)
occupy democrats (-25.86)	oil (-20.99)	islamic (43.65)	redstate news (31.83)

Table 3.7.2: Phrases that Occur in Highly Uncivil Posts listed by Z-score (Top 30).

Posts from Democratic pages (z-score)		Posts from Republican pages (z-score)	
trump (-63.45)	follow occupy (-18.69)	liberal (61.98)	video (25.86)
republican (-45.6)	notmypresident (-18.54)	hillary (59.98)	black (24.93)
donald (-42.01)	rick (-17.73)	obama (48.45)	chick right (23.85)
donald trump (-39.43)	putin (-17.68)	democrat (46.35)	gun control (23.08)
republicans (-37.64)	covid (-17.35)	leftist (40.87)	nancy (23.08)
occupy (-36.69)	president (-17.03)	clinton (39.92)	pathetic (22.86)
occupy democrats (-36.2)	patrick (-16.37)	left (37.51)	nancy pelosi (22.83)
gop (-33.35)	dc tribune (-16.26)	hollywood (33.55)	muslim (22.56)
fucking (-27.73)	american news (-15.96)	gun (32.96)	media (22.5)
fuck (-27.56)	fox news (-15.76)	cnn (31.91)	ridiculous (21.68)
asshole (-26.23)	trump say (-15.69)	pelosi (30.8)	socialist (21.65)
activism (-20.43)	coronavirus (-15.63)	hillary clinton (30.29)	antifa (21.59)
republican party (-20.11)	rude rotten (-15.58)	illegal (29.36)	anti trump (21.51)
right wing (-19.23)	rotten republicans (-15.51)	medium (26.45)	islam (20.96)
fox (-18.76)	share post (-15.45)	chick (26.08)	schumer (20.73)

Later, I applied a word-embedding model and K-means clustering. To be specific, with the lemmatized words above, I trained a neural embedding model (Mikolov et al., 2013) with 50 dimensionality vectors for each word or phrase with a CBOW model. With the 50-dimensional representation for the words and phrases, I applied K-means clustering (k = 30), which iteratively finds optimized positions of the centroids, to identify close words in groups. A similar approach has been used in other work (Muddiman et al., 2020).

Among the 30 clusters of negative or uncivil language identified by the z-scores from feature selection, I identified several dominant topics. Example cases that include both negative and uncivil phrases are presented in Figures 3.8.1 and Figure 3.8.2 and include topics such as gun control/violence, immigration, pandemic/lockdown, partisan news, police brutality/BLM, and left vs right wing. One noticeable difference is that clusters of negative partisan language tend to have clear issue-specific words by political leanings (e.g., gun control vs. gun violence in negative language clustering) while clusters of uncivil partisan language tend to show offensive and insulting phrases (e.g., gun nut in uncivil language clustering). Among these dominant topics, I chose six clusters of gun control/violence, immigration, partisanship, COVID-19, BLM, media as reference themes for the experimental stimuli.

Next, based on posts that contain partisan words with higher or lower incivility across the six political topics (gun, immigration, COVID-19, BLM, media, and partisanship), messages were crafted for a Facebook mock-up environment inspired by actual posts. Table 3.8 displays several actual posts that were used as a reference of negative and uncivil terms. The crafted stimuli messages were assessed for negativity through SentiStrength and incivility through Insult on Google Toxicity. Negative (M = -2.83, ranging -2 to -4) and uncivil (M = -2.83, ranging -2 to -4) content had the same negativity score, while neutral content was closer to zero (M = -0.17, ranging -1 to 1). The uncivil stimuli are more uncivil (M = 0.87, ranging 0.77 to 0.96) than the negative (M = 0.21, ranging 0.04 to 0.56) or neutral stimuli (M = 0.15, ranging 0.002 to 0.34). Full experimental messages with sentiment and incivility scores are in Appendix C.

Table 3.8: Examples of Negative and Uncivil Posts that Inspired the Crafting of Political Message Stimuli.

(Leanings / Topic) Posts (some sentences are truncated)
(Republican / Gun) In California today, the state with strong gun free zones, had children shot and killed at a middle school. A MIDDLE SCHOOL! Wake up California you morons. An

armed officer on site and this doesnt even happen. And if it did the criminal would have been shot dead without any school tragedies. I swear the stupidest people in the states live in California. Conservatives need to pack up their businesses and leave that rat scrotum state for good.

(Democratic / Gun) The "good guy with guns" is such a dumb argument. Do you want to know how to tell a "good guy with a gun" with a "bad guy with a gun"? When the "good guy with a gun" opens fire on you, then you know! (truncated) Without that stockpile of automatic guns, if he used a knife, 59 people would still be alive and 527 would not be injured. Or we can just keep on selling all the guns and all the ammo and all the military grade killing machines designed for speedy mass murder to all the men and praying that ONE of them somewhere near us doesn't lose his shit. We see how well that is working for us as a country right now.

(Republican / Immigration) These Democrat sonsofbitches make me sick. Talk about an actual threat to our Republic. They knowingly break our immigration laws with their goddamn sanctuary cities, where they deliberately harbor and protect known criminal illegal alien murderers, rapists and child molesters that in too many instances have gone on to murder or seriously harm American citizens. Democrats are a clear and present threat to our national security and to the public safety of every American (truncated).

(Democratic / Immigration) The white nationalist monsters running immigration control are trying to hide the crimes against humanity they commit daily in our name. The “Acting” heads of all four immigration agencies must be imprisoned for contempt of court, since they defy court orders by continuing to separate families, to illegally detain people (including children) in horrific, abusive conditions, and to block Congress's constitutional oversight power. A child who spilled soup was told that she would not get any more food until she drank the soup off the floor! The cruelty is demanded from the top (truncated).

(Republican / BLM) SEE WHERE THIS IS GOING...Why not just let the far left create a new dictionary listing all the words and phrases that people of white color are not allowed to use. And if anyone does use those words or phrases they will be deemed racist and subject to have Antifa, and other terrorist groups pay them a visit at their homes and do what violent hate groups do. Waters Classifies 'Rioting' as 'Negative Language ... Against Black People' (truncated).

(Democratic / BLM) This is a disgusting statement from the chairman of the Connecticut Republican Party. Let's be very, very clear: There is no comparison between peaceful protesters seeking racial justice and violent, murderous racists. Alt-right Romano makes comparison between white supremacists and Black Lives Matter (truncated).

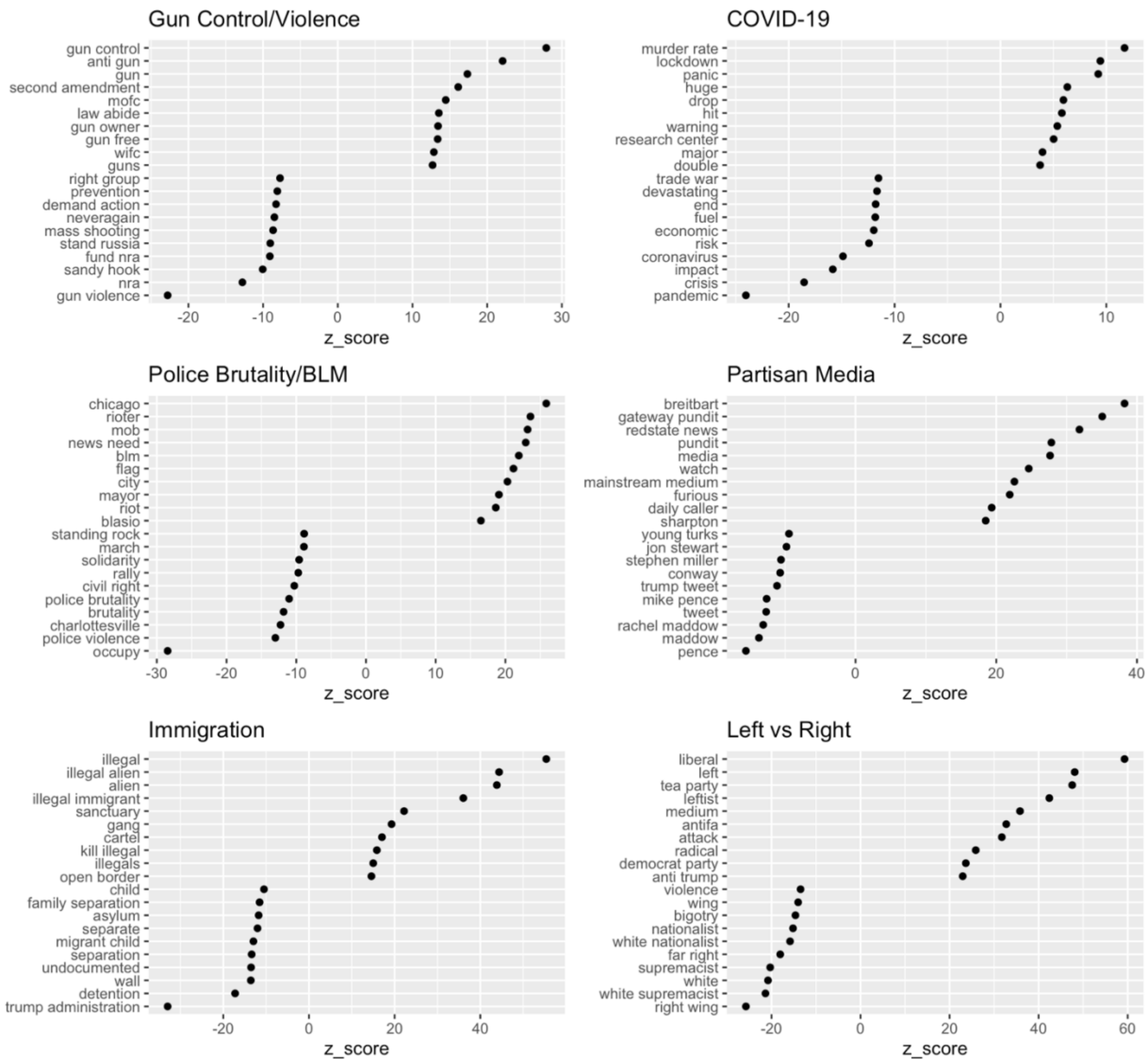


Figure 3.8.1: Clusters of Negative Phrase by Z score with Top 10 Phrases of Each Political Leaning (negative score indicates words used on Democratic-leaning pages; positive score indicates words used on Republican-leaning pages).

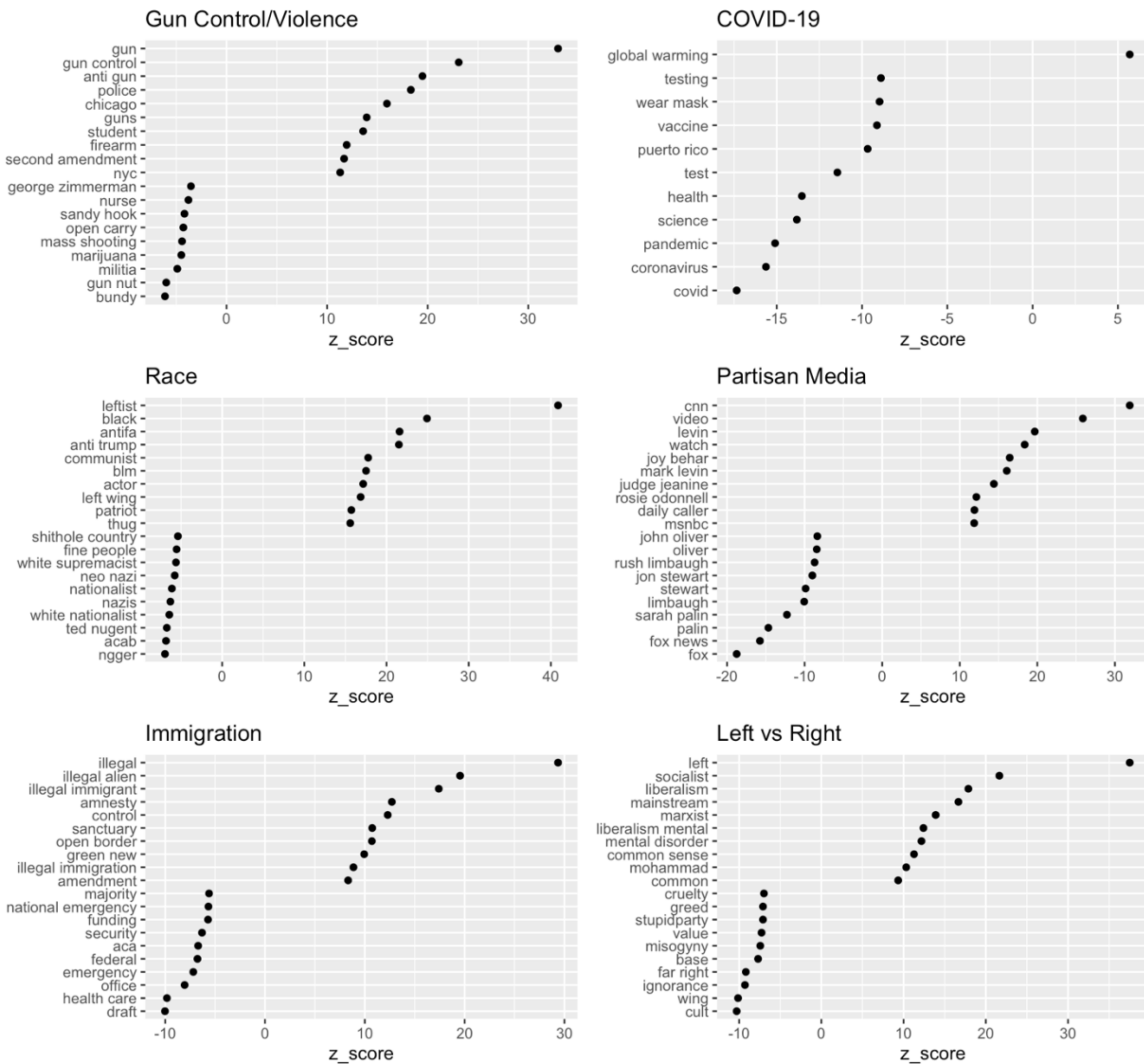


Figure 3.8.2: Clusters of Uncivil Phrase by Z score with Top 10 Phrases of Each Political Leaning (negative score indicates words used on Democratic-leaning pages; positive score indicates words used on Republican-leaning pages).

Political account profiles were made using images and names that signaled political leanings with comparable logos and font. The Republican profile was named “The Free America (TFA)” and used an image of American flags, the Democratic profile was named “Stand Up America (SUA)” and used an image of the Statue of Liberty, and the nonexplicit political account was named “Memos to America (MTA)” and used an image of art supplies. The political account logos are shown in Figure 3.9. A full list of stimuli is in Appendix C.

Figure 3.9: Political Account Logo Images.

Experimental stimuli used different formats including a regular text message post with a news link or related image, a text meme (text message as an image), and a Twitter message (text message from a Twitter post) along with profile of political account. These post types are commonly found on Facebook political pages. Examples of the stimuli used are present in Figures 3.10 to 3.13.

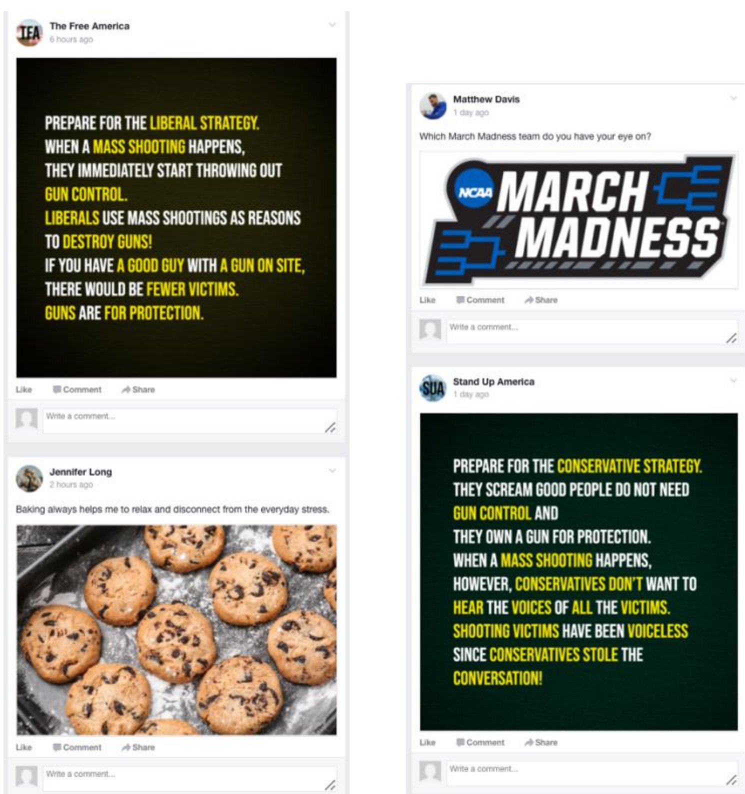


Figure 3.10: Negative Messages about Guns Targeting Democrats (left-top), Republicans (right-bottom) in a Facebook Setting.

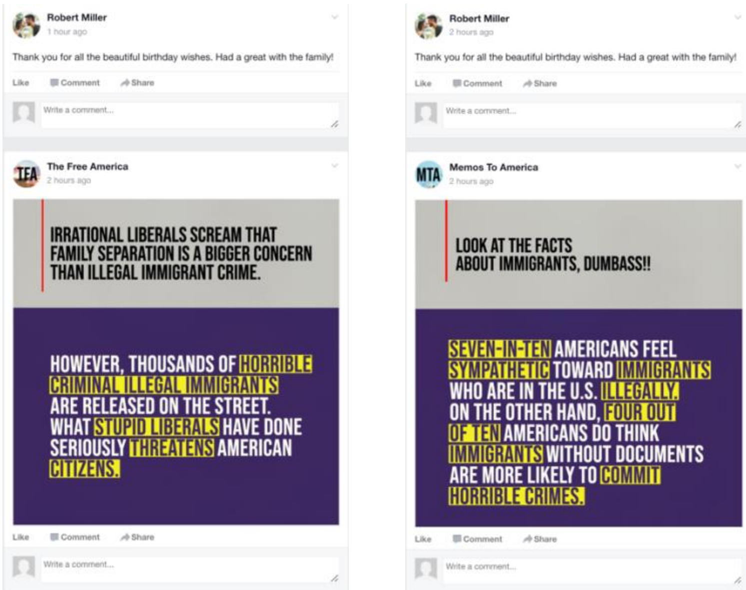


Figure 3.11: Uncivil Messages about Immigration Targeting Democrats (left-bottom) and No Explicit Entities (right-bottom) in a Facebook Setting.

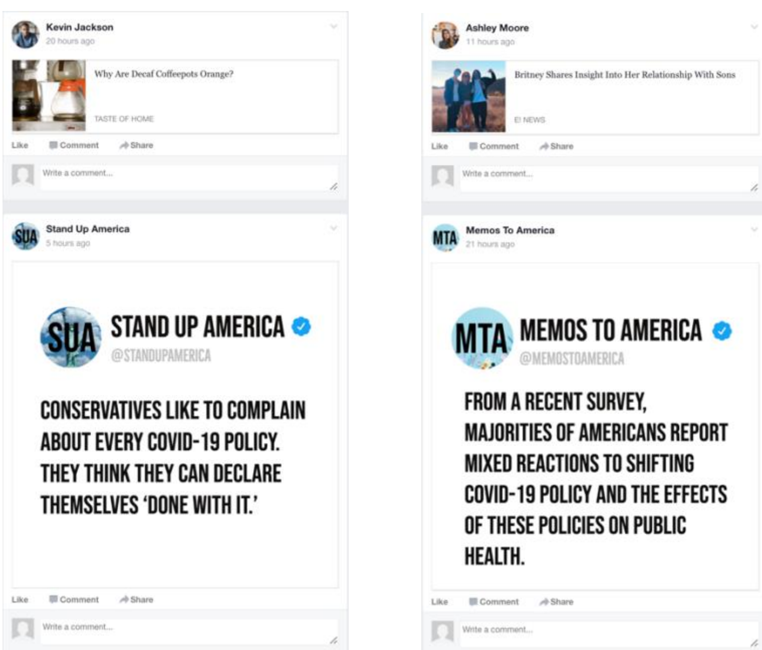


Figure 3.12: Neutral Messages about COVID-19 Targeting Republicans (left-bottom) and No Explicit Entities (right-bottom) in a Facebook Setting.

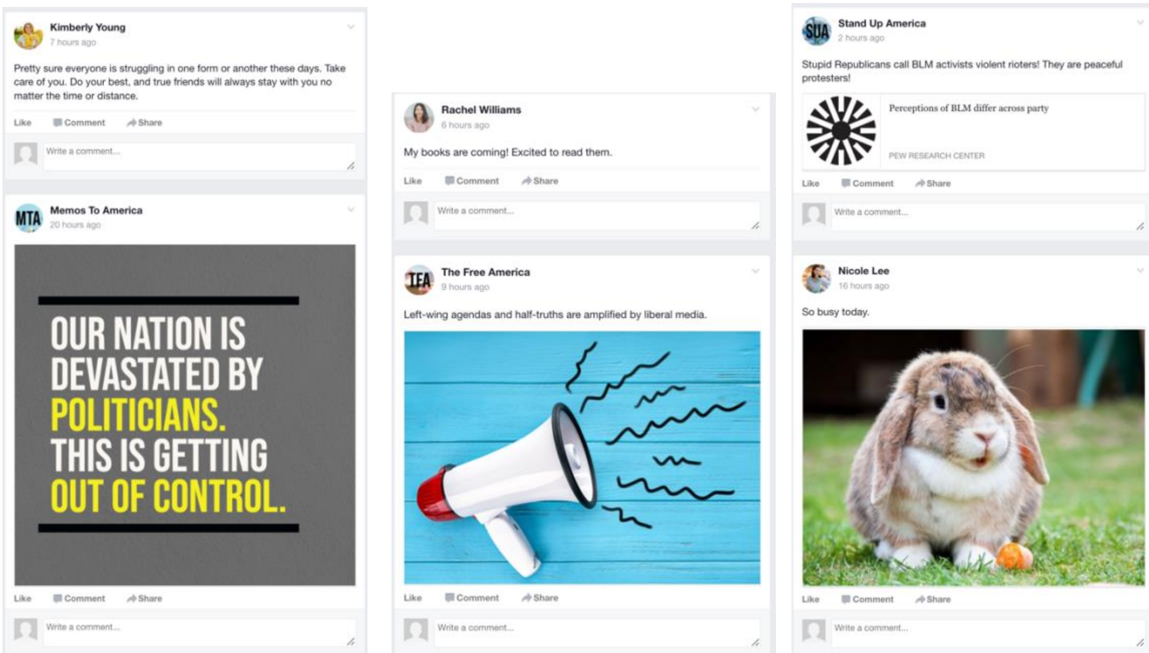


Figure 3.13: Other Stimuli Examples of Negative Messages about Partisanship with No Target (left-bottom), Neutral Message about Media with a Democratic Target (middle-bottom), and Uncivil Message about BLM with a Republican Target (right-top) in a Facebook Setting.

Non-Political Posts. Posts designed to resemble those appearing on personal News Feeds on topics such as a cute animal, a celebrity, sports, food, and movies were also embedded in the experimental setting as demonstrated through Figures 3.10 to 3.13. Some posts such as March Madness and cheer-up messages were added to align with current events. The profile names are random combinations of common first and last names in the U.S. The profile age, gender, and race of the profiles are representative of the U.S. population and the images are from Adobe Creative Cloud.

Dependent Variables

Affective Polarization. Conventional affective polarization measures were asked of participants such as feeling thermometers (from the American National Election Studies, ANES) ranging from 0 (cold) to 100 (warm) toward the Republican [Democratic] Party. Then, positive (intelligent, honest, and open-minded) and negative (closed-minded, lazy, and selfish) stereotypical trait evaluations toward partisans (Iyengar et al., 2012) were asked with 5 scale responses from *strongly agree* (4) to *strongly disagree* (0). Also, emotions of Anger (angry, hostile, disgusted), Enthusiasm (hopeful, pride, enthusiastic), and Anxiety (nervous, anxious, afraid) toward partisans were modified and applied based on Mason's (2016) measures with 4 scale responses from *a great deal* (3) to *not at all* (0). In addition, an open-ended question asked participants to write a paragraph of what they thought or felt about the posts they saw. Full survey items for the individual level analysis are in Appendix D.

For affective polarization, three variables were computed as an absolute value. A partisan absolute feeling thermometer measure was created by an absolute value of the difference between the two parties' net feeling thermometer (*M* = 32.8, *SD* = 28.4, *range* = 0 to 100). A partisan absolute trait evaluation measure was generated by an absolute value of the difference between the two partisans' net trait evaluation (*M* = 1.51, *SD* = 1.80, *range* = 0 to 8). The net trait evaluation was obtained by subtracting average negative trait evaluations from average positive trait evaluations for Democrats [Republicans]. Partisan absolute emotion is crafted by an absolute value of the difference between the two partisans' net emotions (*M* = 1.16, *SD* = 1.38, *range* = 0 to 6). The net emotion was attained by subtracting net Angry emotions on average from Enthusiastic emotions on average for Democrats [Republicans]. I examined the absolute value of affective polarization to see if polarization occurs regardless of its directions by partisanship, with higher values referring to greater polarization.

User Engagement. The experiment was conducted as a replication of the Facebook environment. Participants were able to interact freely with the posts presented, such as Reacting (e.g., Like, Love, Angry, etc.), Sharing, and Commenting. Those activities on the posts were recorded and considered user engagement.

Other Variables. Standard demographic variables were included, such as gender, age, race/ethnicity, education, and income as well as party identification and political ideology.

Independent Variables

Exposure to Language Attributes. Study participants were randomly assigned to see one of three language attributes: negative, uncivil or neutral.

Presence of Political Targets. The Republican profile (TFA) used language targeting Democrats and liberals (the Democratic target condition) while the Democratic profile (SUA) mentioned terms related to Republicans and/or conservatives (the Republican target condition). The nonexplicit political account (MTA) did not use explicit entities associated with partisanship (the non-explicit target condition). Based on these categories, I created a variable of in-group, out-group, and no-explicit target. Note that exposure to in-group targets indicate exposure to counter-attitudinal content while exposure to out-group targets indicates exposure to pro-attitudinal content. Exposure to non-explicit target is used as a reference for presence of political targets.

Analysis

Model Selection for Engagement Metrics. User interactions with posts in the mock Facebook setting were recorded. With participants' IP address on the survey and matching interaction data for the political content only, I created a dataset of how often study participants engaged with each social metric (Comments, Shares, and emotional Reactions) on six political posts. Because many participants did not engage with the political posts, I applied zero inflated models for count variables with Poisson or negative binomial distribution (glmmTMB in R) depending on the likelihood ratio test (lrtest from the lmtest library in R). A summary of the non-engagement rates (respondents who did not interact with any political posts at least once) is present in Table 3.9. On average, participants shared the political posts least frequently and interacted with Like reactions the most.

Table 3.9: Descriptive Statistics for the Engagement Metrics.

Reactions	Non-Engagement Rate (%)	M	SD	Min	Max
All	72.28	1.23	2.36	0	13
Like	75.43	0.76	1.63	0	12
Comments	97.83	0.05	0.40	0	6
Shares	95.87	0.01	0.14	0	3
Angry	91.52	0.07	0.42	0	6
Love	95.54	0.15	0.60	0	6
Sad	95.54	0.06	0.33	0	4
Wow	94.89	0.07	0.37	0	5
Haha	96.30	0.04	0.24	0	3

Model for Affective Polarization. Multivariate regression is applied to capture differences in affective polarization by the experimental conditions of language attributes (negative or uncivil political content with a reference of neutral political content) and presence of political targets (posts with in-group or out-group targets with a reference of those with non-explicit targets). Respondents' party identification is also analyzed to test for partisan differences based on interactions with the experimental conditions.

Chapter 4: Does Negative and Uncivil Language Invite Engagement with Partisan Posts on Facebook?

Whether your News Feed on Facebook is full of news about why Russia invaded Ukraine, memes about the new Spider-man movie, or photos shared by your friends who post every single day, every time you interact with those posts, you feed an algorithm that affects what you see on Facebook. On Facebook, people can encounter news and can be exposed to political information that are inflected by one's personal interest and the behavior of others in one's social network. Although many elements factor into algorithms, such as one's friend lists and the pages followed by friends, social metrics are user-friendly and recurring activities that algorithms use to update the News Feed. My dissertation focuses on interactions with posts from partisan pages featuring negativity and incivility. This process is problematic because more interactions mean higher visibility in Facebook News Feeds.

I propose that negative and uncivil language will invite engagement, such as provoking more discussion through commenting, increasing participation through the spread of the posts, and enhancing Angry reaction clicks. At the same time, negative and uncivil language will discourage certain types of engagement such as Love reactions. News Feeds are more likely to deliver posts that serve one's political leanings, although there are chances that people may accidently encounter something different due to the diversity of their social network. In either case, negative and uncivil posts will strengthen the relationship between partisan content and social engagement, I propose, because they attract biased emotions and urge reactions.

Chapter Overview

In this chapter, I report on the relationship between language and social engagement with posts from partisan Facebook pages. First, H1 to H3 and RQ1 examine how negative posts are associated with Facebook user engagement. Second, hypotheses H6 to H8 and RQ2 look at how uncivil content is linked to user engagement on Facebook. Third, H11 investigates the magnitude of the effect of incivility on Comments, Shares and Angry reactions compared to negativity. RQ3 compares negativity and incivility for the other reactions. Later, RQ4 and RQ5 explore partisan differences in the relationship between Facebook user engagement metrics and political posts that are negative or uncivil. Finally, I discuss the dynamics of negative and uncivil language in political posts as they relate to social engagement.

RESULTS OF WHETHER NEGATIVE CONTENT INCREASED SOCIAL MEDIA ENGAGEMENT

The first set of hypotheses and research question examined how negative posts are associated with Facebook user engagement. Zero-inflated mixed models were used to test the impact of negative content on user engagement after controlling for whether the posts leaned toward Republicans/conservatives (vs. Democrats/liberals) as a dummy variable. Note that as sentiment becomes higher it indicates more positivity and as sentiment becomes lower it means more negativity (e.g., negative coefficients of sentiment mean that posts received more interactions as they became negative and fewer interactions as posts became positive). Evidence shows that more negative content on partisan Facebook pages receives a greater number of Comments (B = -0.09, SE = 0.001, $p < .001$), Shares (B = -0.02, SE = 0.001, $p < .001$), and Angry reactions (B = -0.36, SE = 0.001, $p < .001$) while more negative posts on partisan Facebook pages generate fewer Love reactions (B = 0.24, SE = 0.001, $p < .001$) as shown in Table 4.1. Thus, H1, H2, H3a, and H3b are supported.

Findings also demonstrate that more negative posts on Facebook partisan pages generate fewer Like (B = 0.10, SE = 0.001, $p < .001$) and Haha (B = 0.05, SE = 0.001, $p < .001$) reactions while more negative posts receive a greater number of Sad (B = -0.32, SE = 0.001, $p < .001$) and

Wow (B = -0.16, SE = 0.001, p < .001) reactions after controlling for leanings of the pages (as displayed in Table 4.1), answering RQ1.

Table 4.1: Effects of Negativity on Facebook Social Engagement.

	Comments	Shares	Angry	Love
	B (SE)	B (SE)	B (SE)	B (SE)
Fixed effects:				
Intercept	2.90^{***} (0.05)	5.70^{***} (0.16)	4.12^{***} (0.18)	3.71^{***} (0.12)
Sentiment	-0.09^{***} (0.001)	-0.02^{***} (0.001)	-0.36^{***} (0.001)	0.24^{***} (0.001)
Republican	0.06 (0.07)	-0.36^{**} (0.12)	-0.66^{***} (0.13)	-0.86^{***} (0.10)
Zero-inflation model:				
Intercept	-7.15^{***} (0.01)	-16.97^{***} (2.94)	-16.6^{**} (5.08)	-17.00^{***} (4.76)
Random effects: (Page Name)				
Intercept (SD)	1.38	1.83	1.86	1.73
Number of Groups	1,493	1,493	1,491	1,491
Number of Obs	13,955,869	13,955,869	10,991,537	10,991,537
Log-Likelihood	-44,551,973	-56,336,226	-30,225,470	-28,933,565
AIC	89,103,957	112,672,463	60,450,953	57,867,143
BIC	89,103,989	112,672,495	60,450,985	57,867,174
	Like	Sad	Haha	Wow
	B (SE)	B (SE)	B (SE)	B (SE)
Fixed effects:				
Intercept	4.56^{***} (0.05)	2.87^{***} (0.13)	3.95^{***} (0.23)	1.09^{***} (0.05)
Sentiment	0.10^{***} (0.001)	-0.32^{***} (0.001)	0.05^{***} (0.001)	-0.16^{***} (0.001)
Republican	-0.08 (0.07)	-1.27^{***} (0.10)	-0.49^{***} (0.14)	-0.23^{***} (0.07)
Zero-inflation model:				
Intercept	-19.05^{***} (3.67)	-17.01^{**} (6.47)	-16.28 (3.90)	-4.48^{***} (0.003)
Random effects: (Page Name)				
Intercept (SD)	1.33	1.71	2.16	1.27
Number of Groups	1,493	1,491	1,491	1,491
Number of Obs	13,955,869	10,991,537	10,991,537	10,991,537
Log-Likelihood	-61,868,408	-21,947,373	-27,717,775	-23,472,471
AIC	123,736,829	43,894,757	55,435,563	46,944,955
BIC	123,736,860	43,894,789	55,435,594	46,944,986

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Log-Likelihood, AIC, and BIC are from the GLMMadaptive library for approximate estimates. Sentiment is standardized and centered. Lower sentiment indicates negativity while higher sentiment indicates positivity.

RESULTS OF WHETHER UNCIVIL CONTENT INCREASED SOCIAL MEDIA ENGAGEMENT

Next, zero-inflated mixed models controlling for partisan leanings on Facebook pages show the effects of incivility on user engagement (see Table 4.2). More uncivil posts on Facebook partisan pages generate a greater number of Comments (B = 0.18, SE = 0.001, p < .001), Shares (B = 0.10, SE = 0.001, p < .001), and Angry reactions (B = 0.24, SE = 0.001, p < .001). On the other hand, more uncivil posts produce fewer Love reactions (B = -0.06, SE = 0.001, p < .001). Thus, H6, H7, H8a, and H8b are supported.

RQ2 explores the effects of incivility on Like and other reactions. Findings show that incivility posted on Facebook partisan pages increases Like (B = 0.05, SE = 0.001, p < .001), Sad (B = 0.10, SE = 0.001, p < .001), Haha (B = 0.23, SE = 0.001, p < .001), and Wow (B = 0.11, SE = 0.001, p < .001) reactions, as presented in Table 4.2.

Table 4.2: Effects of Incivility on Facebook Social Engagement.

	Comments	Shares	Angry	Love
	B (SE)	B (SE)	B (SE)	B (SE)
Fixed effects:				
Intercept	2.90^{***} (0.05)	5.68^{***} (0.16)	4.13^{***} (0.17)	3.77^{***} (0.13)
Incivility	0.18^{***} (0.001)	0.10^{***} (0.001)	0.24^{***} (0.001)	-0.06^{***} (0.001)
Republican	0.05 (0.07)	-0.36^{**} (0.12)	-0.66^{***} (0.10)	-0.85^{***} (0.10)
Zero-inflation model:				
Intercept	-6.73^{***} (0.01)	-16.98^{***} (2.99)	-16.50^{***} (5.00)	-16.89^{***} (4.69)
Random effects:				
(Page Name)				
Intercept (SD)	1.40	1.81	1.84	1.75
Number of Groups	1,493	1,493	1,491	1,491
Number of Obs	13,955,869	13,955,869	10,991,537	10,991,537
Log-Likelihood	-44,140,861	-55,784,046	-30,779,569	-28,151,043
AIC	88,281,735	111,568,104	61,559,149	56,302,097
BIC	88,281,767	111,568,136	61,559,181	56,302,129

	Like	Sad	Haha	Wow
	B (SE)	B (SE)	B (SE)	B (SE)
Fixed effects:				
Intercept	4.56^{***} (0.05)	2.91^{***} (0.13)	3.91^{***} (0.24)	1.10^{***} (0.05)
Incivility	0.05^{***} (0.001)	0.10^{***} (0.001)	0.23^{***} (0.001)	0.11^{***} (0.001)
Republican	-0.07 (0.07)	-1.30^{***} (0.10)	-0.45^{***} (0.14)	-0.24^{***} (0.07)
Zero-inflation model:				
Intercept	-19.17^{***} (3.89)	-17.68 (9.32)	-16.38^{***} (4.07)	-4.78^{***} (0.003)
Random effects:				
(Page Name)				
Intercept (SD)	1.33	1.70	2.14	1.26
Number of Groups	1,493	1,491	1,491	1,491
Number of Obs	13,955,869	10,991,537	10,991,537	10,991,537
Log-Likelihood	-55,821,953	-22,026,046	-26,924,816	-44,140,861
AIC	111,643,917	44,052,104	53,849,644	88,281,735
BIC	111,643,949	44,052,136	53,849,676	88,281,767

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Log-Likelihood, AIC, and BIC are from the GLMMadaptive library for approximate estimates. Incivility is standardized and centered.

NEGATIVITY VS. INCIVILITY ON SOCIAL MEDIA ENGAGEMENT

The previous two tables look at the main effects of negativity and incivility without controlling for the other. Yet we would anticipate that sorting out whether the post is negative, controlling for incivility, and vice versa, may produce different patterns. The correlation between sentiment (i.e., SentiStrength) and incivility (i.e., Insult) is -0.257 ($p < .001$), which does not indicate a multicollinearity issue. Including both negativity and incivility as standardized and centered variables, Table 4.3 allows for a comparison of the two effects.

Table 4.3: Comparisons of Negativity and Incivility on Social Engagement.

	Comments	Shares	Angry	Love
	B (SE)	B (SE)	B (SE)	B (SE)
Fixed effects:				
Intercept	2.95^{***} (0.05)	5.66^{***} (0.15)	4.09^{***} (0.17)	3.71^{***} (0.12)
Sentiment	-0.04^{***} (0.001)	0.002^{***} (0.001)	-0.31^{***} (0.001)	0.24^{***} (0.001)
Incivility	0.17^{***} (0.001)	0.09^{***} (0.001)	0.17^{***} (0.001)	0.02^{***} (0.001)
Republican	-0.06 (0.07)	-0.36^{**} (0.11)	-0.63^{***} (0.10)	-0.86^{***} (0.10)

Zero-inflation model:				
Intercept	-6.73*** (0.01)	-16.96*** (3.02)	-16.75** (5.47)	-17.03*** (4.74)
Random effects:				
(Page Name)				
Intercept (SD)	1.40	1.80	1.85	1.72
Number of Groups	1,493	1,493	1,491	1,491
Number of Obs	13,955,869	13,955,869	10,991,537	10,991,537
Log-Likelihood	-43,579,829	-55,522,733	-30,510,505	-29,310,553
AIC	87,159,672	111,045,481	61,021,024	58,621,120
BIC	87,159,709	111,045,518	61,021,061	58,621,157

	Like	Sad	Haha	Wow
	B (SE)	B (SE)	B (SE)	B (SE)
Fixed effects:				
Intercept	4.55*** (0.05)	2.87*** (0.13)	3.89*** (0.22)	1.09*** (0.05)
Sentiment	0.13*** (0.001)	-0.31*** (0.001)	0.12*** (0.001)	-0.14*** (0.001)
Incivility	0.09*** (0.001)	0.02*** (0.001)	0.26*** (0.001)	0.08*** (0.001)
Republican	-0.08 (0.07)	-1.27*** (0.10)	-0.48*** (0.13)	-0.23*** (0.07)
Zero-inflation model:				
Intercept	-18.98*** (3.55)	-17.01** (6.48)	-16.24*** (3.85)	-4.72*** (0.003)
Random effects:				
(Page Name)				
Intercept (SD)	1.33	1.70	2.12	1.26
Number of Groups	1,493	1,491	1,491	1,491
Number of Obs	13,955,869	10,991,537	10,991,537	10,991,537
Log-Likelihood	-60,601,797	-21,947,076	-26,893,475	-43,579,829
AIC	121,203,608	43,894,166	53,786,963	87,159,672
BIC	121,203,645	43,894,203	53,787,001	87,159,709

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Log-Likelihood, AIC, and BIC are from the GLMMadaptive library for approximate estimates. Sentiment and incivility are standardized and centered. Lower sentiment indicates negativity while higher sentiment means positivity.

Incivility has a greater effect on Comments (B = 0.17, SE = 0.001, p < .001) than negativity does (B = -0.04, SE = 0.001, p < .001), and both measures are statistically significant. When both sentiment and incivility measures are included, both negativity and incivility are associated with *increases* in Comments. Figure 4.1.1 and Figure 4.1.2 show the differences of fitted values of Comments in the fixed model based on posts of sentiment and incivility. The slope of incivility is larger on the fitted values of Comments (Figure 4.1.2) than that of sentiment (negativity, Figure 4.1.1). Thus, H11a is supported.

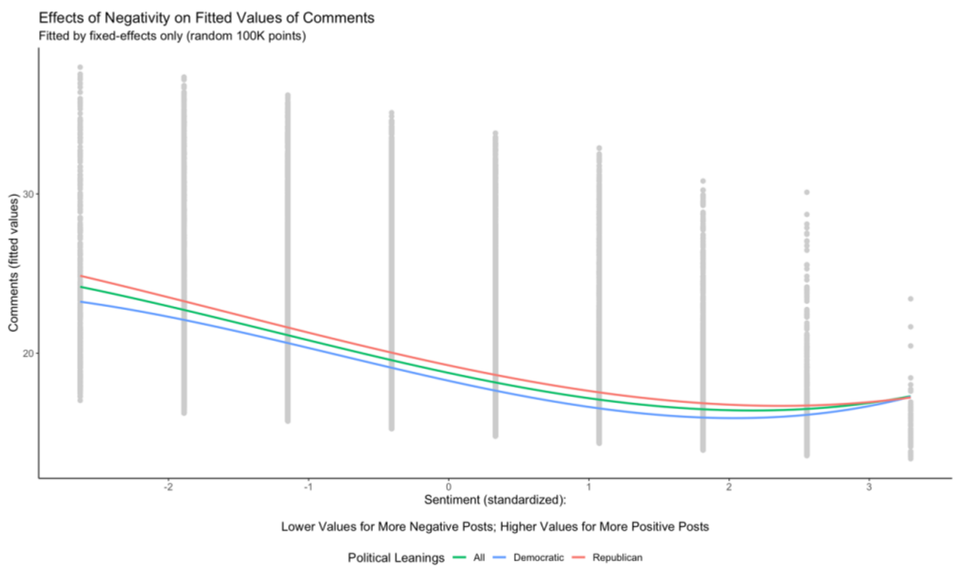


Figure 4.1.1: Effects of Negativity (Lower Values in Sentiment) on the Fitted Values of Comments.

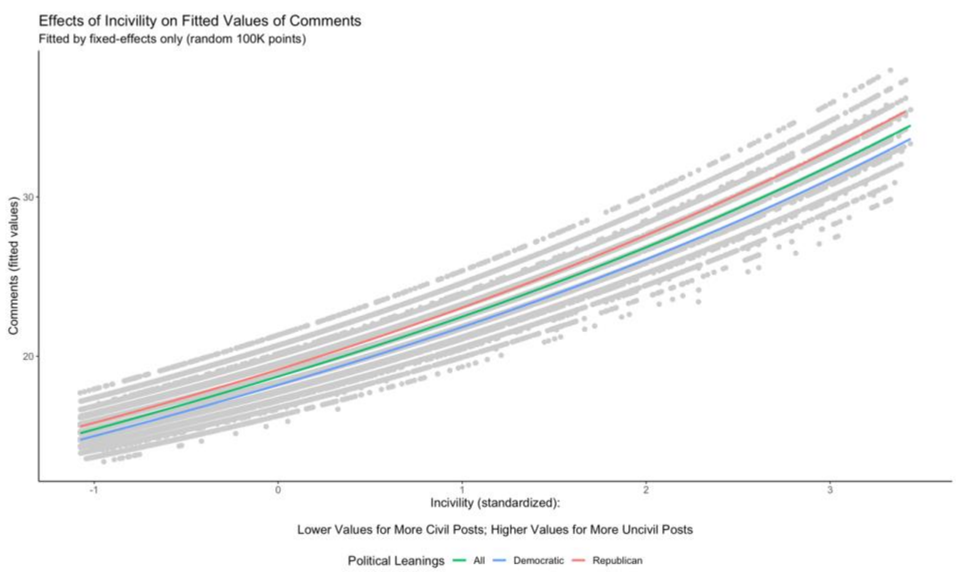


Figure 4.1.2: Comparing Effects of Incivility on the Fitted Values of Comments.

Moreover, when both sentiment and incivility measures are included, incivility is associated with *increases* in Shares ($B = 0.09$, $SE = 0.001$, $p < .001$) while negativity is correlated with *decreases* in Shares ($B = 0.002$, $SE = 0.001$, $p < .001$). Still, the comparison in magnitude shows greater effects of incivility than negativity. Because the estimate of negative posts is very small ($B = 0.002$), the fitted line (green) in Figure 4.2.1. does not show clear direction. Rather, the estimate of uncivil posts is larger ($B = 0.09$), and the fitted line is a positive direction present in Figure 4.2.2. H11b is supported.

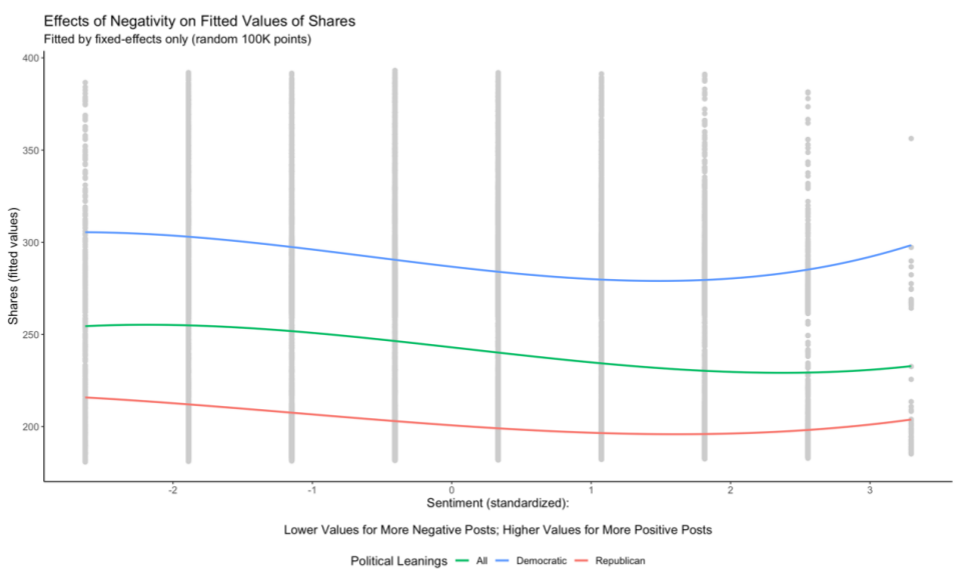


Figure 4.2.1: Comparing Effects of Negativity (Lower Values in Sentiment) on the Fitted Values of Shares.

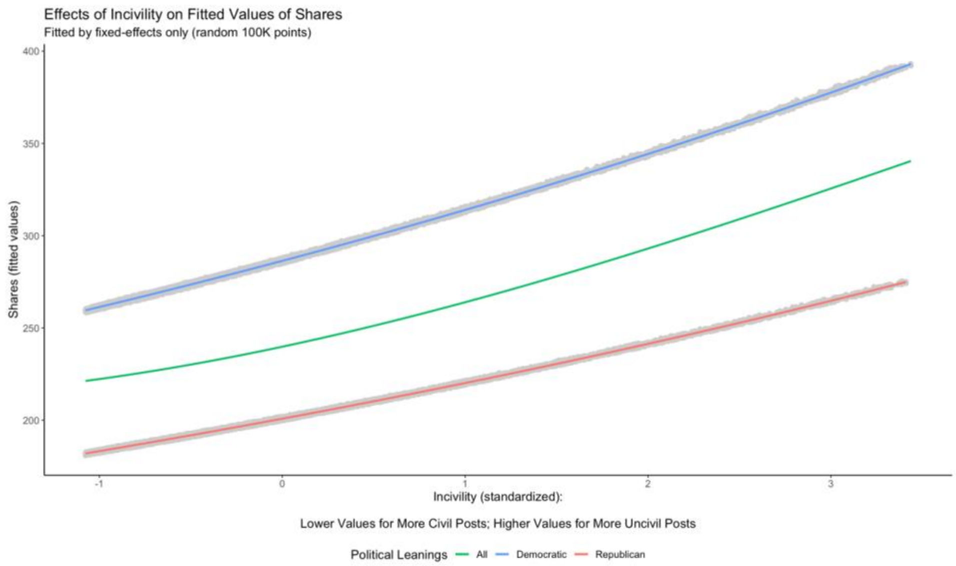


Figure 4.2.2: Comparing Effects of Incivility on the Fitted Values of Shares.

Angry reactions are more likely to be generated by negativity (*B* = -0.31, *SE* = 0.001, *p* < .001) than incivility (*B* = 0.17, *SE* = 0.001, *p* < .001). Figures 4.3.1 and 4.3.2 show overall negativity and incivility with the fitted Angry reactions from the model. As the sentiment of posts becomes more negative, the fitted values of Angry reactions increase. The slope of negativity is greater than that of incivility. Compared to the fitted values of Comments and Shares, it is notable that in Figure 4.3.2, Angry reactions display more banded slopes because of variations from political pages and the original scales of incivility (continuous between 0 to 1) compared to sentiment (ranging -4 to 4 by 1 unit scale changes) before standardizing and centering (i.e., the predicted estimate posts of incivility more spread out than those of sentiment do). H11c is not supported.

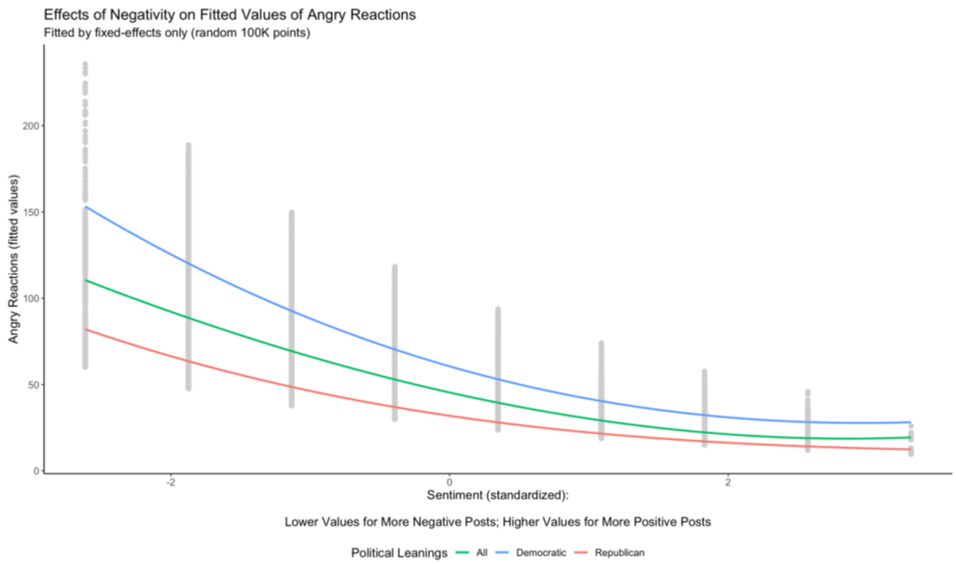


Figure 4.3.1: Comparing Effects of Negativity (Lower Values in Sentiment) on the Fitted Values of Angry Reactions.

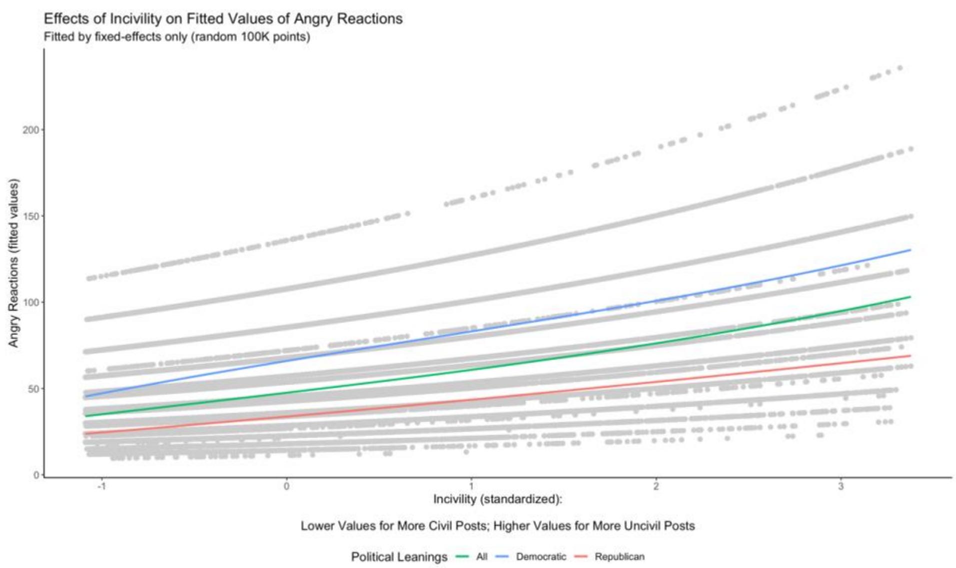


Figure 4.3.2: Comparing Effects of Incivility on the Fitted Values of Angry Reactions.

Findings from Table 4.3 answer RQ3. Although more *negative* posts yield fewer Like (B = 0.13, SE = 0.001, p < .001) and Love (B = 0.24, SE = 0.001, p < .001) reactions, *incivility* generates *more* Like (B = 0.09, SE = 0.001, p < .001) and Love (B = 0.02, SE = 0.001, p < .001) reactions. The magnitude of the effects of negativity are greater than the effects of incivility in terms of Like and Love reactions. Fitted values of Like reactions are presented in Figure 4.4.1 for negativity and in Figure 4.4.2 for incivility. Fitted values of Love reactions are displayed in Figure 4.5.1 for negativity and in Figure 4.5.2 for incivility. Note that fitted values of incivility (Figures 4.4.1 and 4.5.1) showed banded slopes due to pages' variations and differences of original scales between negativity and incivility.

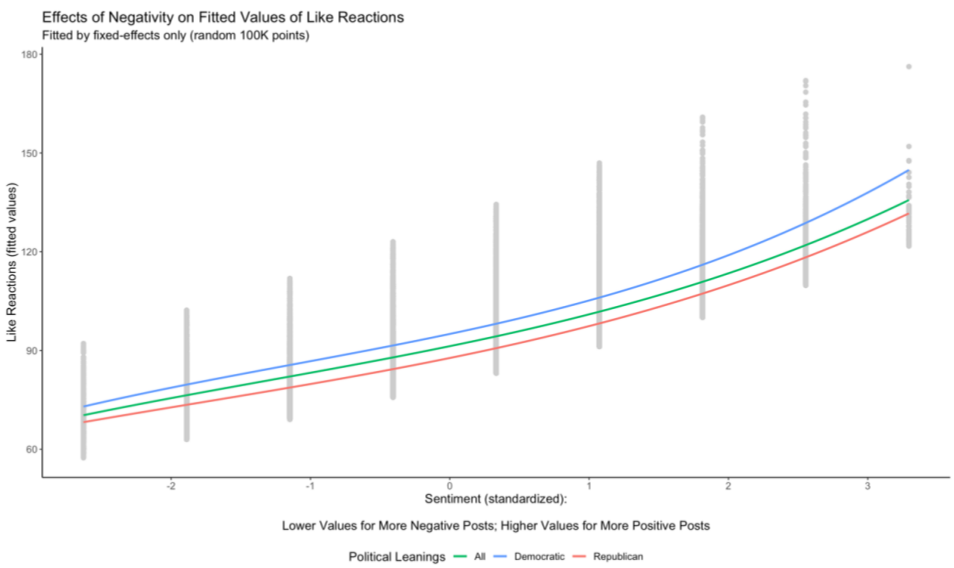


Figure 4.4.1: Effects of Negativity (Lower Values in Sentiment) on the Fitted Values of Like Reactions.

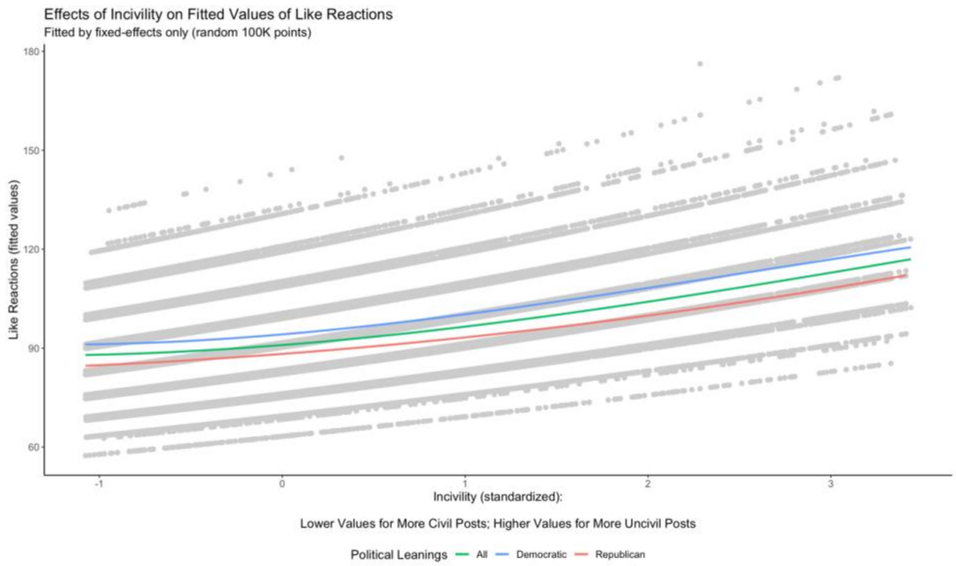


Figure 4.4.2: Comparing Effects of Incivility on the Fitted Values of Like Reactions.

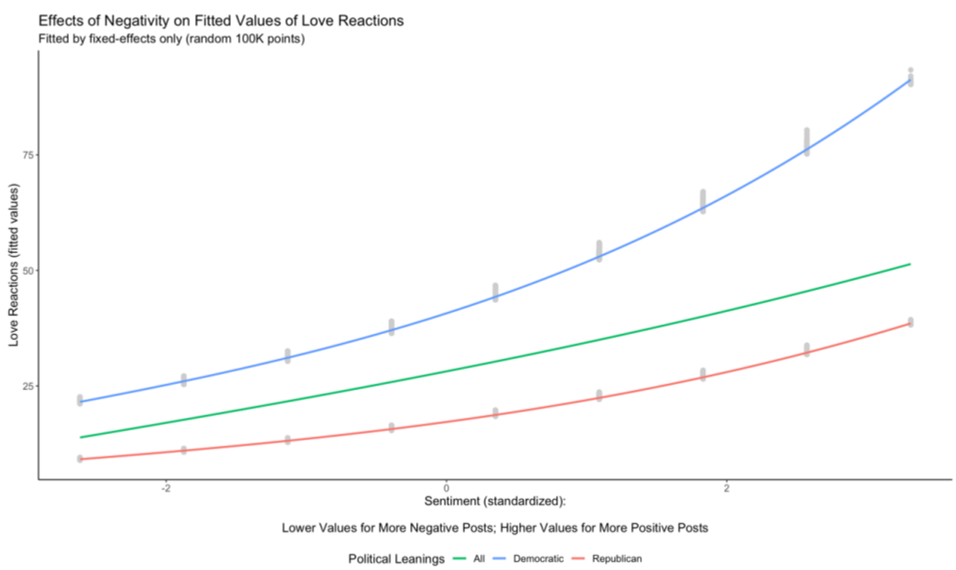


Figure 4.5.1: Effects of Negativity (Lower Values in Sentiment) on the Fitted Values of Love Reactions.

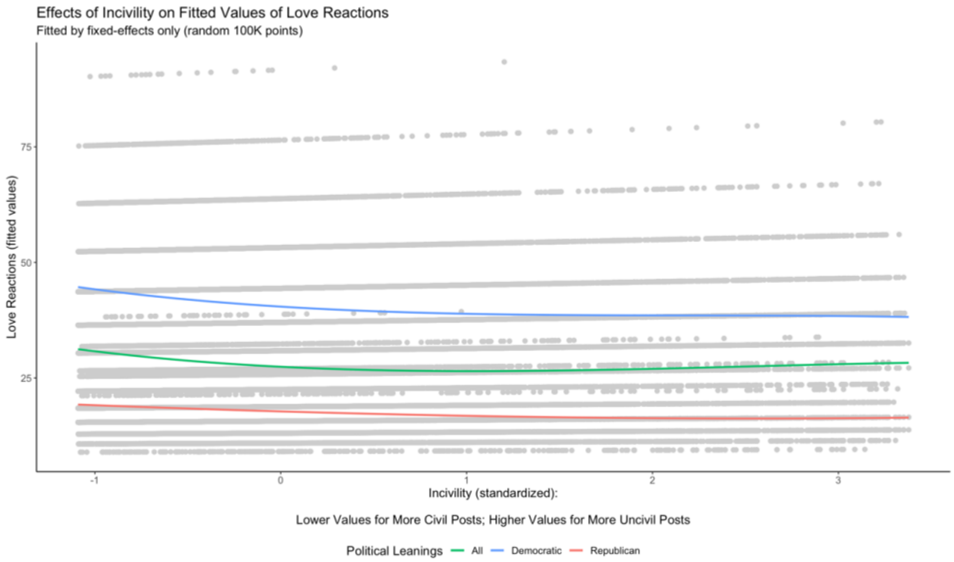


Figure 4.5.2: Comparing Effects of Incivility on the Fitted Values of Love Reactions.

In a similar manner, negativity shows greater effects on Sad (B = -0.31, SE = 0.001, p < .001) and Wow (B = -0.14, SE = 0.001, p < .001) reactions than incivility does (Sad B = 0.02, SE = 0.001, p < .001; Wow B = 0.08, SE = 0.001, p < .001). For better comparisons, Figure 4.6.1 displays fitted values of Sad Reactions by negativity and Figure 4.6.2 shows fitted values of Sad reactions by incivility. Figure 4.7.1 and Figure 4.7.2 also present the estimate of fitted values in Wow reactions by negativity and incivility separately.

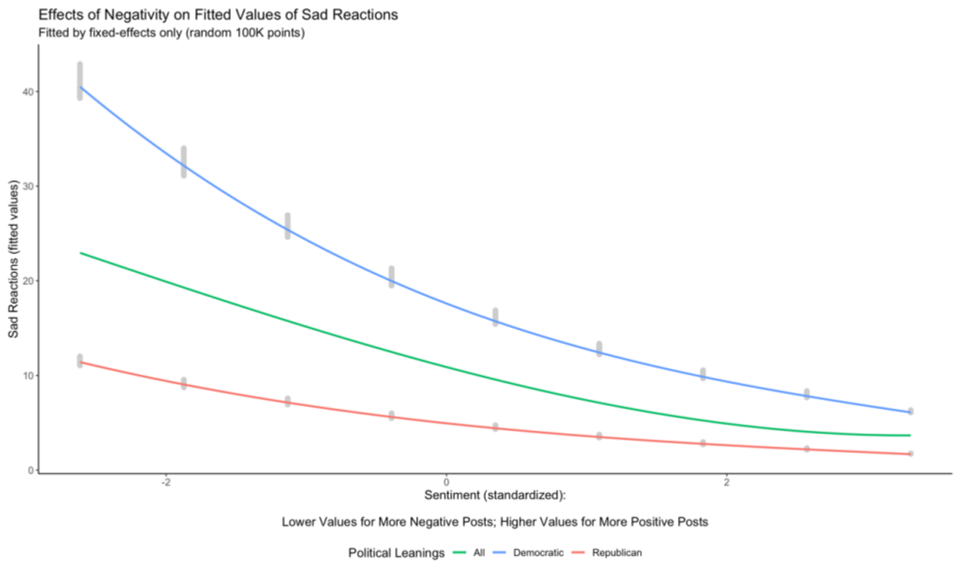


Figure 4.6.1: Effects of Negativity (Lower Values in Sentiment) on the Fitted Values of Sad Reactions.

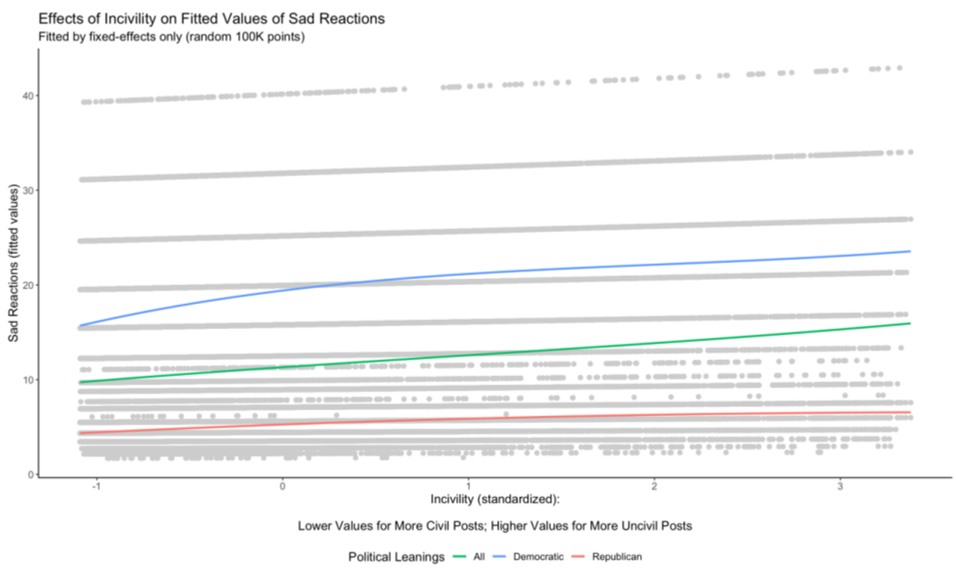


Figure 4.6.2: Comparing Effects of Incivility on the Fitted Values of Sad Reactions.

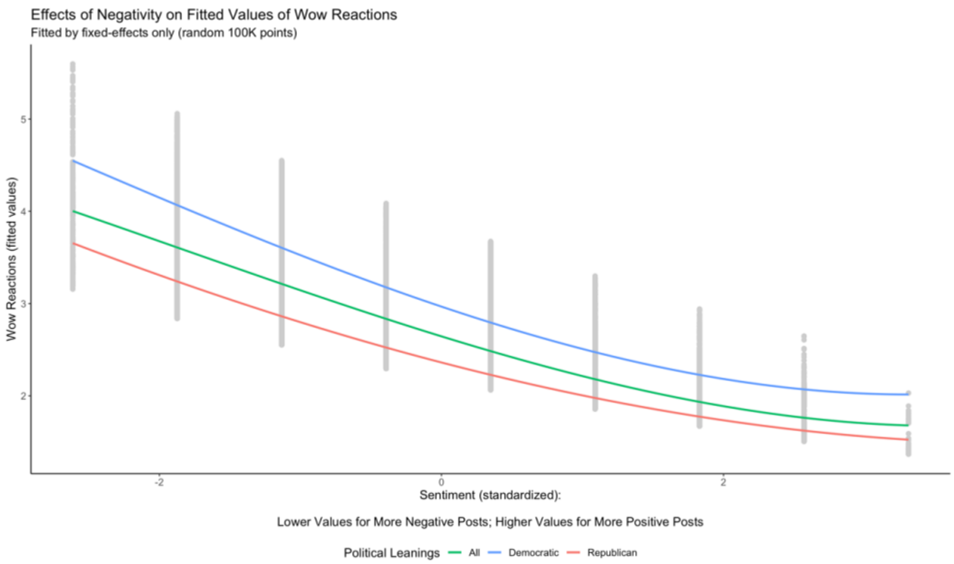


Figure 4.7.1: Effects of Negativity (Lower Values in Sentiment) on the Fitted Values of Wow Reactions.

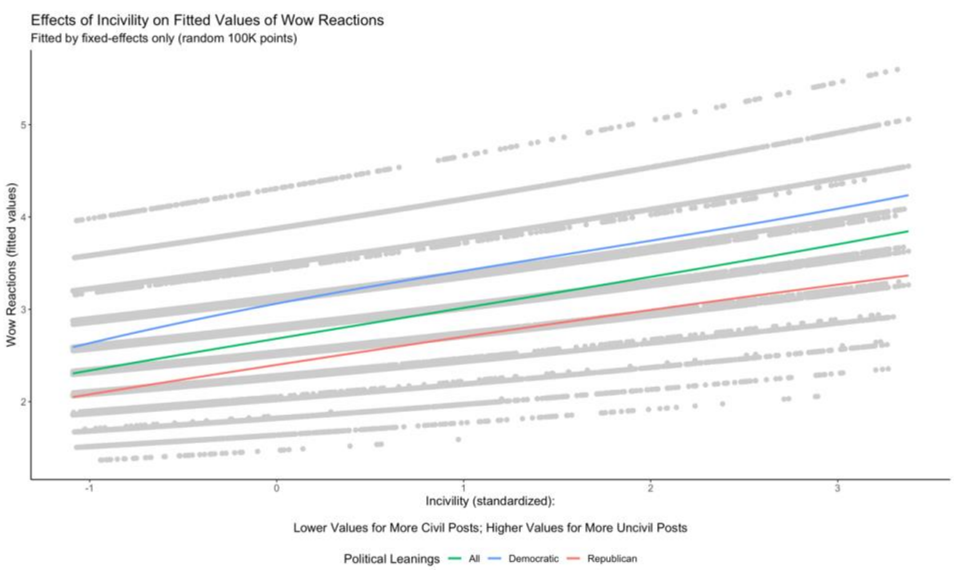


Figure 4.7.2: Comparing Effects of Incivility on the Fitted Values of Wow Reactions.

Finally, incivility has a greater impact on Haha reactions (B = 0.26, SE = 0.001, p < .001) than negativity does (B = 0.12, SE = 0.001, p < .001). Also, posts from Democratic pages are more likely to garner Haha reactions compared to posts from Republican pages (B = -0.48, SE = 0.001,

p < .001). The overall slope of incivility by fitted values of Haha reactions, which presented in Figure 4.8.2, is larger than that of negativity shown in Figure 4.8.1.

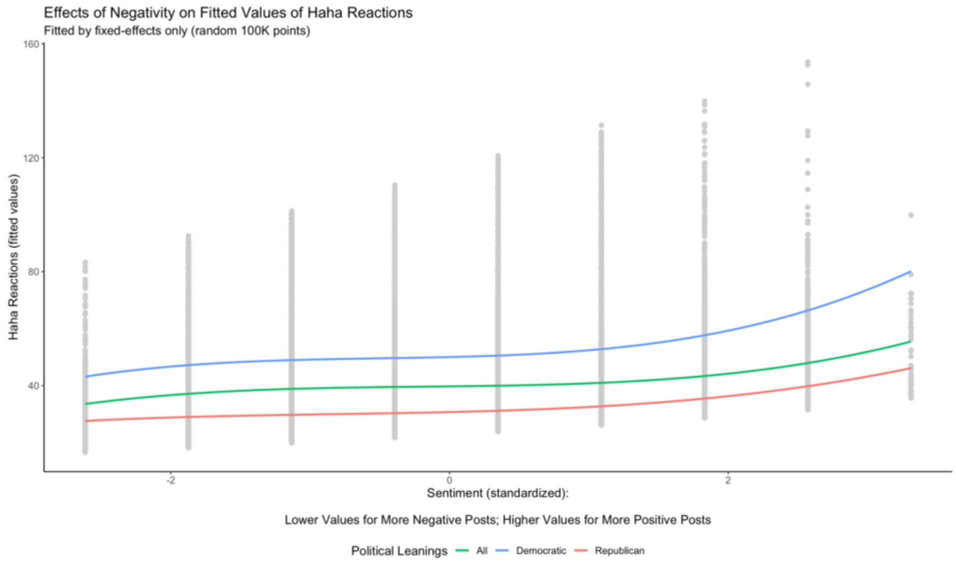


Figure 4.8.1: Effects of Negativity (Lower Values in Sentiment) on the Fitted Values of Haha Reactions.

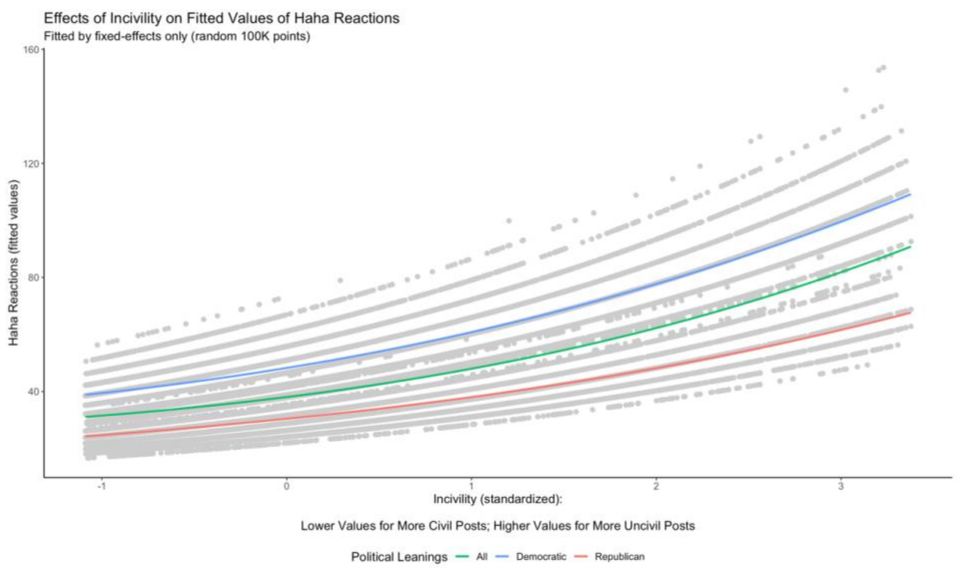


Figure 4.8.2: Comparing Effects of Incivility on the Fitted Values of Haha Reactions.

PARTISAN DIFFERENCES OF SOCIAL ENGAGEMENT

RQ4 and RQ5 examine partisan differences in the relationship between user engagement metrics and Facebook partisan content that is negative or uncivil. As shown in Table 4.4, models were tested by including two interaction terms: (1) political leanings of pages (i.e., whether pages were leaning towards Republicans/conservatives or Democrats/liberals) and sentiment of posts (i.e., lower sentiment means more negativity while higher sentiment refers to more positivity), and (2) political leanings on pages and incivility of posts. Note that following figures present fitted values of the interaction effect models by political leanings.

Table 4.4. Partisan Differences of Negativity and Incivility on Social Engagement.

	Comments	Shares	Angry	Love
	B (SE)	B (SE)	B (SE)	B (SE)
Fixed effects:				
Intercept	2.90*** (0.05)	5.63*** (0.05)	4.09*** (0.17)	3.71*** (0.12)
Sentiment	-0.03*** (0.002)	0.000 (0.001)	-0.30*** (0.001)	0.23*** (0.001)
Incivility	0.18*** (0.002)	0.09*** (0.001)	0.14*** (0.001)	0.04*** (0.001)
Republican	0.06 (0.07)	-0.31** (0.11)	-0.62*** (0.10)	-0.87*** (0.10)
Sentiment x Republican	-0.02*** (0.002)	0.01** (0.001)	-0.04*** (0.002)	0.02*** (0.002)
Incivility x Republican	-0.02*** (0.002)	-0.002 (0.001)	0.05*** (0.002)	-0.04*** (0.002)
Zero-inflation model:				
Intercept	-6.72*** (0.01)	-16.98*** (2.93)	-16.77** (5.47)	-17.03*** (4.73)
Random effects:				
(Page Name)				
Intercept (SD)	1.40	1.80	1.85	1.72
Number of Groups	1,493	1,493	1,491	1,491
Number of Obs	13,955,869	13,955,869	10,991,537	10,991,537
Log Likelihood	-43,446,716	-57,572,567	-30,787,577	-29,456,876
AIC	86,893,450	115,145,151	61,575,171	58,913,771
BIC	86,893,498	115,145,199	61,575,219	58,913,818
	Like	Sad	Haha	Wow
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Fixed effects:				
Intercept	4.55*** (0.05)	2.87*** (0.13)	3.85*** (0.22)	1.09*** (0.05)
Sentiment	0.11*** (0.001)	-0.29*** (0.001)	0.20*** (0.001)	-0.11*** (0.002)

Incivility	0.11*** (0.001)	0.04*** (0.001)	0.28*** (0.001)	0.08*** (0.002)
Republican	-0.08 (0.07)	-1.27*** (0.10)	-0.38** (0.13)	-0.23*** (0.07)
Sentiment x Republican	0.02*** (0.002)	-0.05***(0.002)	-0.14*** (0.002)	-0.05*** (0.002)
Incivility x Republican	-0.04*** (0.002)	-0.04*** (0.002)	-0.05** (0.002)	-0.01*** (0.002)
Zero-inflation model:				
Intercept	-18.97*** (3.52)	-17.01** (6.48)	-16.37*** (3.98)	-4.83*** (0.003)
Random effects:				
(Page Name)				
Intercept (SD)	1.33	1.70	2.12	1.26
Number of Groups	1,493	1,491	1,491	1,491
Number of Obs	13,955,869	10,991,537	10,991,537	10,991,537
Log Likelihood	-61,750,364	-21,946,437	-27,715,994	-23,847,943
AIC	123,500,745	43,892,892	55,432,006	47,695,904
BIC	123,500,793	43,892,939	55,432,054	47,695,952

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Log-Likelihood, AIC, and BIC are from the GLMMadaptive library for approximate estimates. Sentiment and incivility are standardized and centered. Lower sentiment indicates negativity while higher sentiment means positivity.

Metrics Where Engagement Varies by the Posts' Political Leanings

Whether posts are from Republican pages or Democratic pages greatly affects engagement with certain metrics. Compared to posts from Republican pages, posts from Democratic pages are more likely to garner Shares (B = -0.31, SE = 0.11, p < .01) and Angry (B = -0.62, SE = 0.10, p < .001), Love (B = -0.87 SE = 0.10, p < .001), Sad (B = -1.27, SE = 0.10, p < .001), Haha (B = -0.38, SE = 0.13, p < .01), and Wow (B = -0.23, SE = 0.07, p < .001) reactions. That posts from Democratic pages garner more engagement with respect to these metrics compared to posts from Republican pages should be taken into a consideration when interpreting the following figures.[15]

There are no significant differences by posts' political leanings on Comment and Like reactions. In this case, interaction effects between language attributes (negativity and incivility) and posts' political leanings show clearer visual comparisons.

[15] The patterns of significance are identical without the interactions as Table 4.3 shows.

Metrics Where the Effect of Negativity is Bigger for Republican Posts

After controlling for the main effects of negativity, incivility, and posts' political leanings, the interaction effects show that negative posts are more likely to influence Comments, Likes, and Wow reactions when they are from Republican pages compared to when they are from Democratic pages.

Partisan differences were demonstrated in the effects of negativity on Comments (interaction B = -0.02, SE = 0.002, p < .001). When posts were more negative, posts from Republican pages received more Comments compared to posts from Democratic pages. When posts are very positive, posts from Democratic pages gain larger number of Comments compared to posts from Republican pages. The fitted lines are displayed in Figures 4.9.

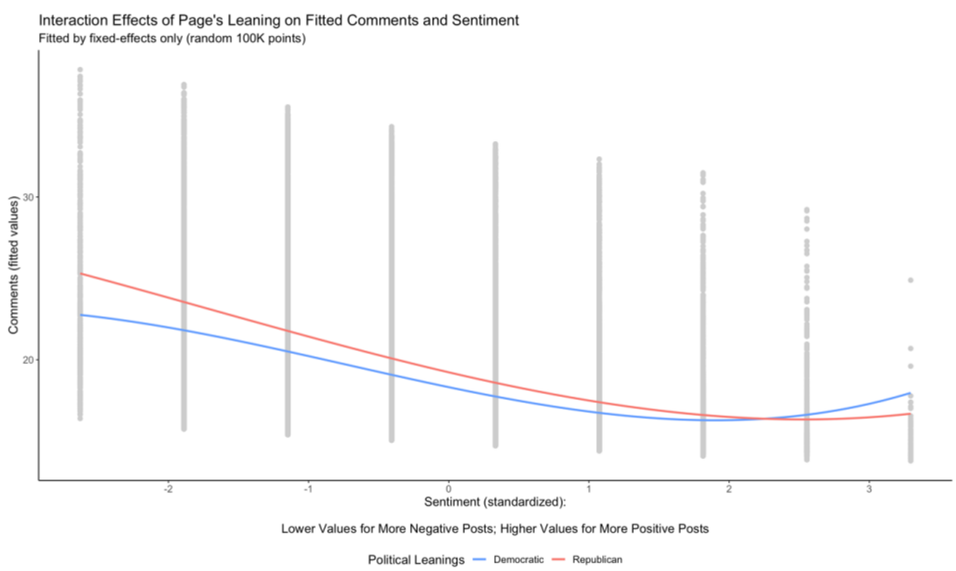


Figure 4.9: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values Comments and Negativity (Lower Values in Sentiment).

Negative posts from Republican pages also gain more Likes (B = 0.02, SE = 0.001, p < .01) than Democratic pages do. When putting both leanings together in a graph (Figure 4.10), the comparison is much clearer. The slope of the Republican line is greater than the line for Democrats, showing that negativity has a greater impact on Likes on Republican posts, even though Democratic posts are more likely to garner Likes overall.

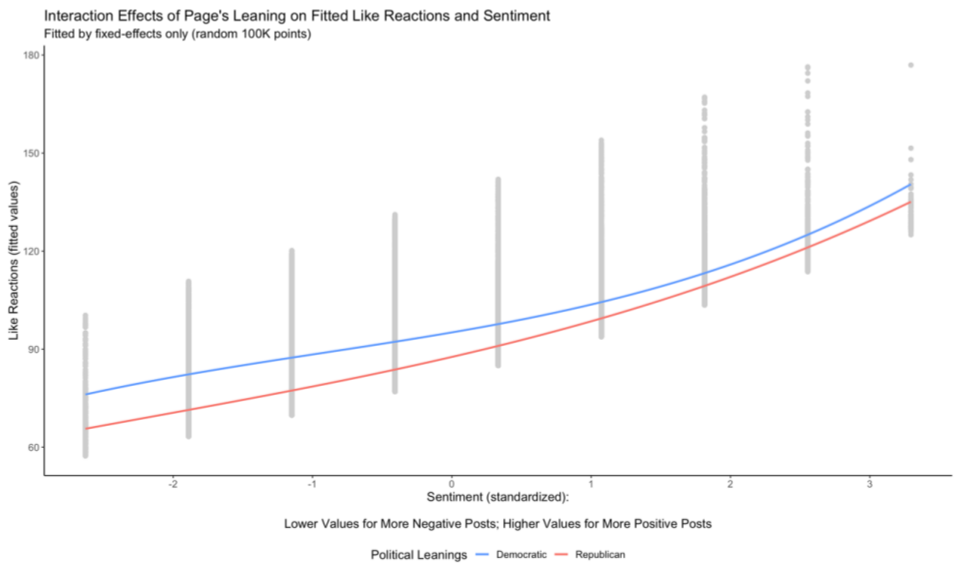


Figure 4.10: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values of Like Reactions and Negativity (Lower Values in Sentiment).

Next, there are also partisan interaction effects on Wow reactions in terms of the negativity of posts (B = -0.05, SE = 0.002, p < .001). Note that there is main effect of posts' political leanings, meaning that Democratic posts are more likely to engage with Wow reactions in general compared to Republican posts (B = -0.23, SE = 0.07, p < .001). Figure 4.11 shows the interaction effects of posts' leanings that higher increases of Wow reactions from Republican posts as posts become negative. The gap of the fitted Wow reactions between the two leanings becomes smaller because of greater increases from Republican posts as posts becomes negative.

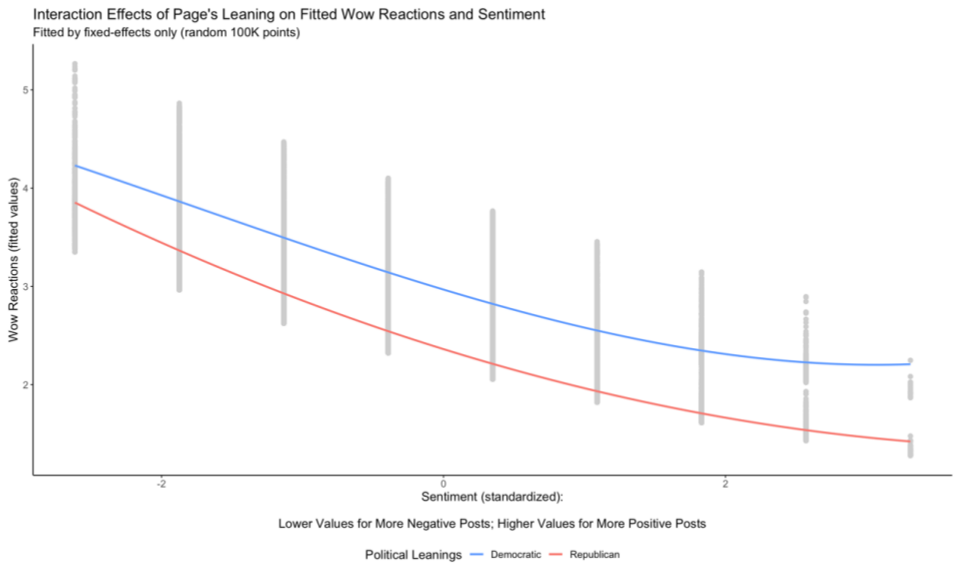


Figure 4.11: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values of Wow Reactions and Negativity (Lower Values in Sentiment).

Metrics Where the Effect of Negativity is Bigger for Democratic Posts

Shares and Angry, Love, Sad, and Haha reactions on posts from Democratic pages are more likely to be influenced by negative posts compared to those from Republican pages. Note that there are the main effects of posts' political leanings in the models, meaning that posts from Democratic pages are more likely to garner shares and these reactions compared to those from Republican pages (for Shares, B = -0.31, SE = 0.11, p < .01; for Angry reactions, B = -0.62, SE = 0.10, p < .001; for Love reactions, B = -0.87, SE = 0.10, p < .001; for Sad reactions, B = -1.27, SE = 0.10, p < .001).

In terms of Shares, partisan differences based on the posts' negativity are found (B = 0.01, SE = 0.001, p < .001). As posts become negative, posts from Democratic pages receive more Shares compared to posts from Republican pages, although the comparison does not display clearly because of the small effect (B = 0.01). When posts are very positive, Shares are slightly increased in posts of both political leanings. Shares with negativity have stronger effects on Democratic posts, which are illustrated in Figure 4.12.

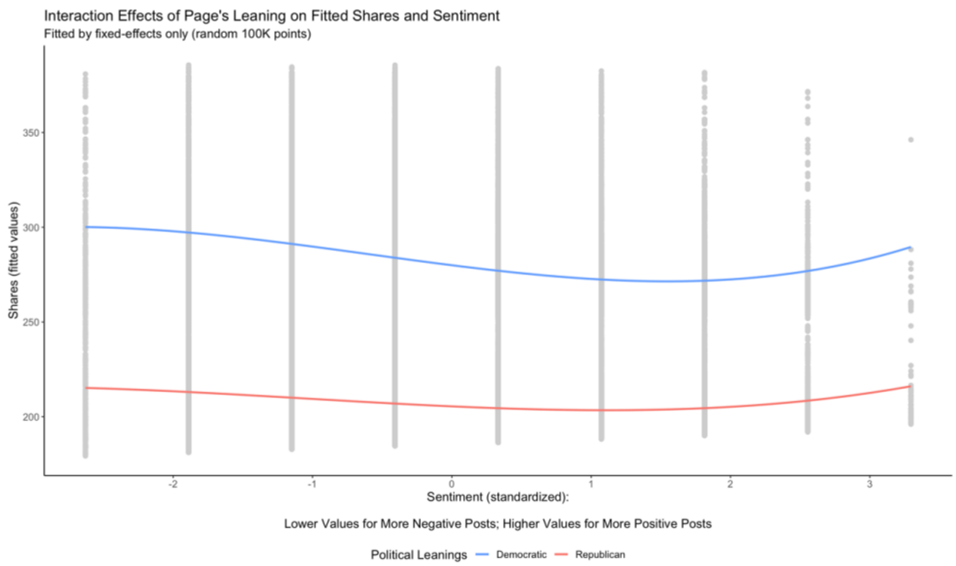


Figure 4.12: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values of Shares and Negativity (Lower Values in Sentiment).

There are partisan differences in the relationship between negativity and Angry reactions (B = -0.04, SE = 0.002, p < .001), as shown in Figure 4.13. When putting both political leanings together, the gap between Angry reactions by the posts' political leanings becomes larger as the posts become negative, showing that posts from Democratic Facebook pages receive more Angry reactions than those from Republican Facebook pages do as posts become more negative.

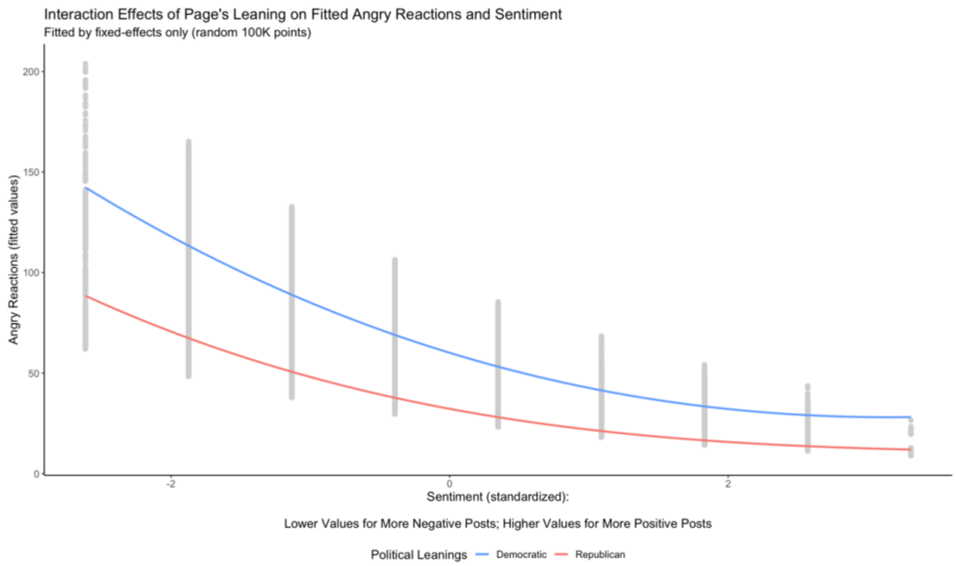


Figure 4.13: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values of Angry Reactions and Negativity (Lower Values in Sentiment).

For Love reactions, there are partisan differences based on the negativity of posts ($B = 0.02$, $SE = 0.002$, $p < .001$). As posts become more negative, posts from Democratic pages gain fewer Love reactions than posts from Republican pages do. The differences of political leanings in terms of posts' negativity are displayed in Figure 4.14, which correspond to overall greater engagement with Love reactions in posts from Democratic pages compared to those from Republican pages.

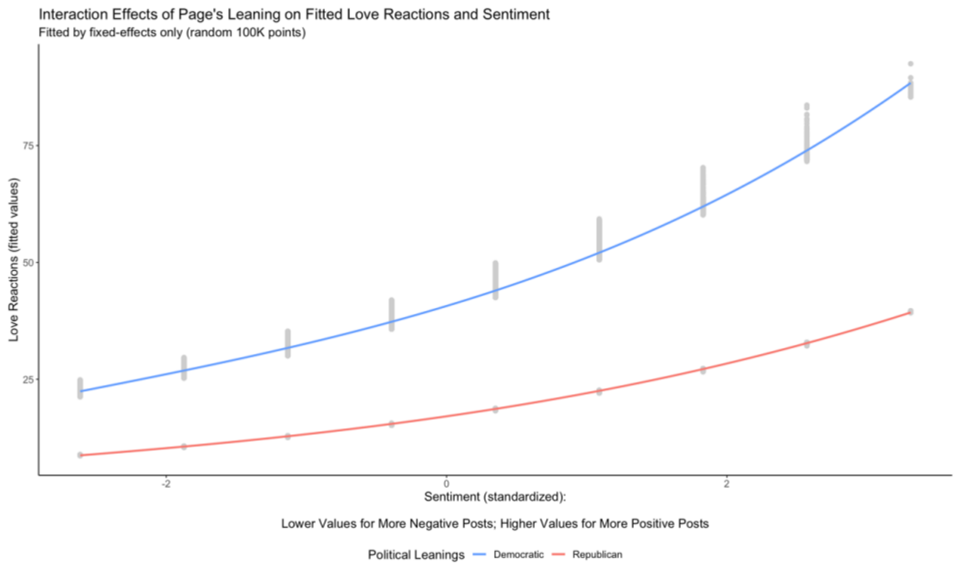


Figure 4.14: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values of Love Reactions and Negativity (Lower Values in Sentiment).

For Sad reactions, a partisan difference is also found with respect to negativity (B = -0.05, SE = 0.002, $p < .001$). As posts become more negative, posts from Democratic pages garner more Sad reactions compared to those from Republican pages, which present in Figure 4.15.

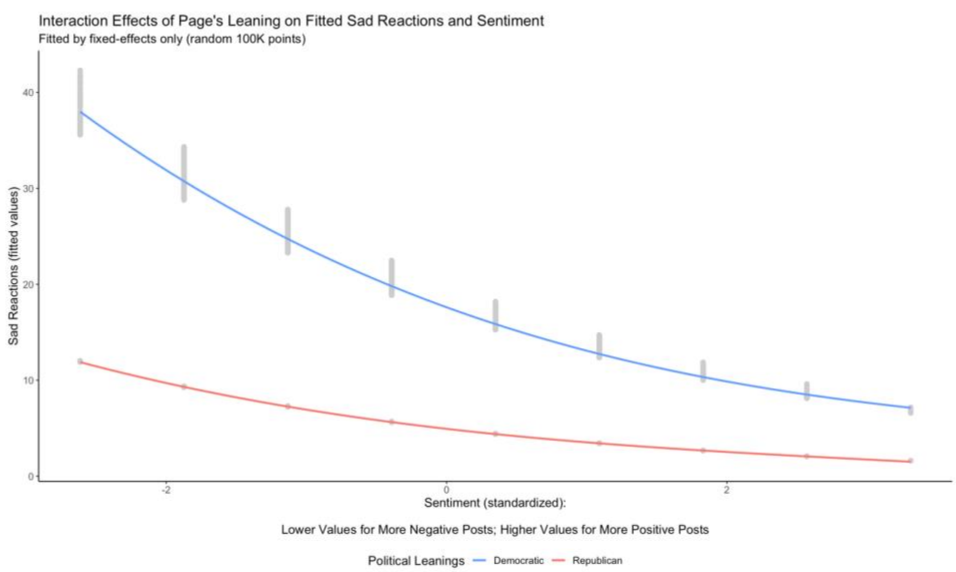


Figure 4.15: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values of Sad Reactions and Negativity (Lower Values in Sentiment).

Finally, Haha reactions also show interaction effects between posts' political leanings and negativity (*B* = -0.14, *SE* = 0.002, $p < .001$). Compared to posts from Republican pages, which garner a relatively constant level of Haha reactions regardless of the posts' negativity, as posts become more negative, Haha reactions on posts from Democratic pages decrease. These patterns are illustrated in Figure 4.16.

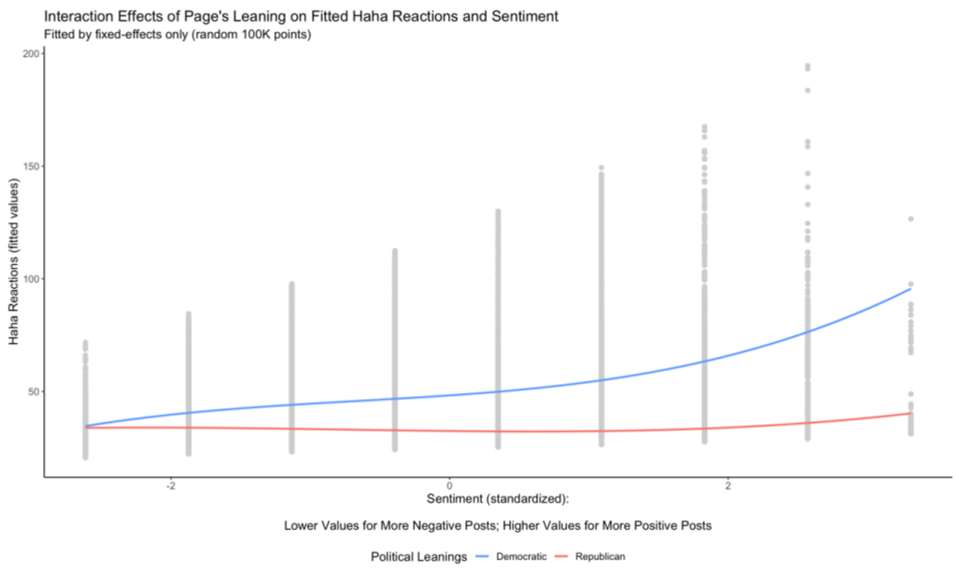


Figure 4.16: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values of Haha Reactions and Negativity (Lower Values in Sentiment).

Metrics Where the Effect of Incivility is Bigger for Republican Posts

Love reactions are predicted by an interaction between incivility and the political leanings of the page (*B* = -0.04, *SE* = 0.002, $p < .001$). Statistical findings indicate that Love reactions on posts from Republican pages decrease as posts become uncivil while Love reactions on posts from Democratic pages remain relatively constant as post become uncivil. As posts become civil on Democratic pages, Love reactions increase. Because the effects are very small, I looked at the estimates at two different incivility levels: on Democratic Facebook pages, posts with less than -1 on incivility are predicted to garner 40.55 Love reactions on average and posts with more than 3 on incivility are predicted to garner 41.66 Love reactions on average – a difference of -1.11. For Republican Facebook pages, posts with less than -1 on incivility are predicted to garner 18.70 Love reactions on average and posts with more than 3 on incivility are predicted to garner 14.87 Love reactions on average – a difference of -3.83. Because the Republican difference is greater than the Democratic difference in magnitude, it shows that the slope is greater for Republicans, indicating that posts from Republican pages lose more Love reactions than those from Democratic pages do as posts become uncivil. These patterns are displayed in Figure 4.17.

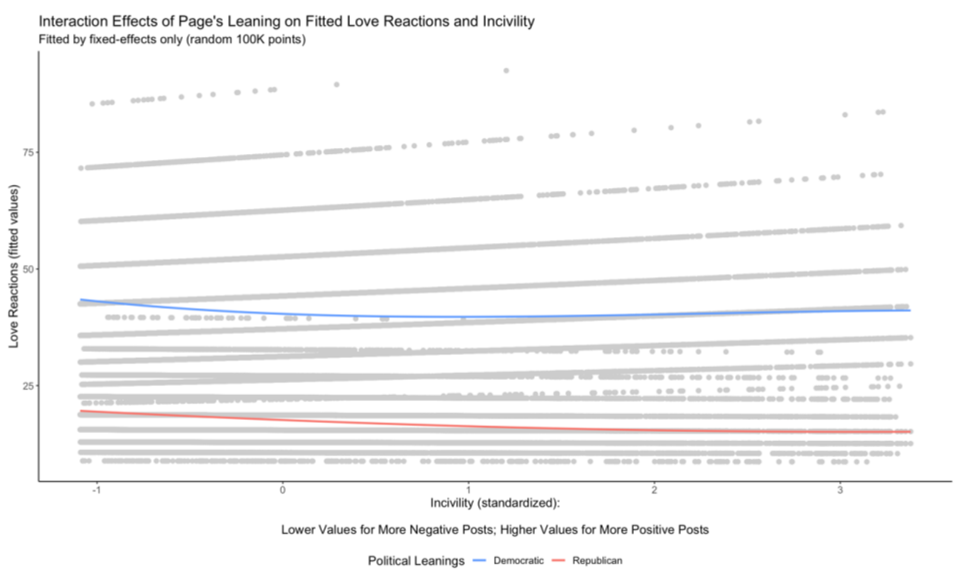


Figure 4.17: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values of Love Reactions and Incivility.

Metrics Where the Effect of Incivility is Bigger for Democratic Posts

Comments, Like, Angry, Sad, and Haha reactions show stronger effects of uncivil posts on Democratic pages than Republican pages. Posts from Democratic pages receive more Comments when posts were more uncivil ($B = -0.02$, $SE = 0.002$, $p < .001$) compared to those from Republican pages, although the interaction effect is small. The gap of Comments fitted values on posts between

Democratic pages and Republican pages becomes smaller as posts become more uncivil as shown in Figure 4.18.

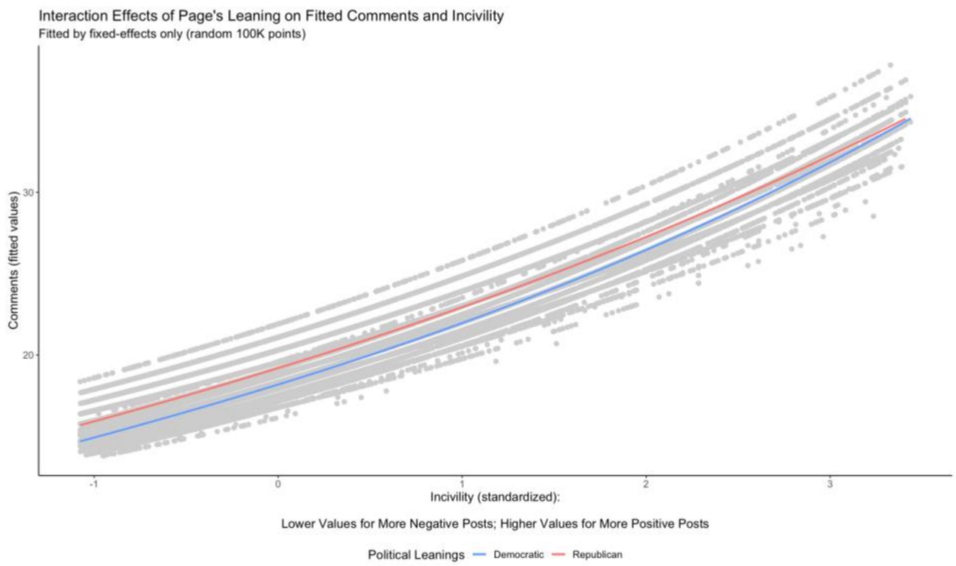


Figure 4.18: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values Comments and Incivility.

Like reactions also are predicted by an interaction effect between incivility and the political leanings of content (*B* = -0.04, *SE* = 0.002, $p < .001$). When posts become uncivil, posts from Democratic pages receive a greater number of Like reactions compared to those from Republican pages (see Figure 4.19).

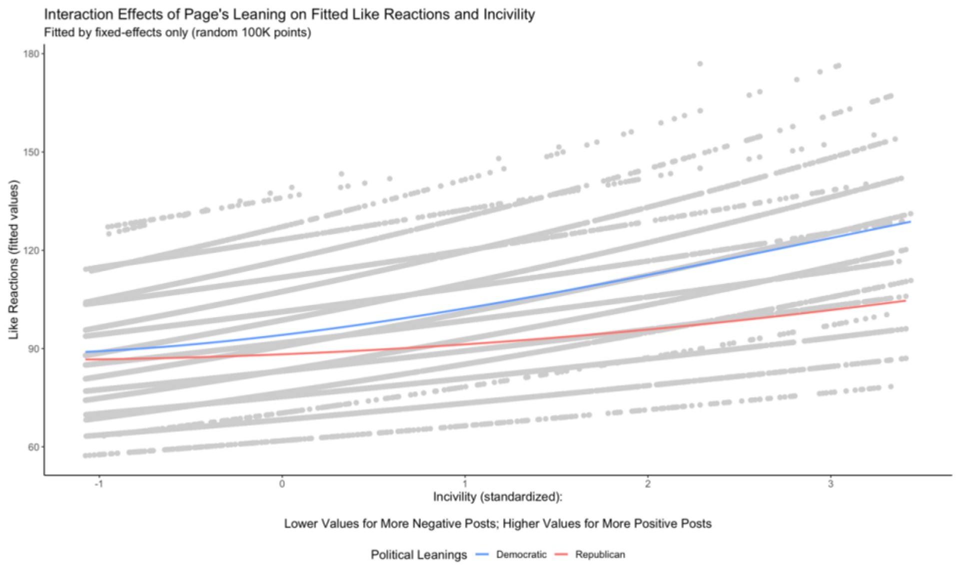


Figure 4.19: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values of Like Reactions and Incivility.

Comparison of the estimates at the two different incivility is reported for Angry reactions due to small estimates (*B* = -0.05, *SE* = 0.002, $p < .001$) as displayed in Figure 4.20. On Democratic Facebook pages, posts with less than -1 on incivility are predicted to garner 50.42 Angry reactions on average and posts with more than 3 on incivility are predicted to garner 113.33 Angry reactions on average – a difference of 62.91. For Republican Facebook pages, posts with less than -1 on incivility are predicted to garner 25.26 Angry reactions on average and posts with more than 3 on incivility are predicted to garner 73.07 Angry reactions on average – a difference of 47.81. Because the Democratic difference is greater than the Republican difference in magnitude, it shows that the slope is greater for posts on Democratic pages.

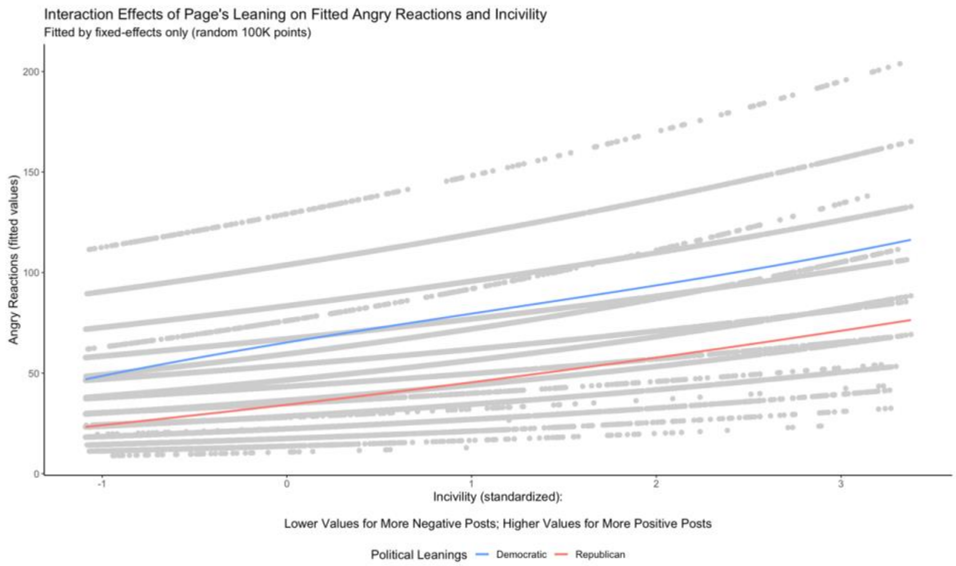


Figure 4.20: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values Angry Reactions Incivility.

Similarly, uncivil posts on Democratic pages received more Sad reactions (B = -0.04, SE = 0.002, p < .001) than those on Republican pages as Figure 4.21 demonstrates.

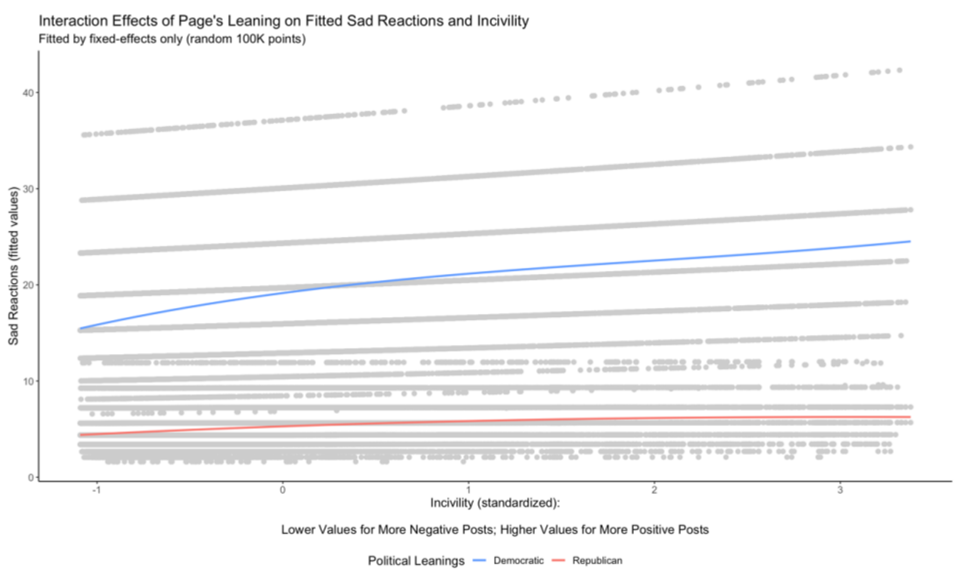


Figure 4.21: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values of Sad Reactions and Incivility.

Uncivil posts from Republican pages received a smaller number of Haha reactions than those from Democratic pages did (B = -0.05 SE = 0.002, p < .001) as shown in Figure 4.22.

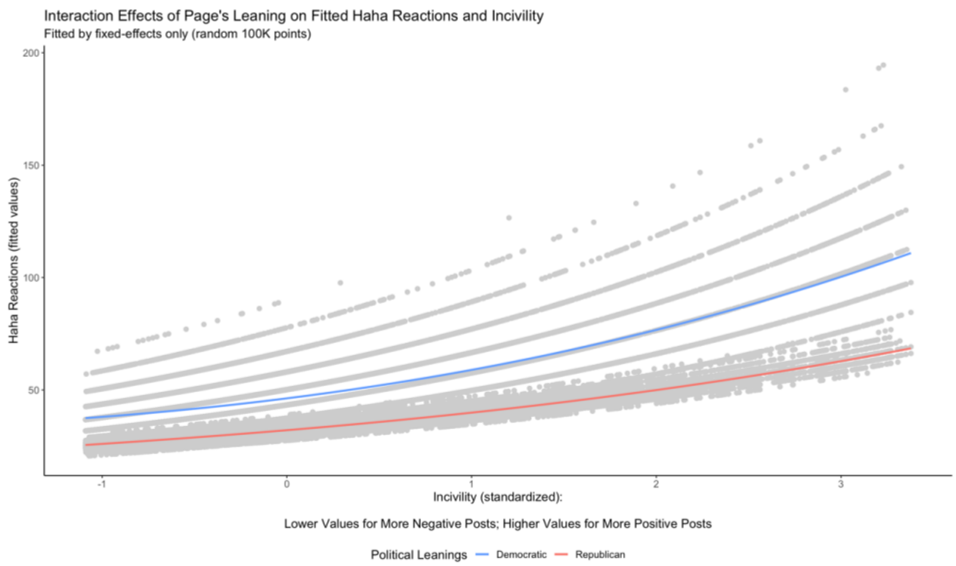


Figure 4.22: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values of Haha Reactions and Incivility.

For Wow reactions, there are partisan differences in reaction to incivility (B = -0.01, SE = 0.004, p < .001) although the effect is again small. The gap between the fitted lines and posts' political leanings becomes slightly larger as posts become uncivil (present in Figure 4.23), with a larger slope for posts on Democratic pages than Republican pages.

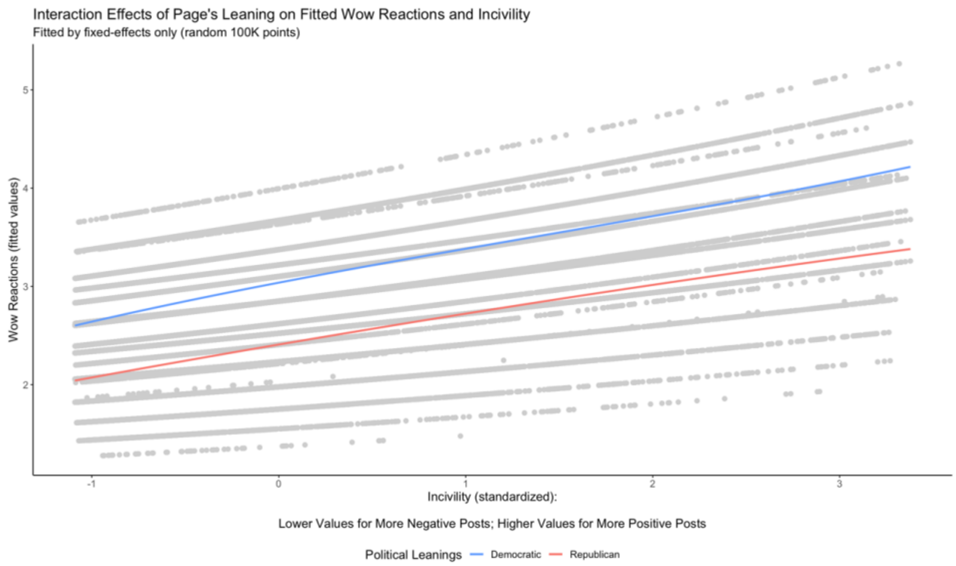


Figure 4.23: Interaction Effects of Page's Political Leanings on the Relationship between the Fitted Values of Wow Reactions and Incivility.

Metrics Where the Effect of Incivility by Political Leanings is Mixed

The interaction between the incivility of posts and political leanings of pages does not significantly predict the number of Shares (B = -0.002, SE = 0.001, p = .14).

DISCUSSION

This chapter examines whether and how negative and uncivil language play a role in promoting social engagement with posts on partisan pages. Negativity and incivility alone significantly affect all social engagement metrics, but whether those language attributes encourage or discourage engagement depended on the type of engagement. Relative to other reactions, commenting on and sharing posts requires greater motivation and effort. When it comes to negative or uncivil posts alone, people comment on and share those posts, which is consistent with evidence that negativity and incivility lead to engaging with online discourse (e.g., De León & Trilling, 2021; Masullo Chen & Lu, 2017; Wang & Silva, 2018). As more engagement happens, it leads to higher visibility due to Facebook algorithms. This supports a possible route to higher visibility of politically malignant content compared to harmless content. In terms of Comments and Shares, negativity and incivility work in the same directions of promoting partisan content.

Although relatively less effort is needed for clicking emotional reactions (e.g., Love and Angry), emotional reactions also contribute to the promotion of posts and may reinforce News Feed algorithms on Facebook. Looking at each separate reaction, how people react to negativity and incivility vary. For Angry, Sad, Wow, and Love reactions, negativity or incivility alone show similar patterns. To be specific, greater negativity and incivility garner more Angry, Sad, and Wow reactions on political content while they garner fewer Love reactions. Other research also finds that negative political news increases Angry and decreases Love reactions (Eberl et al., 2020; Jost et al., 2020), which are clear indications of disapproval and approval of political content. De León & Trilling (2021) further confirm the effects of negative coverage in the 2018 Mexican election on Facebook in response to Sad, Wow Angry, and Love reactions. They explain Wow reactions as negative expression of disbelief, instead of amazement. Because negative emotions include fear, sadness, anxiety, and anger, it makes sense that negative political content can lead to Sad and Wow reactions. When considering negativity includes incivility, uncivil posts can also ignite emotional responses that lead people to react with Sad and Wow. It is possible that Sad and Wow reactions can be seen as intended reactions by post creators. Partisan content consists of rhetorical devices such as hyperbole, metaphor, satire, and humor as well as adversarial and insulting language against the opposite party (Sturm Wilkerson et al., 2021). Rhetorical devices such as hyperbole and metaphor in negative and uncivil partisan content can incite Angry, Sad, and Wow reactions.

On the other hand, Like and Haha reactions show different patterns for negative versus uncivil posts. Uncivil posts encourage engagement with Like and Haha reactions and *positive* posts do the same. These findings demonstrate that incivility is different from negativity. In the same way as Sad and Wow reactions, Like and Haha reactions also can function as intended reactions contrived by political rhetoric such as metaphor, satire, and humor. Differences in how negativity and incivility work on partisan posts is important; here, incivility contributes to increased algorithmic visibility for these two reaction metrics.

Comparisons between the magnitudes of the effects of negativity and incivility suggest different dynamics in predicting social engagement. Negativity is a stronger influencer than incivility for encouraging Angry, Sad, Wow reactions and for discouraging Love and Like reactions. For instance, Table 4.5 shows some examples of negative, but not necessarily uncivil, posts that received a high number of Angry, Sad and Wow reactions. With more choices to express negative emotions (compared to when only Like existed before February 2016), negative content from hyper-partisan pages spread out and update News Feed with more visibility.

Table 4.5. Examples of Negative Posts by Social Engagement Metrics.

Leanings	Reaction (n)	Posts (some parts are truncated)
Republican	Angry (6,636)	You will be completely DISGUSTED! THIS is What Traitor Mccain Did Moments After He Killed Obamacare Repeal
Democratic	Angry (18,560)	SPEECHLESS 😠 This large group of voters arrived ON TIME to vote, but were locked out of the Expo Center due to a THIRTY-MINUTE WAIT TIME TO PARK after Kentucky shut down 95% of polling places and forced largely-black counties to vote in ONE polling location. THIS IS VOTER SUPPRESSION 😠😠😠 ... (truncated)
Republican	Wow (569)	Who do you think is responsible for this? Our vote is Hillary Clinton FBI Agent Suspected in Hillary Email Leaks Found Dead in Apparent Murder-Suicide An FBI agent believed to be responsible for the latest email leaks was found dead in an apparent murder-suicide early Saturday morning, according to police.
Democratic	Wow (1824)	BREAKING NOW: Treasury Officials Just Found Russian Money Trail to Trump Campaign Investigators followed the money and found something devastating.
Republican	Sad (5,032)	She is sounding off and furious! Are you with her? Boston bombing survivor who lost leg "livid" after Tsarnaev death sentence overturned ...

Democratic	Sad (6,240)	Hurricane Michael has destroyed lives, homes, schools, marinas, and a whole town. Mexico Beach, Florida: "It's Gone" - News & Guts Media When the sun came up today in Florida and Georgia the devastation was worst than people ever imagined. The drone footage above was taken this morning in Panama City. Hurricane Michael was as ferocious as they come. The worst hurricane Florida's panhandle has ever seen. It came ashore as a category...

For Comments, Shares, and Haha reactions, on the other hand, incivility is a stronger predictor than negativity of engaging with political posts. As discussed earlier, incivility on political content arouses an uncivil response in subsequent comments (Kim et al., 2021; Masullo Chen & Lu, 2017) and increases chances of being shared (Wang & Silva, 2018). Also, Haha reactions function differently from other emotional reactions when taking into consideration both negativity and incivility: negativity discourages Haha reactions (instead, positivity encourages of Haha reactions) but incivility promotes Haha reactions. Uncivil posts can include humor, satire, and sarcasm, which may increase Haha reactions and play a role in inducing higher visibility on Facebook News Feed. This pattern can be problematic because satirical news is more likely to make people engage in partisan selective exposure (Knobloch-Westerwick, & Lavis, 2017; Stroud & Muddiman, 2013). Table 4.6 displays example posts that received the greater number of Comments, Shares and Haha reactions, which are the most uncivil. Taken together, commenting, sharing and Haha reactions affect the priority of uncivil partisan content, which may reinforce one's algorithms for political information, and lead to more exposure to this type of content.

Table 4.6. Examples of Uncivil Posts by Social Engagement Metrics.

Leanings	Reaction (n)	Posts (some parts are truncated)
Republican	Comments (13,967)	Rob Reiner Claims President Trump 'Is Causing People in NY To Die' - READ THE STORY: (truncated) "It's no surprise that this President only cares about himself. But goddamn, people are dying. If you can't be human, get the f** out of the way and

		let people who are take over" -Rob Reiner POLITICAL INSIDER
Democratic	Comments (22,180)	YOU CAN'T FIX STUPID: WATCH: Amidst exploding COVID cases in Texas, GOP Texas Lt. Gov Dan Patrick goes on Fox "News" and unleashes on public health expert Anthony Fauci, saying "He doesn't know what he's talking about. I don't need his advice anymore." Texas lt. governor hits Fauci: 'He doesn't know what he's talking about' Texas Lt. Gov. Dan Patrick (R) on Tuesday lambasted Anthony Fauci over his assessment of the recent surge in coronavirus cases in the U.S., claiming that his state did not need advice from the nation's top infectious disease expert.
Republican	Shares (76,424)	Timeline Photos BEFORE YOU TRASH THIS MAN GOOGLE 'CLINTONSCANDALS SCA
Democratic	Shares (147,865)	This one says it all... Timeline Photos TRUMP SUPPORTERS KEEP MESSAGING US THAT THEY "CAN'T WAIT TO GET AMERICA BACK!" BACK TO WHAT? SEGREGATION? COAT-HANGER ABORTIONS? INTERNMENT CAMPS? BAREFOOT AND PREGNANT? LYNCHINGS? CHILD LABOR? WHITE SUPREMACY? SERIOUSLY, EITHER SAY WHAT YOU MEAN, OR SHUT THE HELL UP.
Republican	Haha (20,942)	YIKES! 😂😂😂 #ThinkForYourself THE NEXT TIME YOU FEEL STUPID OR UNINFORMED... TURNING POINT USA **中 JUST REMEMBER, THERE ARE PEOPLE WHO GET THEIR NEWS FROM HIM
Democratic	Haha (20,109)	Only an idiotic trump voter would cheer Obamacare repeal. A Trump Fan Cheering Obamacare Repeal Just Found Out He's On Obamacare. Hilarity Ensues He couldn't handle the truth.

Finally, I explored partisan social engagement differences in response to negative and uncivil posts. One overarching finding is that posts from Democratic pages garner more engagement for nearly all metrics except for Comments and Like reactions compared to posts from Republican pages. Although interaction effects between language attributes and posts' political leanings are significant (except for incivility and Shares), estimates of the interaction effects are mostly small.

I looked at how differently partisans react to political content through social engagement. Compared to posts on Democratic pages, those on Republican pages lose more Likes and gain more Comments and Wow reactions when posts become negative. On the other hand, as posts become negative Democratic posts receive fewer Love reactions and more Shares and Angry, Sad, and Haha reactions compared to Republican posts. When political content is negative, Republicans are more likely to engage in relatively traditional ways such as Likes and Comments (but not Shares) while Democrats are more likely to engage in relatively new ways of interacting via the more recently introduced reactions (except for Wow) and Shares. Although there are some exceptions (Wow reactions for Republican posts and Shares for Democratic posts), reactions to political content may reflect openness to new ideas, which is one of the perspectives that explains ideological differences. Newer technical features on Facebook may be more embraced by Democrats and liberals because they have more "fluid" outlooks rather than "fixed" outlooks; in other words, Democrats may be excited by new things while it may be more likely for Republicans and conservatives to feel comfortable with familiar and predictable ways of engaging on the platform (Hetherington & Weiler, 2018). Wow reactions, however, saw a sharper slope for posts on Republican Facebook pages relative to Democratic Facebook pages. Because Wow reactions often indicate disbelief rather than amazement on political content (De León & Trilling, 2021), Republicans may prefer Wow reactions to express disbelief when reacting to negative political posts. On the other hand, Shares are more responsive to negativity for posts on Democratic pages, an effect which happens in two ways. As posts become negative and when posts are very positive, Democrats are more likely to share political content compared to posts with middle levels of sentiment while Republicans' sharing behavior is not predicted by its sentiment. Prior work on Twitter finds that liberal users are more likely to engage in cross-ideological dissemination

(Barberá, et al., 2015), which could explain the greater share rate of Democratic posts – even if it is negative content toward Democratic party.

In terms of uncivil political content, posts on Democratic pages garner more Comments and Like, Angry, Sad, Haha, and Wow reactions as incivility increases relative to posts on Republican pages. This is consistent with findings that toxic comments attract more Likes on liberal outlets than conservative or neutral outlets (Kim et al., 2021). Partisan differences for Love reactions are small. Shares of uncivil political content do not show partisan differences.

In closing, the main findings of this chapter can be summarized as follows: negativity and incivility, in general, tend to promote engagement with partisan posts. Yet, the magnitude of negativity and incivility varies by the metric and partisanship of the page. These phenomena may boost the presence of such political content in personal News Feeds, which I investigate in the next chapter as possibly leading to polarization. Chapter 5 will examine whether language attributes in political content correspond with affective polarization among the public.

Chapter 5: Does Negative and Uncivil Language Affect Partisan Polarization in the American Public?

Jesse ardently follows and participates in politics. She votes in every election and encourages others to do the same. She subscribes to online news and watches cable news that aligns with her world view. Lately, she has attended protests and marches to support what she believes is right.

Taylor, on the other hand, is a different sort of citizen. She is also passionate about politics but her major playground is social media. She often leaves comments on news that she encounters, which mostly comes from several Facebook pages that she follows. She likes to share political memes with her friends because she thinks it's hilarious. She constantly tags her posts with hashtags on Twitter. She checks social media on her phone more than 12 times a day.

Who do you think influences the public and democracy more? Whatever your answer is, it is likely that Jesse and Taylor are not two different people. Rather, it is likely that Jesse would do the same activities that Taylor does as an active political junkie in contemporary times. What people like Taylor create, share, and engage with is hyper-partisan content that is mostly biased in favor of one's own groups and against opposing groups. I argue that one particular type of political content – negative and uncivil content – leads to polarization among the public. People are exposed to negative and uncivil language in political content that spreads thanks to networked algorithms. By examining negativity, incivility, and partisan targets in political posts, I assess whether language affects affective polarization among the public.

CHAPTER OVERVIEW

In the first two sections of this chapter, I investigate whether negative and uncivil content correspond with affective polarization in the public (H4 and H9). In the third section, I examine

H12, which proposes that incivility will have stronger effects on aggregated affective polarization than negativity. The estimates of negativity and incivility are compared through statistical tests. In the fourth section, how political targets mentioned in posts with negative and uncivil content affect polarization toward the two parties is investigated. As H5 and H10 proposed, the effects of posts mentioning out-groups only and mentioning both in- and -out groups may predict affective polarization. In the fifth section, I analyze how the effects of negative (RQ6) and uncivil (RQ7) content accompanied by political targets within posts vary for Democratic and Republican pages. In the final section, this chapter's findings are interpreted and discussed.

RESULTS OF WHETHER NEGATIVE LANGUAGE CORRESPONDS WITH POLARIZATION

H4 proposed that lagged and aggregated negativity in posts on partisan Facebook pages will correlate with aggregated affective polarization. This section will analyze the absolute value of the affective polarization measures for both partisan sentiment and feeling thermometer, so that higher values indicate more polarization and smaller values indicate less polarization irrespective of whether the polarization favors Democrats or Republicans. The residuals of affective polarization measures are modified by the best ARFIMA model fit. With exogenous variables of aggregated negativity by quarter, lagged negativity, and economic perceptions as a control variable (CCI), lagged negativity did not predict affective polarization using the ARFIMA residuals (0, 0.46, 1) for partisan sentiment (B = -7.56, SE = 16.87, p = .66) or using the ARFIMA residuals (1, 0.33, 2) for the feeling thermometer measure (B = 13.33, SE = 23.07, p = .57). No significant contemporaneous relationships were found for negativity in posts from partisan pages on either the partisan sentiment or feeling thermometer measures. Aggregated negativity in posts on partisan pages alone does not have an impact on affective polarization. The results are summarized in Table 5.1.1.

Table 5.1.1: Effects of Lagged Negativity of Partisan Content on Aggregated Absolute Value of Affective Polarization.

Aggregated Sentiment by All Posts	Partisan Sentiment	Partisan Feeling Thermometer
	B (SE)	B (SE)
Intercept	54.75 (69.10)	-110.66 (94.68)
Sentiment	10.40 (18.75)	7.28 (24.95)
Sentiment (lagged)	-7.56 (16.87)	13.33 (23.07)
CCI	-0.56 (0.69)	1.10 (0.94)
R^2	0.04	0.04
Number of Observations	52	47

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Sentiment is standardized and centered before aggregation. Lower sentiment indicates negativity while higher sentiment means positivity.

I also looked at the effects of aggregated negativity by the political leanings of the posts on affective polarization, using the relative sentiment and feeling thermometer measures in Table 5.1.2. Here, as values move away from zero, affective polarization favors one political party such that more positive values indicate greater favorability toward Republicans and more negative values indicate greater favorability toward Democrats. Lagged negativity of posts from Democratic pages marginally affects the residuals of ARFIMA (1, 0.38, 0) on the feeling thermometer measure (B = -48.89, SE = 28.67, p = .096), which is presented in Figure 5.1. An increase of lagged negativity of posts from Democratic pages leads to more favorable feelings toward the Republican Party. Lagged positivity of posts from Democratic pages is relatively constant on the fitted value of affective polarization (residuals of partisan relative feeling thermometer analysis). I offer possible explanations for this at the close of this chapter. The estimate of lagged negativity of posts from Democratic pages on the residuals of ARFIMA (0, 0.45, 1) on relative sentiment is in the same direction, but not statistically significant (B = -15.11, SE = 22.05, p = .50). Lagged negativity of posts from Republican pages is not significant for either the relative sentiment (B = 22.20, SE = 18.19, p = .23) or feeling thermometer measures (B = 28.39, *SE* = 21.04, *p* = .18). Although there is some evidence of negative posts on Democratic pages affecting polarization, H4 is not supported.

Table 5.1.2: Effects of Lagged Negativity of Partisan Content on Aggregated Relative Value of Affective Polarization by Political Leanings.

Aggregated Sentiment by Post Leanings	Partisan Sentiment	Partisan Feeling Thermometer
	B (SE)	B (SE)
Intercept	-59.78 (78.25)	152.61 (110.02)
Sentiment of Democratic Posts	-2.96 (22.54)	5.95 (28.09)
Sentiment of Democratic Posts (lagged)	-15.11 (22.05)	-48.89^{+} (28.67)
Sentiment of Republican Posts	-8.43 (18.85)	-14.94 (21.89)
Sentiment of Republican Posts (lagged)	22.20 (18.19)	28.39 (21.04)
CCI	0.61 (0.78)	-1.53 (1.10)
R^2	0.09	0.12
Number of Obs	52	47

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Sentiment is standardized and centered before aggregation. Lower sentiment indicates negativity while higher sentiment means positivity.

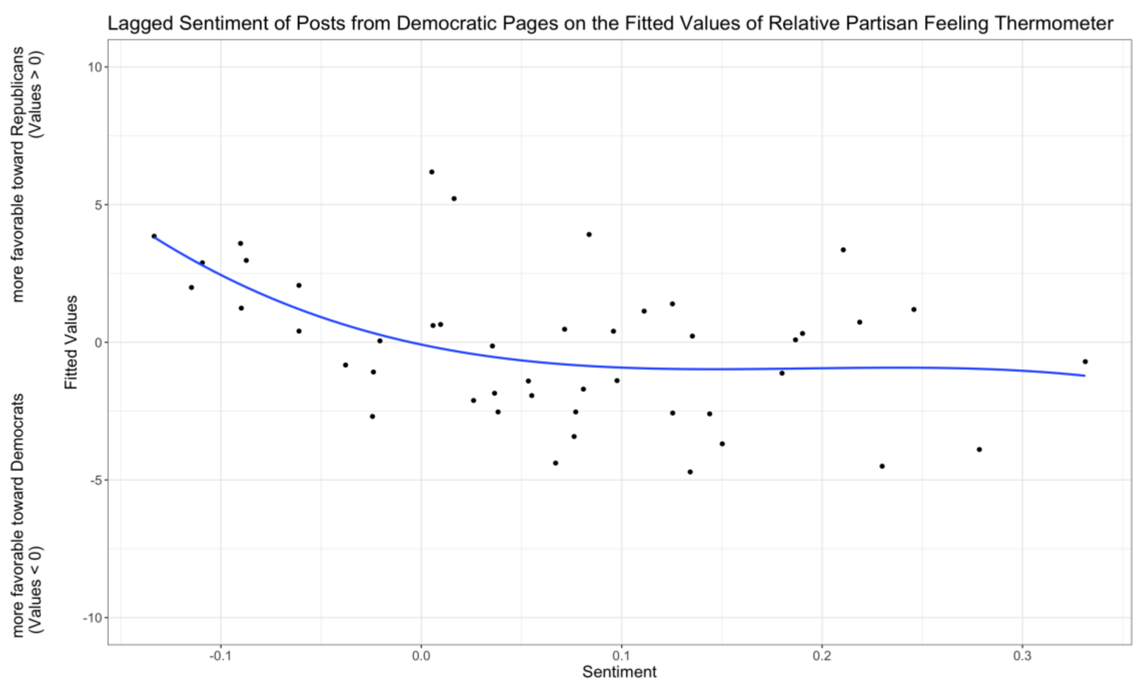


Figure 5.1: Lagged Sentiment of Posts from Democratic Pages on the Fitted Values of Affective Polarization (Partisan Relative Feeling Thermometer).

RESULTS OF WHETHER UNCIVIL LANGUAGE CORRESPONDS WITH POLARIZATION

H9 asked whether the incivility of partisan content posted on Facebook led to affective polarization, which is captured by the absolute value of the affective polarization with higher values meaning more polarization and smaller values meaning less polarization. With external regressors of aggregated incivility by quarter, lagged incivility, and economic perceptions as a control variable (CCI), there is no statistically significant evidence of a relationship between lagged incivility and affective polarization (using partisan sentiment as residuals of ARFIMA (0, 0.46, 1), B = 9.17, SE = 20.31, p = .65; or using the partisan feeling thermometer as residuals of ARFIMA (1, 0.33, 2), B = 12.80, SE = 27.25, p = .64). No significant relationship is found for contemporaneous incivility on either the partisan sentiment or feeling thermometer measures. Aggregated incivility by all posts on partisan pages alone does not influence affective polarization. Results are summarized in Table 5.2.1.

Table 5.2.1: Effects of Lagged Incivility of Partisan Content on Aggregated Absolute Value of Affective Polarization.

Aggregated Sentiment by All Posts	Partisan Sentiment	Partisan Feeling Thermometer
	B (SE)	B (SE)
Intercept	46.43 (96.70)	25.57 (140.12)
Incivility	-11.92 (20.89)	-9.79 (27.90)
Incivility (lagged)	9.17 (20.31)	12.80 (27.25)
CCI	-0.48 (0.96)	-0.25 (1.39)
R^2	0.04	0.01
Number of Observations	52	47

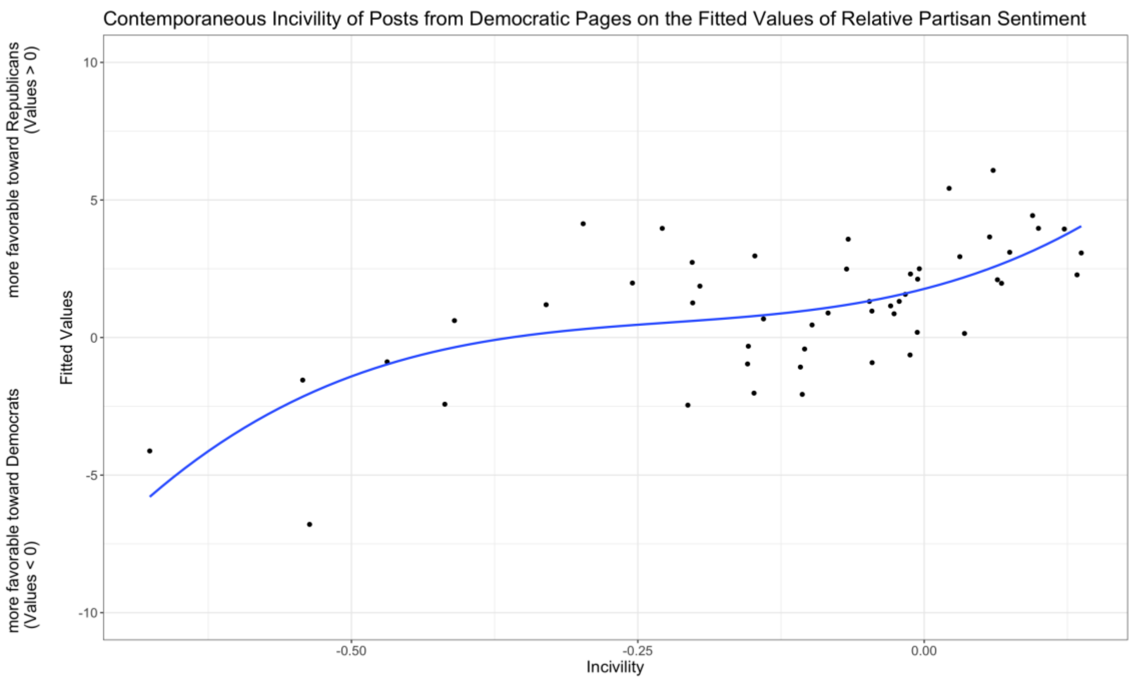
Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Incivility is standardized and centered before aggregation.

As with sentiment, I further tested H9 by including aggregated uncivil posts and examining polarization based on the relative measure. There is no evidence that lagged incivility influences affective polarization (residuals of ARFIMA (0, 0.45, 1) on partisan relative sentiment; residuals of ARFIMA (1, 0.38, 0) on relative feeling thermometer), as summarized in Table 5.2.2. However, there is evidence of contemporaneous effects of incivility within posts from Democratic pages on both the partisan sentiment (*B* = 32.92, *SE* = 17.99, *p* = .07) and feeling thermometer (*B* = 51.20, *SE* = 19.60, $p < .05$) measures of affective polarization, although the former is only marginally significant. An increase in uncivil posts from Democratic pages is associated with *increased* relative sentiment and feeling thermometer measures, meaning more favorable feelings toward the Republican Party and unfavorable feelings toward the Democratic Party. To be more specific, both graphs in Figure 5.2. show that civil content on Democratic pages is related to more favorable feelings toward the Democratic Party (i.e., relative score of affective polarization is less than 0) and uncivil content on Democratic pages is connected to more favorable feelings toward the Republican Party (i.e., relative score of affective polarization is greater than 0). Lagged incivility is not correlated with affective polarization. As the only results were from the contemporaneous effects, rather than the lagged effects proposed in the hypothesis, H9 is not supported.

Table 5.2.2: Effects of Lagged Incivility of Partisan Content on Aggregated Relative Value of Affective Polarization by Political Leanings.

Aggregated Incivility by Post Leanings	Partisan Relative Sentiment	Partisan Relative Feeling Thermometer
	B (SE)	B (SE)
Intercept	-133.99 (117.36)	-283.97^{+} (142.98)
Incivility of Democratic Posts	32.92^{+} (17.99)	51.20^{*} (19.60)
Incivility of Democratic Posts (lagged)	-12.10 (16.89)	-2.75 (18.52)
Incivility of Republican Posts	-32.31 (20.99)	-33.32 (22.43)
Incivility of Republican Posts (lagged)	8.08 (23.57)	-34.77 (25.54)
CCI	1.35 (1.17)	2.80^{+} (1.42)
R^2	0.14	0.34
Number of Obs	52	47

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Incivility is standardized and centered before aggregation.



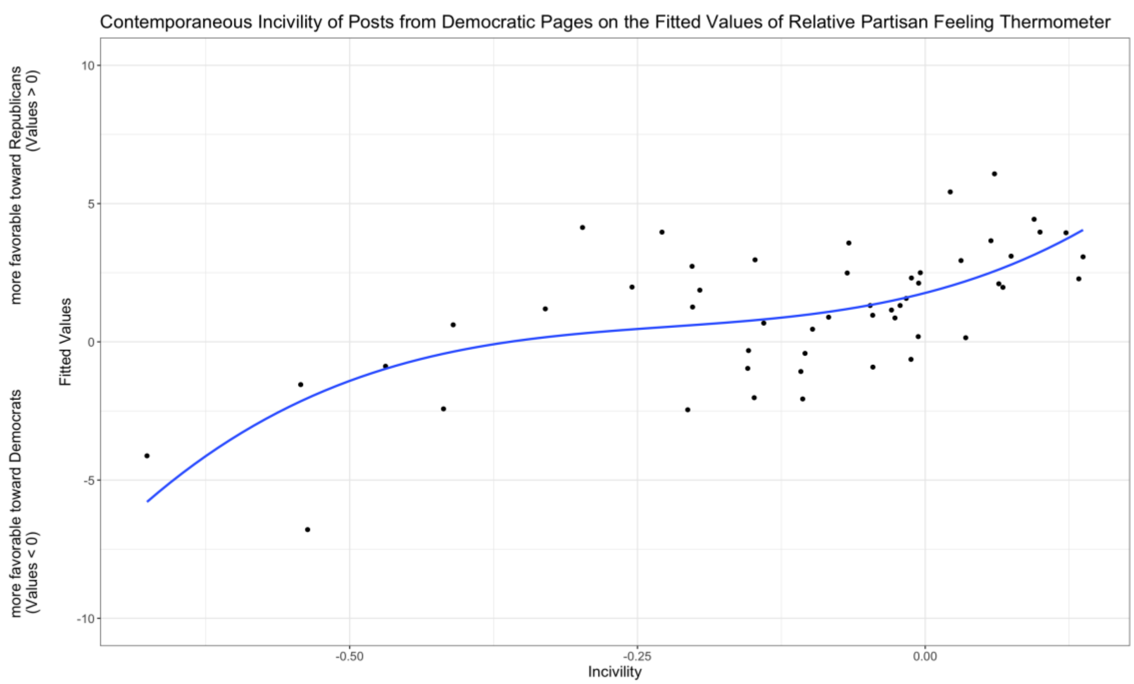


Figure 5.2: Contemporaneous Incivility of Posts from Democratic Pages on the Fitted Values of Affective Polarization: Partisan Relative Sentiment (top) and Feeling Thermometer (bottom).

NEGATIVITY VS. INCIVILITY ON AFFECTIVE POLARIZATION

H12 predicted a stronger relationship between the aggregated absolute value of affective polarization and incivility compared to the relationship between aggregated absolute value of the affective polarization and negativity. Residuals of ARFIMA models with external regressors of sentiment, incivility, and CCI were examined, and are shown in Table 5.3.1. There is marginally significant evidence that negativity and incivility influence the absolute feeling thermometer (negativity, B = 38.14, SE = 20.29, p = .07; incivility, B = 27.20, SE =16.09, p = .10) but no evidence on the absolute sentiment measure (negativity, B = 2.17, SE = 15.98, p = .89; incivility, B = -4.37, SE = 12.72, p = .73).[16] On the partisan feeling thermometer, aggregated negativity is associated with reduced affective polarization, and aggregated incivility is related to increased affective polarization.

Then I conducted a test for differences between the two coefficients of negativity and incivility for a direct comparison of the two estimates in a model where more than one estimate is not statistically significant (Gelman & Stern, 2006). Results show that there is no significant difference between negativity and incivility on either the partisan feeling thermometer measure (B = 10.95, 95% CI [-14.14, 36.03], for estimation; $F(1, 43)$ = 0.77, p = 0.38, for linear model comparisons) or the sentiment measure (B = 6.54, 95% CI [-12.01, 25.09], for estimation; $F(1, 48)$ = 0.77, p = 0.48, for linear model comparisons). There is no clear evidence that aggregated incivility has stronger effects on the absolute score of the affective polarization than aggregated negativity does.

Table 5.3.1: Comparisons of Negativity and Incivility on Aggregated Absolute Value of Affective Polarization.

[16] Models with lagged values produced non-significant results.

Aggregated Sentiment by All Posts	Partisan Sentiment	Partisan Feeling Thermometer
	B (SE)	B (SE)
Intercept	17.01 (88.87)	73.02 (118.22)
Sentiment	2.17 (15.98)	38.14^{+} (20.29)
Incivility	-4.37 (12.72)	27.20^{+} (16.09)
CCI	-0.19 (0.89)	-0.73 (1.18)
R^2	0.05	0.08
Number of Observations	52	47

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Sentiment and incivility are standardized and centered before aggregation. Lower sentiment indicates negativity while higher sentiment means positivity.

I continued to test if the relationship between negativity/incivility and affective polarization, using relative measures, differs depending on posts' political leanings, which is displayed in Table 5.3.2. Examining the effects of negativity and incivility in posts from Democratic pages only, incivility is stronger in magnitude (for partisan sentiment, B = 38.89, SE = 16.00, $p < .05$; for partisan feeling thermometer, B = 45.54, SE = 18.35, $p < .05$) than negativity (for partisan sentiment, B = 3.08, SE = 16.07, p = .85; for partisan feeling thermometer, B = -2.44, SE = 19.43, p = .90). Significance tests confirm these differences (partisan sentiment, B = 35.82, 95% CI [4.31, 67.33], for estimation; $F(1, 46) = 5.23$, $p < .05$, for linear model comparisons; partisan feeling thermometer, B = 47.98, 95% CI [9.63, 86.33], for estimation; $F(1, 41) = 6.38$, $p < .05$, for linear model comparisons). This pattern of polarization indicates increased favorability toward the Republican Party and/or increased unfavorability toward the Democratic Party. To be specific, both graphs in Figure 5.3.1 confirm that civil content from Democratic pages is related to more favorable feelings toward the Democratic Party and uncivil content from Democratic pages is connected to more favorable feelings toward the Republican Party, which shows a type of boomerang shape as a flat middle and curved edges.

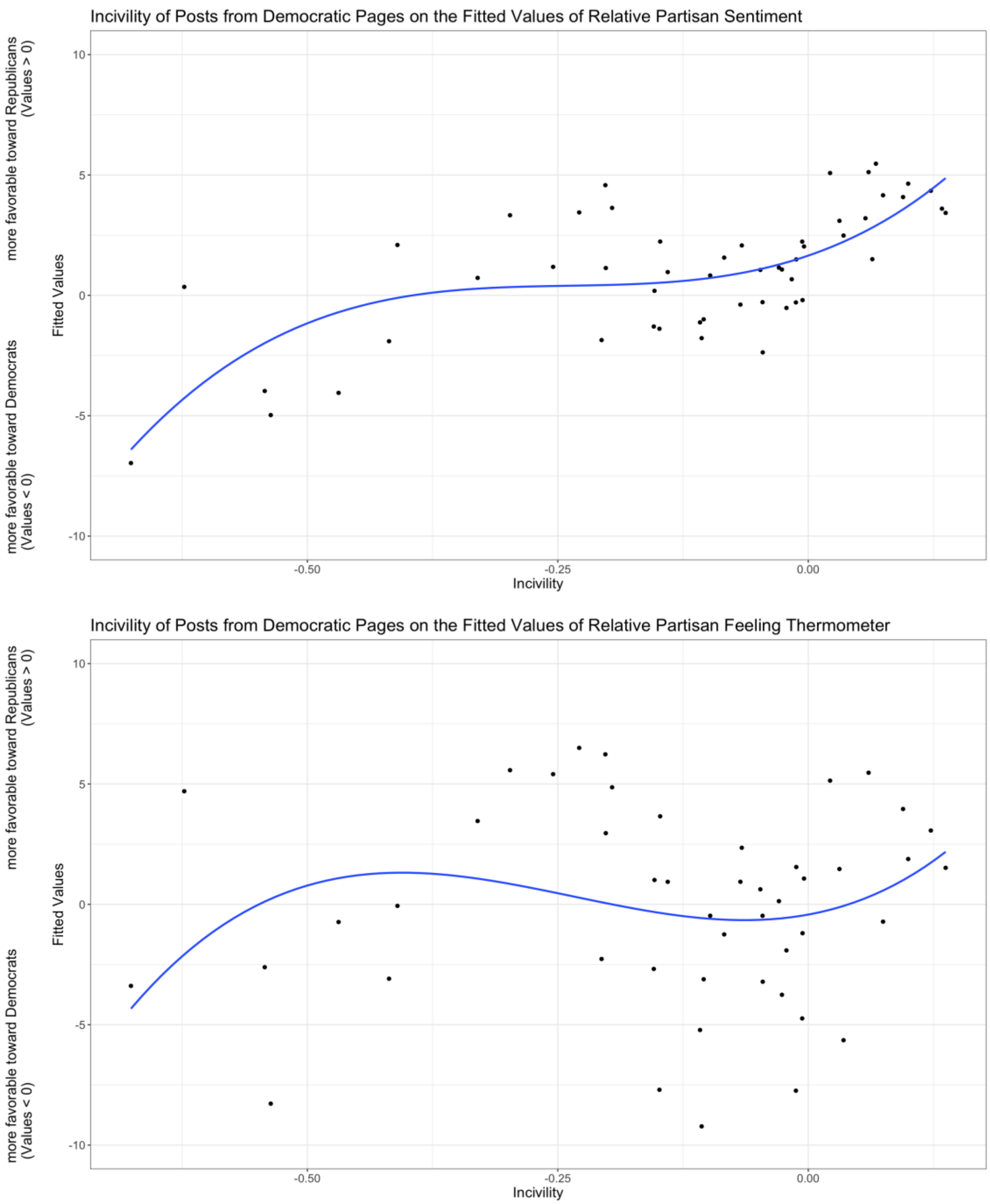


Figure 5.3.1: Incivility of Posts from Democratic Pages on the Fitted Values of Affective Polarization: Partisan Relative Sentiment (top) and Feeling Thermometer (bottom).

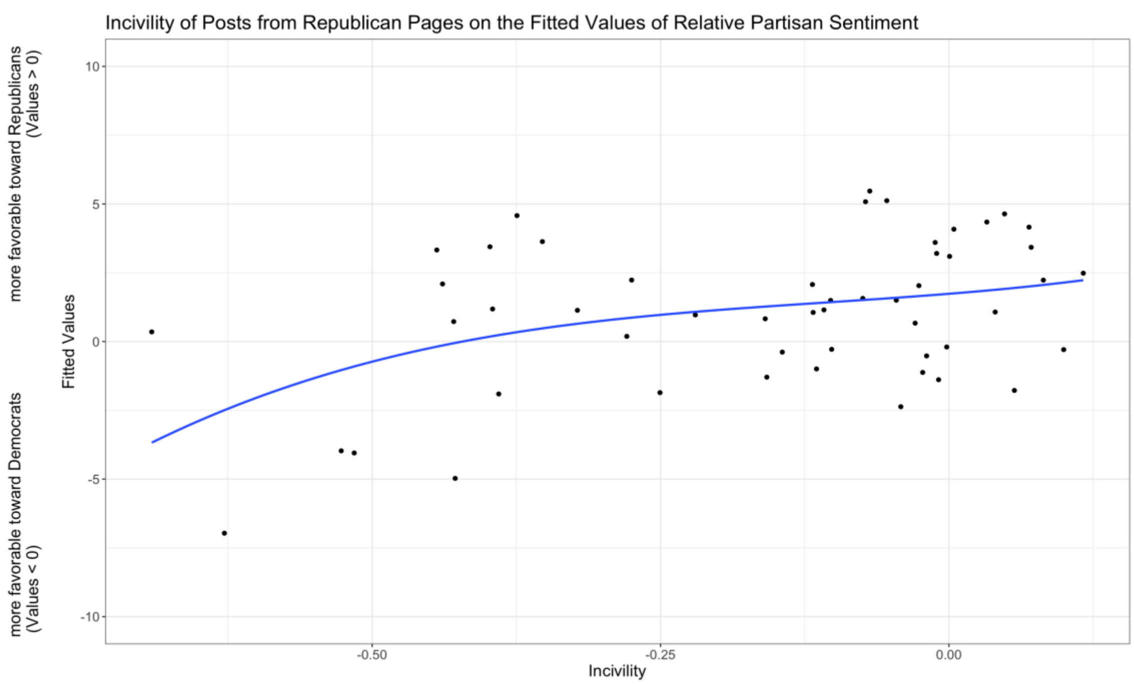
Table 5.3.2: Comparisons of Negativity and Incivility on Aggregated Relative Value of Affective Polarization by Political Leanings.

Aggregated Incivility by Post Leanings	Partisan Sentiment	Partisan Feeling Thermometer
	B (SE)	B (SE)
Intercept	-62.48 (100.54)	-143.71 (127.28)
Sentiment of Democratic Posts	3.08 (16.07)	-2.44 (19.43)
Incivility of Democratic Posts	38.89* (16.00)	45.54* (18.35)
Sentiment of Republican Posts	18.28 (12.57)	3.27 (13.91)
Incivility of Republican Posts	-21.92^{+} (12.91)	-54.47*** (13.98)
CCI	0.63 (1.00)	1.41 (1.27)
R^2	0.18	0.31
Number of Observations	52	47

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Sentiment and incivility are standardized and centered before aggregation. Lower sentiment indicates negativity while higher sentiment means positivity.

Next, looking at the effects of negativity and incivility in posts from Republican pages only, incivility has an impact on affective polarization (for partisan sentiment, B = -21.92, SE = 12.91, p = .096; for partisan feeling thermometer, B = -54.47, SE = 13.98, p < .001) but negativity does not (for partisan sentiment, B = 18.28, SE = 12.57, p = .19; for partisan feeling thermometer, B = 3.27, SE = 13.91, p = .72). Findings are summarized in Table 5.3.2. Testing for estimate comparisons demonstrates that there are significant differences between negativity and incivility in Republican posts on partisan sentiment (B = -40.2, 95% CI [-73.93, -6.47], for estimation; $F(1, 46) = 6.76$, $p < .05$, for linear model comparisons) and the partisan feeling thermometer (B = -57.74, 95% CI [-94.55, -20.92], for estimation; $F(1, 41) = 10.03$, $p < .01$, for linear model comparisons). Incivility in posts from Republican pages has a stronger influence on affective polarization than negativity. Increased incivility of posts in Republican pages is correlated with increased unfavorability toward the Republican Party and/or increased favorability toward the Democratic Party, which is also a type of boomerang shape as a flat middle and curved edges.

Figure 5.3.2 shows that uncivil content from Republican pages is related to more favorable feeling toward the Democratic Party on the relative feeling thermometer (bottom in Figure 5.3.2) while civil content from Republican pages is relatively constant. These patterns are not clearly present on the relative partisan sentiment (top in Figure 5.3.2). H12, proposing a stronger effect for incivility on affective polarization than negativity, is conditionally supported when aggregating negativity and incivility by posts' political leanings. These findings indicate that incivility in posts from Democratic pages relates with more favorable feelings toward the Republican Party and less favorable feelings toward the Democratic Party. The same thing happens for posts from Republican pages, where public opinion becomes more favorable toward the Democratic Party and less favorable toward the Republican Party.



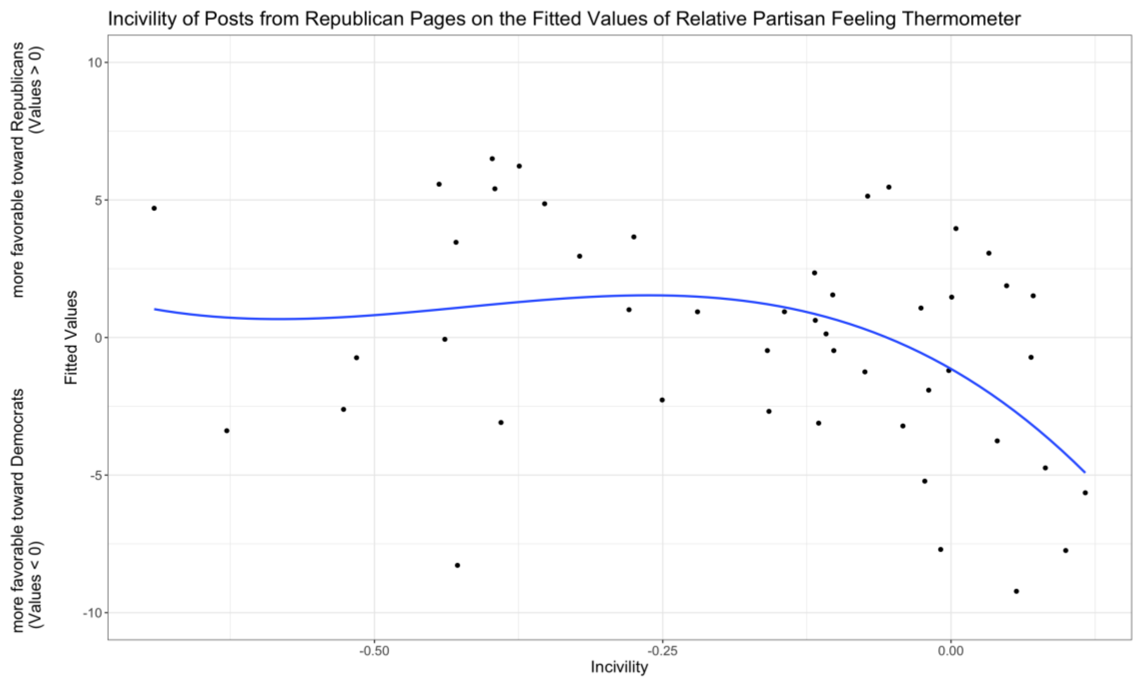


Figure 5.3.2: Incivility of Posts from Republican Pages on the Fitted Values of Affective Polarization: Partisan Relative Sentiment (top) and Feeling Thermometer (bottom).

EFFECTS OF NEGATIVITY AND INCIVILITY BY POLITICAL TARGETS ON POLARIZATION

I proposed that negativity and incivility, when accompanied by mentions of a political out-group only or when accompanied by mentions of both in- and out-groups (H5, H10) would be related to affective polarization. With independent variables of negativity by targets (out-group only, in-group only, without explicit entities, and both in-/out-groups) and CCI, regression models with residuals of ARFIMA predicting the absolute value of the affective polarization were constructed.

Negativity associated with out-group entities only does not significantly relate with affective polarization (for partisan sentiment measure, *B* = 18.15, *SE* = 16.23, *p* = .27; for partisan feeling thermometer measure, *B* = 13.43, *SE* = 19.99, *p* = .51). Findings are summarized in Table 5.4.1. H5a is not supported.

Next, negative content mentioning both political in-/out-groups has a marginally significant effect on partisan sentiment measure (*B* = -19.32, *SE* = 9.86, *p* = .06). The marginal effect on partisan sentiment means that negative posts (sentiment lower than 0) mentioning both in-/out-groups are correlated with increased affective polarization. There is no evidence of such a relationship using the partisan feeling thermometer measure (*B* = -15.41, *SE* = 11.96, *p* = .20), although the estimate is the same direction. In the partisan feeling thermometer model, there is a marginal effect of negativity associated with political in-group entities only (*B* = 38.40, *SE* =20.55, *p* = .07) meaning that negative political posts (sentiment lower than 0) mentioning in-group entities are associated with reduced affective polarization. H5b is not supported.

Table 5.4.1: Effects of Negativity with Partisan Targets on Aggregated Absolute Value of Affective Polarization.

Aggregated Sentiment by All Posts	Partisan Sentiment	Partisan Feeling Thermometer
	B (SE)	B (SE)
Intercept	38.34 (78.19)	-90.63 (109.03)
Sentiment: Out-group Entities Only	18.15 (16.23)	13.43 (19.99)
Sentiment: In-group Entities Only	20.12 (16.60)	38.40^{+} (20.55)
Sentiment: No Explicit Entities	-15.59 (16.77)	-14.00 (21.30)
Sentiment: Both In-/Out-group Entities	-19.32^{+} (9.86)	-15.41 (11.96)
CCI	-0.43 (0.79)	0.83 (1.10)
R^2	0.12	0.11
Number of Observations	52	47

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Sentiment is standardized and centered before aggregation. Lower sentiment indicates negativity while higher sentiment means positivity.

Turning to uncivil content, there is evidence that uncivil content targeted toward political out-groups relates to affective polarization on the partisan feeling thermometer measure (*B* = 38.06, *SE* =13.05, *p* < .01). This means that increased incivility toward political out-groups is associated with increased affective polarization using the partisan feeling thermometer measure. Uncivil content accompanied by mentions of out-groups does not affect partisan sentiment (*B* =

12.16, *SE* = 10.67, p = .26) although the direction of the estimate is consistent with the findings from the feeling thermometer. Findings are included in Table 5.4.2. Given the partial evidence from the feeling thermometer measure, H10a is partially confirmed.

Table 5.4.2: Effects of Incivility with Partisan Targets on Aggregated Absolute Value of Affective Polarization.

Aggregated Sentiment by Partisan Target	Partisan Sentiment	Partisan Feeling Thermometer
	B (SE)	B (SE)
Intercept	120.38 (105.63)	228.94 (152.56)
Incivility: Out-group Entities Only	12.16 (10.67)	38.06^{**} (13.05)
Incivility: In-group Entities Only	26.08 (17.31)	19.20 (21.98)
Incivility: No Explicit Entities	8.58 (10.21)	-8.50 (12.91)
Incivility: Both In-/Out-group Entities	-29.73^{*} (11.44)	-27.20^{+} (13.83)
CCI	-1.10 (1.05)	-2.29 (1.50)
R^2	0.17	0.18
Number of Observations	52	47

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Incivility is standardized and centered before aggregation.

There is consistent evidence that incivility toward both in-/out-groups relates to affective polarization, especially on the partisan sentiment measure (B = -29.73, *SE* =11.44, $p < .05$) and a marginally significant relationship for the partisan feeling thermometer (B = -27.20, *SE* =13.83, p = .06). These findings indicate that increased incivility toward both political in-/out-groups is correlated with reduced affective polarization. Findings are displayed in Table 5.4.2. H10b is confirmed.

PARTISAN DIFFERENCES OF POLITICAL TARGETS

RQ6 and RQ7 investigate partisan differences in the relationship between negative and uncivil posts by political targets and affective polarization, controlling for economic perceptions.

Negative content (sentiment lower than 0) mentioning both in-/out-groups in Republican posts is associated with increased affective polarization when measured using the partisan sentiment measure (B = -10.40, SE =5.13, p < .05). The direction is consistent for the partisan feeling thermometer, but it is not statistically significant (B = -9.16, SE = 6.66, p = .18). On the other hand, there is no evidence of political targets on negative content from Democratic pages, except for a marginal effect of negative posts with in-group entities on partisan sentiment (B = 28.33, SE = 15.45, p = .07), indicating that negative posts (sentiment lower than 0) about in-group targets is correlated with reduced affective polarization, as measured by partisan sentiment. RQ6 is answered and results are summarized in Table 5.5.1.

Table 5.5.1: Partisan Differences in Negativity with Partisan Targets on Aggregated Absolute Value of Affective Polarization by Political Leanings.

Aggregated Sentiment by Republican Posts Only	Partisan Sentiment	Partisan Feeling Thermometer
	B (SE)	B (SE)
Intercept	40.58 (62.47)	-61.07 (86.82)
Sentiment: Out-group Entities Only	17.70 (14.63)	7.02 (19.95)
Sentiment: In-group Entities Only	-6.79 (8.89)	0.66 (12.14)
Sentiment: No Explicit Entities	-4.06 (14.99)	7.32 (21.71)
Sentiment: Both In-/Out-group Entities	-10.40* (5.13)	-9.16 (6.66)
CCI	-0.40 (0.63)	0.60 (0.87)
R^2	0.19	0.06
Number of Observations	52	47
Aggregated Sentiment by Democratic Posts Only	**Partisan Sentiment**	**Partisan Feeling Thermometer**
	B (SE)	B (SE)
Intercept	138.23 (89.75)	-41.64 (126.28)
Sentiment: Out-group Entities Only	-0.63 (16.29)	5.06 (20.86)
Sentiment: In-group Entities Only	28.33+ (15.45)	20.22 (20.43)
Sentiment: No Explicit Entities	-20.06 (15.10)	-4.18 (19.74)
Sentiment: Both In-/Out-group Entities	-9.00 (9.87)	-3.31 (12.32)
CCI	-1.47 (0.92)	0.36 (1.29)
R^2	0.11	0.04

Number of Observations	52	47

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Sentiment is standardized and centered before aggregation. Lower sentiment indicates negativity while higher sentiment means positivity.

Uncivil content mentioning political out-groups is related to affective polarization for Republican posts only. Increased incivility toward political out-groups is related to increased affective polarization for both the absolute values of partisan sentiment (B = 14.29, SE =5.67, $p < .05$) and feeling thermometer (B = 25.33, SE =7.10, $p < .001$). Moreover, incivility accompanied by both in-/out-group entities is correlated with reduced affective polarization as measured by partisan sentiment (B = -16.03, SE = 6.57, $p < .05$) only. On the other hand, there is no significant evidence that uncivil content with political targets among Democratic posts is related to aggregated affective polarization. RQ7 is answered and findings are summarized in Table 5.5.2.

Table 5.5.2: Partisan Differences in Incivility with Partisan Targets on Aggregated Absolute Value of Affective Polarization by Political Leanings.

Aggregated Sentiment by Republican Posts Only	Partisan Sentiment	Partisan Feeling Thermometer
	B (SE)	B (SE)
Intercept	206.16 (104.35)	242.23 (149.08)
Incivility: Out-group Entities Only	14.29* (5.67)	25.33*** (7.10)
Incivility: In-group Entities Only	9.70 (9.69)	-2.47 (12.55)
Incivility: No Explicit Entities	12.42 (8.68)	8.85 (10.75)
Incivility: Both In-/Out-group Entities	-16.03* (6.57)	-3.66 (7.65)
CCI	-2.02^{+} (1.03)	-2.44 (1.47)
R^2	0.23	0.28
Number of Observations	52	47
Aggregated Sentiment by Democratic Posts Only	**Partisan Sentiment**	**Partisan Feeling Thermometer**
	B (SE)	B (SE)
Intercept	-44.78 (90.87)	-15.30 (127.22)
Incivility: Out-group Entities Only	-6.79 (11.71)	17.13 (14.62)
Incivility: In-group Entities Only	10.58 (21.05)	28.54 (26.11)
Incivility: No Explicit Entities	4.16 (11.05)	-22.55 (15.30)
Incivility: Both In-/Out-group Entities	-8.29 (14.78)	-20.02 (18.02)

CCI	0.51 (0.93)	0.17 (1.29)
R^2	0.10	0.09
Number of Observations	52	47

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Incivility is standardized and centered before aggregation.

DISCUSSION

In this chapter, I investigated an overarching question of whether language attributes – negativity, incivility, and political targets – on Facebook political posts are linked to affective polarization. To begin, I proposed that negative or uncivil language on political content may lead to affective polarization. The degree of negative or uncivil content from all partisan posts in the previous quarter does not directly lead to affective polarization. When both negative and uncivil content are included in the analysis, there are marginally significant relationships predicting affective polarization whereby negativity corresponds with lower polarization and incivility greater polarization, both when using the partisan feeling thermometer as the outcome variable. Yet these effects are not significantly different from each other. These findings provide at best modest evidence for the deleterious effects of uncivil content and no evidence of polarizing effects of negative content.

Measuring affective polarization as a relative measure with positive values indicating support for Republicans and negative values indicating support for Democrats yielded few significant relationships. Quarterly lagged negativity from posts on Democratic pages marginally leads to affective polarization and only contemporaneous effects of uncivil posts on Democratic pages are associated with affective polarization. The direction of the effect is that greater negativity and incivility on Democratic pages relates with higher favorability toward the Republican Party. Although the data at hand cannot definitively explain the result, several explanations are possible. Perhaps those running Democratic pages either respond to or anticipate more favorable opinion swings toward Republicans and publish more uncivil content. Perhaps uncivil content from Democratic pages circulates among Republicans and increases their in-party favorability. Regardless of the explanation, when taken as a whole, the evidence that negativity or incivility within posts leads to affective polarization is not clear-cut given the inconsistent patterns.

When negativity and incivility are aggregated by political leanings and tested in the same model, I found that (1) uncivil political content from Democratic and Republican pages is associated with affective polarization and (2) incivility is more effective than negativity in explaining affective polarization. The incivility estimates show that as incivility in posts on Democratic pages increases, public opinion is more favorable toward the Republican Party (or possibly less favorable toward the Democratic Party due to the relative nature of the affective polarization measures I applied). The incivility estimates for posts on Republican pages suggest that as the incivility of content from Republican pages increases, public opinion turns unfavorably toward the Republican Party (or possibly favorably toward the Democratic Party due to the relative dynamics from the measures in my dissertation). That is, increasing incivility from both partisan pages corresponds with public opinion moving away from the party using more incivility in their posts. As these are aggregate data, caution is warranted in advancing individual-level explanations. It could be that there is public backlash to incivility. Alternatively, it could be that partisans increasingly use incivility when they feel that public opinion turns against them, the reverse causal argument. The data as presented here cannot sort out among alternative explanations. What we can learn is that aggregate affective polarization is connected to incivility expressed on partisan pages online.

When the political targets (out-groups, in-groups, both in- and out-groups, and non-explicit or no targets) are considered, incivility in posts that also mention the out-group only or both in-and out-group targets have a much clearer relationship with affective polarization. As proposed in the hypotheses, incivility mentioning out-group targets or both in- and out-group targets is associated with polarization. Examples of uncivil content with presence of political targets are in Table 5.6. Uncivil content including out-group targets is associated with increased affective polarization and uncivil content including out-group targets along with in-groups is related to reduced affective polarization. These findings suggest that out-group hatred, in isolation, may be more problematic to affective polarization of the public opinion than in-group love is. Although it uses an individual level of analysis, previous research shows that partisans do not necessarily opt for harming the opposing party and rather opt in helping their own party (Lelks & Westwood, 2017) but under some conditions such as moral threats to their values or identities they are likely to choose the former (Amira et al., 2021). Perhaps outgroup hatred, by itself, leads people to think in starker terms. The prevalence of incivility toward out-groups may influence the broader public to greater threat compared to incivility with or without other political targets.

Table 5.6. Examples of Uncivil Posts by Presence of Political Targets.

Leanings	Targets	Posts (targets in bold)
Republican	Both in-/out groups	Another Hollywood idiot turned into a clown by exposing his ignorance. Thanks, **Ted Cruz**! **Ted Cruz** Wins Twitter With Comeback to Luke Skywalker's Butthurt Over Ending **Obama's** 'Net Neutrality' "Reject the dark side."
Democratic	Both in-/out groups	We Are Anti- **Donald J. Trump** Because We Are Not Fools Our Thanks To: Winning **Democrats** Timeline Photos YES, I'M **ANTI-TRUMP**. NO, I'M NOT MUSLIM, I'M NOT ILLEGAL, & I'M NOT LIVING OFF THE GOVERNMENT, WINNING **DEMOCRATS** I'M JUST NOT STUPID.
Republican	Out-groups only	DON'T BE LIKE THIS GUY... GREG MOVED OUT OF CALIFORNIA BECAUSE OF THE HIGH TAXES AND CRAZY **LEFTIST** POLICIES. GREG KEPT VOTING FOR **DEMOCRATS** & RUINED HIS NEW CITY WITH THE

		SAME CRAZY LEFTIST POLICIES. GREG IS AN IDIOT. DON'T BE LIKE GREG.
Democratic	Out-groups only	This guy is sub-human. Americans Against **Trump Donald Trump**, A Good President? He's Not Even A Good Person. He's so pathetically stupid. Ugh...

Finally, partisan differences in the relationships between negative or uncivil content and affective polarization were apparent. Uncivil posts from Republican pages targeting both in- and out groups correspond with lower partisan affective polarization while negative posts targeting both in- and out groups and uncivil posts targeting out-groups only from Republicans pages correspond with greater partisan affective polarization. In contrast, negativity associated with political in-groups in Democratic posts marginally relates with partisan affective polarization, while incivility does not. The relationship between incivility and out-group mentions in Republican posts and affective polarization could be because (1) the presence of out-groups from uncivil Republican posts is more effective in influencing what the public feels about the political parties, which may correspond to Republicans' straightforward and clear communication style compared to Democrats (Jarvis, 2004; Tetlock, 1983), and (2) Republicans as the main audience of Republican pages may be more susceptible to incivility targeting out-groups, which is consistent with research that conservatives are more likely to be influenced by negative information (Hibbing et al., 2014).

In sum, I found some evidence that language attributes are connected to aggregated public opinion about affective polarization. The clearest result, found for both measures of affective polarization, as well as for posts from Republican and Democratic pages, is that incivility in partisan posts leads people to polarize in the direction of the political opposition. It also is noteworthy that incivility paired with outgroup mentions resulted in more polarization whereas incivility paired with mentions of both in- and out-groups resulted in lower levels of polarization.

At the aggregate level, there is evidence of a connection between language use on partisan Facebook pages and polarization among the public. In the next chapter, I examine how people interact with partisan content by varying language attributes of social media posts, testing the direct causal relationship of language on affective polarization in an experimental setting and at the individual level.

Chapter 6: Do People Engage with Negative and Uncivil Political Content on Facebook and Polarize?

Here are some True or False questions about recent facts about politics and the media. Read each of the following statements and guess whether it is True or False.

(1) 70% of U.S. social media users never or rarely post or share content about political or social issues.

(2) Political extremists in the U.S. are the most active group of people in national politics, participating at higher rates in acts such as voting, attending a political rally and campaign event in person and virtually, and creating and engaging with political content on social media.

(3) 23% of social media users in the U.S. say social media led them to change their views on political issues.

The answers are True, True, and True according to Pew Research Center. For (1), only 9% of social media users actively participate in sharing content about political or social issues (Mcclain, 2021). The main reasons why people do not share political or social posts are that they don't want the things that they posted used against them, they don't want to be attacked for their political views or to offend others, and they don't have anything to add or not much interest (Mcclain, 2021). For (2), political discourse on Twitter is driven by a small number of those who have extreme political ideologies (Blazina, 2022). And (3) is True. One-quarter of adult social media users who report having changed their views about politics because of something they saw on social media (Perrin, 2020).

Given that only a small number of people actively engage with political content on social media and most people do not want to engage with political content, how do a substantial number of people on social media change their views in politics? Do they become more polarized? What

I examine in this chapter fills gaps in how engagement with political posts using negative, uncivil language targeted at political groups leads to affective polarization.

CHAPTER OVERVIEW

In this chapter, I examined if, at the individual level, negative and uncivil content garners more engagement and whether language attributes influence affective polarization. Conducting an experiment enables me to identify a causal relationship between language and affective polarization. In the next two sections, I report on whether those who viewed negative or uncivil language in political content are more likely to engage (H13) and affectively polarize (H14), compared to those who saw neutral language in political content. Then, I look at whether the political targets of the messages influence social engagement (H15) and affective polarization (H16) in the following two sections. Later, H17 and H18, which propose stronger effects of incivility over negativity in social engagement and polarization, are tested. Research questions of whether there are differences in social engagement (RQ8) and affective polarization (RQ9) based on mentions of in-groups and out-groups are explored. Finally, I analyze an open-ended question about how respondents in the experiment felt and thought about the political posts that they saw.

RESULTS OF WHETHER NEGATIVE AND UNCIVIL LANGUAGE CORRESPOND WITH SOCIAL ENGAGEMENT

Compared to neutral language, I proposed that negative or uncivil language will receive more Comments (H13a), Shares (H13b), and Angry reactions (H13c) after controlling for the presence of political targets. Social engagement metrics are analyzed in Table 6.1. Compared to neutral posts, uncivil posts garner significantly more Comments (B = 1.40, SE = 0.60, $p < .05$). For Shares, there is no evidence that either negative (B = 0.54, SE = 0.76, p = .48) or uncivil (B = -0.31, SE = 0.85, p = .72) content receives more engagement than neutral content does. Finally, only uncivil content gains a greater number of Angry reactions compared to neutral content (B = 1.13, SE = 0.38, $p < .01$). No evidence is found that negative content receives more Angry reactions relative to neutral content (B = 0.58, SE = 0.43, p = .18). H13a and H13c are supported while H13b is not. Other social engagement metrics do not show statistically significant differences by language attributes except for a marginal effect on Sad reactions, where there is less engagement with negative content compared to neutral content (B = -0.77, SE = 0.41, p = .06).

Table 6.1: Effects of Negativity, Incivility, and Political Targets on Social Engagement.

	Comments	Shares	Angry	Love
	B (SE)	B (SE)	B (SE)	B (SE)
Fixed effects:				
Intercept	-0.27 (0.61)	-1.01 (1.01)	-1.47^{**} (0.49)	-0.24 (0.50)
(ref. Neutral terms)				
Negative terms	0.97 (0.62)	0.54 (0.76)	0.58 (0.43)	0.22 (0.28)
Uncivil terms	1.40^{*} (0.60)	-0.31 (0.85)	1.13^{**} (0.38)	0.35 (0.37)
(ref. No explicit target)				
Out-group target	-0.03 (0.48)	-0.01 (0.76)	0.00 (0.52)	0.10 (0.35)
In-group target	-0.78 (0.49)	-1.09 (0.94)	1.27^{**} (0.40)	0.12 (0.33)
Zero-inflation model:				
Intercept	3.44^{***} (0.27)	3.06^{***} (0.68)	2.38^{***} (0.22)	1.77^{***} (0.42)
Number of Obs	920	920	920	920
Log-Likelihood	-123	-64	-192.9	-360
AIC	260.1	140	397.9	734
BIC	293.8	168.9	426.8	767.8
	Like	Sad	Haha	Wow
	B (SE)	B (SE)	B (SE)	B (SE)
Fixed effects:				
Intercept	0.87^{***} (0.11)	0.11 (0.35)	-1.07^{+} (0.55)	-0.57 (0.38)
(ref. Neutral terms)				
Negative terms	0.12 (0.12)	-0.77^{+} (0.41)	0.18 (0.42)	0.16 (0.36)
Uncivil terms	0.12 (0.13)	-0.49 (0.38)	-0.17 (0.46)	-0.12 (0.39)
(ref. No explicit target)				
Out-group target	0.01 (0.13)	-0.45 (0.39)	0.03 (0.41)	0.02 (0.40)
In-group target	0.18 (0.12)	-0.52 (0.41)	-0.52 (0.47)	0.56 (0.35)
Zero-inflation model:				

Intercept	0.98^{***} (0.09)	2.13^{***} (0.27)	1.79^{***} (0.44)	2.19^{***} (0.23)
Number of Obs	920	920	920	920
Log-Likelihood	-937.5	-197.1	-161.3	-225.9
AIC	1889.1	408.2	334.7	463.8
BIC	1922.8	441.9	363.6	492.8

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Model fit by a generalized linear mixed model (GLMM) using glmmTMB in R. Family parameter (probability distribution) is determined based on likelihood ratio tests (either Poisson or negative binomial).

RESULTS OF WHETHER POLITICAL TARGETS CORRESPOND WITH SOCIAL ENGAGEMENT

H15 proposed that the presence of political targets would influence social engagement, holding the language attributes constant. Note that in-group target posts are counter-attitudinal while out-group target posts are pro-attitudinal. Table 6.1 shows that compared to posts with non-explicit targets, posts with in-group (B = -0.78, SE = 0.49, p = .11) or out-group (B = -0.03, SE = 0.48, p = .95) targets do not receive more Comments. H15a is not confirmed. Also, there is no evidence of an effect of out-group targets on greater Shares compared to posts with non-explicit targets (B = -0.01, SE = 0.76, p = .99). H15b is not supported. On the other hand, political posts with in-group targets (counter-attitudinal posts) garner more Angry reactions compared to posts with non-explicit targets (B = 1.27, SE = 0.40, p < .01). H15c is confirmed. Other forms of social engagement do not show statistical differences based on the presence of political targets.

RESULTS OF WHETHER NEGATIVE AND UNCIVIL LANGUAGE CORRESPOND WITH POLARIZATION

H14 proposed that partisan content with negative or uncivil terms will yield more affective polarization. Holding the presence of political targets constant, respondents who were exposed to negative political posts polarized compared to those who were exposed to neutral political posts in terms of partisan absolute emotions (B = 0.28, SE = 0.11, p < .05), absolute feeling thermometer (B = 4.56, SE = 2.29, p < .05), and absolute trait evaluation (B = 0.27, SE = 0.15, p = .07). Although the effects on the partisan trait evaluation are marginally significant, the other two measures are statistically significant. H14 is supported that negative language results in greater affective polarization, as the findings in Table 6.2.1 show.

Table 6.2.1: Effects of Negativity, Incivility, and Political Targets on Absolute Value of Affective Polarization.

Absolute Measures of Affective Polarization	Partisan emotions	Partisan feeling thermometer	Partisan trait evaluation
	B (SE)	B (SE)	B (SE)
Intercept	1.00^{***} (0.10)	29.86^{***} (2.09)	1.29^{***} (0.13)
(ref. Neutral terms)			
Negative terms	0.28^{*} (0.11)	4.56^{*} (2.29)	0.27^{+} (0.15)
Uncivil terms	0.16 (0.11)	1.66 (2.30)	0.19 (0.15)
(ref. No explicit target)			
Out-group target	0.07 (0.11)	1.85 (2.30)	0.14 (0.15)
In-group target	0.00 (0.11)	0.71 (2.28)	0.07 (0.15)
Number of Obs	915	915	914
R^2	0.007	0.005	0.005

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

I further examined the effects of negative or uncivil language on political content by looking at interaction effects with partisanship. I summarized the interaction effect between the experiment conditions and party identification after controlling for the presence of political targets in Table 6.2.2. There are consistent patterns for how language attributes affect affective polarization by the respondent's party identification. Differences exited for Republicans compared to Democrats. Republicans who were exposed to negative political content polarized in terms of partisan absolute emotions compared to those exposed to neutral content (B = 0.51, SE = 0.24, p < .05). Likewise, only for Republicans, compared to those who viewed neutral content, those who viewed negative content polarized (B = 0.74, SE = 0.31, p < .05) and those who saw uncivil content marginally polarized (B = 0.58, SE = 0.31, p = .06) based on the partisan absolute trait evaluation measure. For the partisan absolute feeling thermometer, there is no significant evidence between language attributes and party identification although the effects for Republicans of negative content showed the same direction with other two polarization measures (*B* = 6.95, *SE* = 4.85, *p* = .15). These patterns are visualized in Figure 6.1. Compared to neutral content, negative content polarizes Republicans while neutral or negative language does not affect Democrats. H14 is further confirmed for Republicans only.

Table 6.2.2: Effects of Negative and Uncivil Language by Partisanship on Absolute Value of Affective Polarization.

Absolute Measures of Affective Polarization	Partisan emotions	Partisan feeling thermometer	Partisan trait evaluation
	B (SE)	B (SE)	B (SE)
Intercept	1.08^{***} (0.12)	32.16^{***} (2.38)	1.48^{***} (0.15)
(ref. Neutral terms)			
Negative terms	0.11 (0.14)	2.23 (2.80)	0.02 (0.18)
Uncivil terms	0.09 (0.14)	0.73 (2.82)	-0.01 (0.18)
(ref. No explicit target)			
Out-group target	0.07 (0.11)	1.81 (2.30)	0.15 (0.15)
In-group target	0.01 (0.11)	0.72 (2.28)	0.08 (0.14)
Republican	-0.26 (0.17)	-6.83^{*} (3.43)	-0.60^{**} (0.22)
Negative terms x Republican	0.51^{*} (0.24)	6.95 (4.85)	0.74^{*} (0.31)
Uncivil terms x Republican	0.22 (0.24)	2.85 (4.85)	0.58^{+} (0.31)
Number of Obs	912	912	911
R^2	0.012	0.011	0.014

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

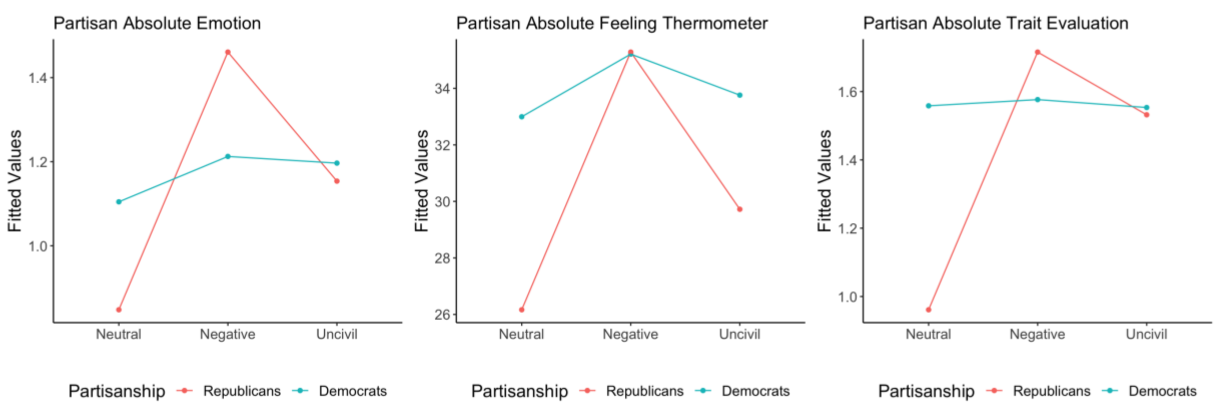


Figure 6.1: Interaction Effects between Language Attributes and Party Identification on the Fitted Values of Absolute Affective Polarization (Partisan Emotions, left; Partisan Feeling Thermometer, middle; Partisan Trait Evaluation, right).

RESULTS OF WHETHER POLITICAL TARGETS CORRESPOND WITH POLARIZATION

I next test whether that presence of explicit political targets (in-group or out-group) influences affective polarization after controlling for the language attributes (H16). There is no evidence that affective polarization varies by political target after controlling for the language attributes of negativity or incivility as shown in Table 6.2.1.

I continued to examine the effects of political targets by partisanship on affective polarization as shown in Table 6.3. There are consistent but marginal patterns over how presence of political targets influences affective polarization by the respondent's party identification. There are not consistent patterns for Democrats by presence of political targets. On the other hand, Republicans who were exposed to posts with in-group targets depolarized compared to those who were exposed to posts with non-explicit targets on the partisan absolute emotions (B = -0.51, SE = 0.23, $p < .05$), absolute feeling thermometer (B = -8.43, SE = 4.81, p = .08), and absolute trait evaluation (B = -0.51, SE = 0.31, p = .10), although the effects were marginal on the partisan feeling thermometer and trait evaluation measures. Those partisan differences are displayed in Figure 6.2. H16 is not supported.

Table 6.3: Effects of Political Targets by Partisanship on Absolute Value of Affective Polarization.

Absolute Measures of Affective Polarization	Partisan emotions	Partisan feeling thermometer	Partisan trait evaluation
	B (SE)	B (SE)	B (SE)
Intercept	0.91^{***} (0.12)	29.91^{***} (2.39)	1.22^{***} (0.15)
(ref. Neutral terms)			
Negative terms	0.27^{*} (0.11)	4.41^{+} (2.29)	0.26^{+} (0.15)
Uncivil terms	0.16 (0.11)	1.59 (2.29)	0.18 (0.15)
(ref. No explicit target)			
Out-group target	0.18 (0.14)	2.69 (2.83)	0.34^{+} (0.18)
In-group target	0.17 (0.14)	3.47 (2.80)	0.24 (0.18)
Republican	0.25 (0.16)	0.07 (3.37)	0.20 (0.21)
Out-group target x Republican	-0.31 (0.24)	-2.58 (4.85)	-0.58^{+} (0.31)

In-group target x Republican	-0.51* (0.23)	-8.43+ (4.81)	-0.51+ (0.31)
Number of Obs	912	912	911
R^2	0.013	0.012	0.011

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

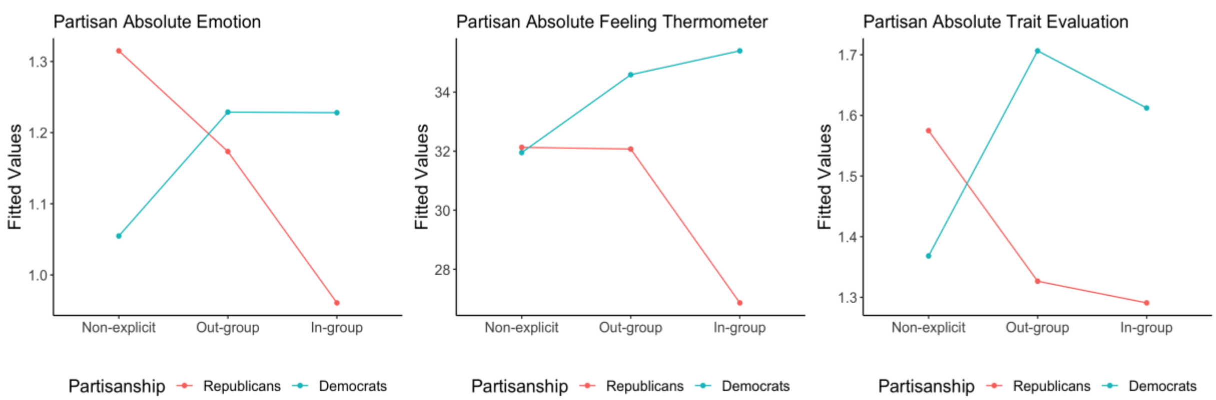


Figure 6.2: Interaction Effects between Presence of Political Targets and Party Identification on the Fitted Values of Absolute Affective Polarization (Partisan Emotions, left; Partisan Feeling Thermometer, middle; Partisan Trait Evaluation, right).

DIFFERENCES OF POLITICAL LANGUAGE: INCIVILITY VS. NEGATIVITY

I proposed that Comments (H17a), Shares (H17b), and Angry reactions (H17c) would be more strongly predicted by uncivil posts than negative posts. Holding the presence of political targets constant, each metric is tested in turn as shown in Table 6.4. Compared to negative content, uncivil content receives marginally greater number of Angry reactions (B = 0.77, SE = 0.42, p = .07) while there is no evidence of stronger effects of incivility over negativity for Comments (B = 0.77, SE = 0.59, p = .19) or Shares (B = -0.44, SE = 0.91, p = .63). H17c is marginally confirmed but H17a and H17b are not supported. Except for the finding that uncivil content receives marginally fewer Haha reactions than negative content (B = -1.08, SE = 0.60, p = .07), other reactions do not differ between uncivil and negative content.

Table 6.4: Negativity vs. Incivility and In-group vs. Out-group on Social Engagement.

	Comments	Shares	Angry	Love

	B (SE)	B (SE)	B (SE)	B (SE)
Fixed effects:				
Intercept	0.67 (0.42)	-3.71^{***} (0.62)	-1.07^{+} (0.60)	-0.04 (0.64)
(ref. Negative terms)				
Uncivil terms	0.77 (0.59)	-0.44 (0.91)	0.77^{+} (0.42)	-0.10 (0.44)
(ref. Out-group target)				
In-group target	-1.26^{*} (0.64)	-1.42 (1.12)	1.34^{*} (0.56)	-0.06 (0.40)
Zero-inflation model:				
Intercept	3.60^{***} (0.39)	-16.5 (18212)	2.51^{***} (0.20)	1.57^{+} (0.84)
Number of Obs	405	405	405	405
Log-Likelihood	-48.5	-25.8	-100.6	-159.4
AIC	105.1	59.7	209.3	328.8
BIC	121.1	75.7	225.3	348.8

	Like	Sad	Haha	Wow
	B (SE)	B (SE)	B (SE)	B (SE)
Fixed effects:				
Intercept	1.05^{***} (0.13)	-0.72 (0.72)	-1.43 (1.02)	-0.15 (0.43)
(ref. Negative terms)				
Uncivil terms	0.00 (0.15)	-0.30 (0.67)	-1.08^{+} (0.60)	-0.36 (0.45)
(ref. Out-group target)				
In-group target	0.18 (0.15)	-1.64^{*} (0.82)	-1.07^{+} (0.60)	0.30 (0.45)
Zero-inflation model:				
Intercept	1.23^{***} (0.13)	2.01^{**} (0.74)	0.61 (1.44)	2.30^{***} (0.31)
Number of Obs	405	405	405	405
Log-Likelihood	-376.9	-49.6	-61	-104.5
AIC	763.9	107.3	130.1	217
BIC	783.9	123.3	146.1	233

Note: $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$. Model fit by a generalized linear mixed model (GLMM) using glmmTMB in R. Family parameter (probability distribution) is determined based on likelihood ratio tests (either Poisson or negative binomial).

Next, I proposed stronger effects of uncivil content over negative content on affective polarization (H18). Those who are exposed to uncivil posts do not polarize more than those who are exposed to negative posts, leading me to reject H18. Findings are summarized in Table 6.5.1.

Table 6.5.1: Negativity vs. Incivility and In-group vs. Out-group on Absolute Value of Affective Polarization.

Absolute Measures of Affective Polarization	Partisan emotions	Partisan feeling thermometer	Partisan trait evaluation
	B (SE)	B (SE)	B (SE)
Intercept	1.38*** (0.13)	35.58*** (2.56)	1.74*** (0.16)
(ref. Negative terms)			
Uncivil terms	-0.14 (0.15)	-3.32 (2.93)	-0.06 (0.19)
(ref. Out-group target)			
In-group target	-0.06 (0.15)	-0.14 (2.93)	-0.12 (0.19)
Number of Obs	402	402	402
R^2	0.003	0.003	0.001

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Interactions effects of negativity and incivility by partisanship also do not show differences in the effects of negativity and incivility on affective polarization, as Table 6.5.2 shows. H18 is not supported.

Table 6.5.2: Negativity vs. Incivility by Partisanship on Absolute Value of Affective Polarization.

Absolute Measures of Affective Polarization	Partisan emotions	Partisan feeling thermometer	Partisan trait evaluation
	B (SE)	B (SE)	B (SE)
Intercept	1.35*** (0.15)	35.57*** (2.93)	1.67*** (0.19)
(ref. Negative terms)			
Uncivil terms	-0.02 (0.18)	-0.54 (3.54)	0.12 (0.23)
(ref. Out-group target)			
In-group target	0.09 (0.23)	-0.01 (4.50)	0.21 (0.29)
Republican	-0.06 (0.15)	-0.13 (2.92)	-0.12 (0.19)
Uncivil terms x Republican	-0.38 (0.32)	-8.46 (6.26)	-0.55 (0.40)
Number of Obs	400	400	400
R^2	0.007	0.013	0.006

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

DIFFERENCES OF POLITICAL TARGETS: IN-GROUP VS. OUT-GROUP

Differences between in-group and out-group targets are explored in relation to social engagement (RQ8) and affective polarization (RQ9). I examined differences between in-group and out-group targets by social engagement, as posed in RQ8. After controlling for the language attributes, posts with in-group target receive significantly fewer comments than do posts with out-group targets (*B* = -1.26, *SE* = 0.64, $p < .05$). Angry reactions also show significant in-group and out-group differences (*B* = 1.34, *SE* = 0.56, $p < .05$), indicating stronger effects of political posts for in-group targets than out-group targets. Sad reactions demonstrate significantly less engagement with in-group targeting posts than out-group targeting posts (*B* = -1.64, *SE* = 0.82, $p < .05$) and Haha reactions display marginally less engagement with posts with in-group targets than those with out-group targets (*B* = -1.07, *SE* = 0.60, $p = .08$). Other metrics does not show much difference. Findings are presented in Table 6.4.

Whether political posts with out-group or in-group targets influence affective polarization is investigated in Table 6.5.1 (RQ9). There is no difference between in-group or out-group targets of political posts on affective polarization.

Interaction effects of in- and out-group targets by partisanship on affective polarization do not show partisan differences (RQ9). There is no evidence of partisan differences by political targets, which is displayed in Table 6.6.

Table 6.6: In-group vs. Out-group Targets by Partisanship on Absolute Value of Affective Polarization.

Absolute Measures of Affective Polarization	Partisan emotions	Partisan feeling thermometer	Partisan trait evaluation
	B (SE)	B (SE)	B (SE)
Intercept	1.34^{***} (0.15)	35.88^{***} (2.94)	1.75^{***} (0.19)
(ref. Negative terms)			
Uncivil terms	-0.13 (0.15)	-3.20 (2.92)	-0.06 (0.19)
(ref. Out-group terms)			
In-group target	0.08 (0.18)	1.89 (3.55)	-0.10 (0.23)
Republican	0.11 (0.22)	-1.12 (4.44)	-0.04 (0.28)
In-group x Republican	-0.43 (0.32)	-6.48 (6.26)	-0.07 (0.40)
Number of Obs	400	400	400
R^2	0.008	0.011	0.002

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

ANALYSIS OF OPEN-ENDED QUESTIONS: HOW PARTICIPANTS FEEL AND THINK ABOUT THE POLITICAL POSTS THEY SAW

Participants were asked to write a short paragraph about how they thought or felt about the posts they saw. I applied the sentiment and incivility models used for the aggregate analysis to these responses. This analysis can provide demonstrate whether the experimental stimuli affected people's responses to the political content that they encountered as part of the experiment.

Table 6.7 summarizes participants' comments by experimental condition. Overall, language attributes made people feel and think differently in response to the posts they encountered. People wrote different responses depending on what they saw in terms of how negative their comments were and how uncivil their comments were. Those viewing negative or uncivil posts wrote with more negativity (e.g., "This person seems to hate liberals. They seem to be a conservative. They just moan and complain about liberal policies. It's annoying.") than those viewing neutral posts (e.g., "The results of weapon viciousness are more unavoidable and influence whole networks, families and youngsters.") (for negative posts, B = -0.18, SE = 0.09, $p < .05$; for uncivil posts, B = -0.23, SE = 0.09, $p < .05$). Those viewing uncivil posts also wrote more uncivil open-ended responses (e.g., "GD idiotic left leaning liberal talking about complete nonsense in EVERY POST! DUMBER THAN A BOX OF ROCKS.") than those viewing neutral (e.g., "they seemed pretty typical of facebook, one of the reasons I avoid it except for specialized groups. Seeing the dumb things some people believe is quite disappointing and in some cases, many cases, harmful.") (B = 0.04, SE = 0.01, $p < .01$). Examples above are some of the most negative or uncivil responses by experimental conditions for comparisons. In addition, which targets were present affected the amount of incivility included in their open-ended responses (B = 0.03, SE = 0.01, $p < .05$). Those viewing in-group targets used more incivility than those viewing non-explicit targets.

Table 6.7: Negativity and Incivility of Responses by Negativity, Incivility and Political Targets of Political Posts.

	Response Sentiment	Response Incivility
	B (SE)	B (SE)
Intercept	0.57^{***} (0.08)	0.07^{***} (0.01)
(ref. Neutral terms)		
Negative terms	-0.18^{*} (0.09)	0.01 (0.01)
Uncivil terms	-0.23^{*} (0.09)	0.04^{**} (0.01)
(ref. No explicit target)		
Out-group target	0.07 (0.09)	0.00 (0.01)
In-group target	0.00 (0.09)	0.03^{*} (0.01)
Number of Obs	915	907
R^2	0.008	0.019

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Discussion

This chapter scrutinized whether people engage more with posts containing negative or uncivil terms compared to posts with impartial terms and whether exposure to those posts leads to affective political divisions. Using an experimental setting that resembles a Facebook News Feed, people's interactions with political content was unobtrusively tracked and how those who saw different political posts with varying language attributes polarized was analyzed. When different forms of social engagement are tested separately holding the presence of political targets constant, participants who saw uncivil posts left more Comments and more Angry reactions compared to those who viewed neutral posts. Those who were exposed to negative content left fewer Sad reactions than those who were exposed to neutral content (although it's marginally significant). There are no differences among other reactions. Negative posts do not receive more engagement than neutral posts. I did not find much evidence of stronger effects for incivility over negativity on social engagement in the individual level, except that uncivil posts receive marginally more Angry reactions than negative posts after controlling for the presence of political targets.

The next question is whether exposure to negative or uncivil language results in affective polarization. Overall, exposure to negative terms has a polarizing effect compared to exposure to neutral terms after holding presence of political targets constant. There is no evidence for uncivil terms. Partisan differences explain these dynamics more clearly, as might be anticipated based on how the polarization measures were constructed where polarization exists with higher values indicating greater polarization. Among those who identified as Republican, participants who encountered negative posts polarized more than those who received neutral posts. These patterns are found on emotions and trait evaluation toward the two partisans. Effects of incivility are not found for Republicans except for a marginal effect on partisan trait evaluation. Participants who identified as Democrats do not show any differences by the posts' negativity or incivility. Negativity is more effective for Republicans than Democrats. This may be because conservatives are more attentive to negative information than liberals are (Carraro et al., 2011; Dodd et al., 2012; Hibbing et al., 2014), which may have resulted in asymmetrical polarization of emotional evaluations toward the partisans.

For the effects as political targets, participants do not show more social engagement with political posts depending on the political target (in-groups or out-groups) after controlling for the language attributes (i.e., negative, uncivil, and neutral content). Only Angry reactions received more engagement from those who were exposed to posts with in-group targets compared to posts with non-explicit targets. Because posts mentioning in-groups were counter-attitudinal, those who were exposed to those posts may cause angry emotions, which leads to react to more Angry reactions.

Subsequently, I examined whether the presence of political targets would have polarizing effects. There is no evidence of polarization resulting from posts with political in-group or out-group targets in general. However, among Republicans, those who were exposed to political content targeting in-groups depolarized more than those who viewed political content without explicit targets did. These patterns are found on the partisan emotion, feeling thermometer, and trait evaluation, but the two latter measures are marginal. Exposure to targeting political in-groups depolarizes as counter-attitudinal information does. No difference is found between effects of in-group and out-group targets. Partisan differences of political target presence of in-groups and out-groups on affective polarization are not found either.

From the comparison of in-group vs. out-group targets in political content, posts with in-group targets receive more engagement as well as Angry reactions compared to those with out-group targets. As related research suggests that partisan extremists are likely to suppress antagonistic rhetoric directed toward their own party than to promote antagonistic rhetoric directed toward their opposing party (Lelkes & Westwood, 2017), it is possible that political content targeting in-groups received more Angry reactions as a token of disapproval. On the other hand, in-group target posts receive significantly fewer Comments and Sad reactions, and generate marginally less Haha reactions compared to out-group target posts. I also find that negative or uncivil posts with out-group targets (pro-attitudinal content) are more engaging than those with in-group targets (counter-attitudinal content). These findings support that pro-attitudinal information attracts political engagement and participation more than counter-attitudinal information (e.g., Matthes et al., 2019; Pedersen et al., 2021).

Finally, I analyzed open-ended responses that participants authored about how they think or feel about the posts they saw. The degree of negativity or incivility in their responses corresponds with whether participants saw negative or uncivil posts. Those who saw posts targeting their in-groups (counter-attitudinal messages) expressed strong aversion or disregard. On the other hand, those who saw posts targeting their out-groups (pro-attitudinal messages) responded by agreeing with the posts or displaying less intense reactions (e.g., “the posts are political in nature”). Some said that the Feed had too many political posts unlike their News Feed while others said the News Feed was similar to their own.

A majority of participants did not interact with political posts and preferred to engage with non-political posts. A subset of respondents mentioned that they would block the political posts from their News Feed or did ignore the posts in the study, as I counted the responses with explicit opinions for opting out political content (88 out of 920) although it is not a content analysis. However, I found evidence of greater interaction with posts that are highly negative, uncivil, or attacking political targets in this chapter that adds to the evidence from earlier chapters. In the next chapter, I will reconcile findings across the empirical chapters and discuss the broader implications of this work.

Chapter 7: How Language Responds to Polarization

Social media platforms such as Facebook and Twitter have become important political actors. Facebook has removed a substantial number of political pages (Tynan, 2018) and officially announced that they will permanently reduce political content appearing in people's timelines, mentioning that users don't want political content to take over their News Feed (Facebook, 2021). Twitter also removed hundreds of political accounts between July and December 2020 for violating its policy against hateful conduct (Wagner, 2021). Twitter regulates political ads and prohibits the promotion of political content. These efforts could block the negativity and incivility that give rise to polarization. Given that platforms are taking actions to reduce politically harmful content, is everything actually okay?

In this dissertation, I propose that language usage on social media drives affective polarization, which is influenced by two processes: (1) negativity and incivility in political content attract more engagement, which means that this content is then algorithmically prioritized and more visible to other users, and (2) exposure to negative and uncivil content leads to affective polarization. I do not argue that deleting or blocking polarizing content and accounts would be of no use; rather, it might be helpful in addressing the malignant effects of polarization. Polarization is bad because it undermines essentials of democracy such as tolerance for opposing perspectives, empathy, and creation of public consensus on issues.

However, addressing polarization requires fully understanding the dynamics of polarization and identifying the possible causes. Only then we can better discuss how to attend to polarization through systematic on-platform changes or regulation of social media platforms. Current attempts and efforts to change social media are a good step in creating an environment with less polarization. I hope this dissertation contributes to this endeavor.

In this final chapter, I summarize the findings and highlight the implications of the results. Chapters 4 thru 6 provide evidence of negative or uncivil language affecting social engagement and affective polarization in both aggregated and individual levels of analysis. Putting the findings all together helps to answer the overarching questions proposed at the beginning of this dissertation, namely how and what we can learn from language about polarization. Table 7.1 summarizes the findings across the chapters.

Table 7.1: Summary of Findings Across the Chapters 4, 5 and 6.

	Aggregate Analysis (Chs. 4 & 5)	Individual Analysis (Ch. 6)
Negativity & Engagement	Negativity (a) increased Comments, and Angry, Sad, and Wow reactions, and (b) decreased shares and Love, Like, and Haha reactions (Table 4.3). Effects of negativity were stronger for Comments, Likes, and Wows on Republican pages. Effects of negativity were stronger for Shares and Angry, Love, Sad, and Haha reactions on Democratic pages.	No statistically significant findings
Incivility & Engagement	Incivility (a) increased Comments, Shares, Angry, Love, Like, Sad, Haha, and Wow reactions (Table 4.3). Effects of incivility were stronger for Love on Republican pages. Effects of incivility were stronger for Comments, Like, Sad, and Haha reactions on Democratic pages.	Compared to neutrality, incivility increased social engagement in Comments and Angry reactions (Table 6.1).
Negativity & Polarization	More negative posts from Republican pages mentioning both in-/out-group entities corresponded with increased polarization (Table 5.5.1).	Negativity increased polarization for the partisan emotion and feeling thermometer measures (Table 6.2.1).

		The effect appeared for Republicans on partisan emotions and trait evaluation but not for Democrats (Table 6.2.2)
Incivility & Polarization	More uncivil posts from Democratic pages corresponded with greater favorability toward Republicans on both polarization measures (Table 5.3.2; Figure 5.3.1). More uncivil posts from Republican pages corresponded with greater favorability toward Democrats on the partisan feeling thermometer measure (Table 5.3.2; Figure 5.3.2). More uncivil posts that mentioned out-group entities corresponded with increased affective polarization for the partisan feeling thermometer measure (Table 5.4.2) and more uncivil posts that mentioned both in-/out-group entities corresponded with decreased affective polarization for the partisan sentiment measure (Table 5.4.2). The effect appeared for Republican posts (on both polarization measures) but not for Democratic posts (Table 5.5.2).	No statistically significant findings (Depolarizing effects of presence of in-group targets were found for Republicans on partisan emotions compared to that of non-explicit targets in Table 6.3).

HOW LANGUAGE CAN ANSWER THE DYNAMICS OF POLARIZATION

Language is not merely written text. Rather, language is a type of complex information. It is a comprehensive, flexible, and evolving indicator of the political and social atmosphere, which needs interdisciplinary understandings across communication, political science, psychology, computer science, and many other relevant areas. In this dissertation, I pointed out several language features that could help us understand the current levels of political polarization, which is a serious problem for American democracy.

First Route to Polarization: Negative and Uncivil Language Attracts More Engagement Leading to More Visibility

In this dissertation, I proposed that language featuring negativity and incivility in political content attracts more engagement on social media, which influences algorithms that populate News Feeds on Facebook such that this type of content is boosted and more visible. Greater visibility for such content creates more chances of exposure, which may ultimately lead to polarization. In Chapter 4, the findings clearly demonstrate that negative and uncivil language used in political posts attracts more social engagement than the absence of these attributes. Negativity alone receives a greater number of Comments, Shares, and Angry, Sad and Wow reactions. Incivility alone garners more Comments, Shares, and Angry, Like, Sad, Haha, and Wow reactions. Given that posts that appear within News Feeds are decided by connections (i.e., who users are friends with on social network) and activities (i.e., how users and their friends/family interact to posts and how many interactions a post received) (Facebook, n.d.), both negativity and incivility in Facebook hyper-partisan pages influence what users and their friends see on on Facebook at the aggregated level.

In Chapter 6, the individual level of analysis also finds higher engagement with uncivil posts compared to neutral posts. Those who were exposed to uncivil posts were more likely to engage with Comments and Angry reactions than those exposed to neutral posts, which indicates criticism and disapproval. The individual level of analysis confirms higher levels of engagement with incivility, which can lead to greater algorithmic prioritization of this content.

Second Route to Polarization: Negative and Uncivil Language is Consonant with How People Think of Politics Now

This dissertation also examined the direct effects of exposure to negative and uncivil content on polarization. In Chapter 5, negativity and incivility in political content aggregated on a quarterly basis do not necessarily affect how people feel about the two political parties. Instead, contemporaneous effects of both negativity and incivility correspond with how polarized people feel, although the effects were only marginally significant. Much clearer evidence shows up when political posts are aggregated by the source's partisan leanings. When aggregated by posts' political leanings, uncivil posts from both Republican and Democratic pages are related to affective polarization. Negative posts with both in-/out-group targets from Republican pages are also associated with increased polarization. A possible explanation is that people are drawn to mediated information about conflict (Mutz, 2006b; Soroka, 2014; Trussler, & Soroka, 2014), negativity and incivility accompanied by in-/out-group targets may lead to decreased respect and intense affective reactions toward the opposing party. Also, for Republicans, it is possible that compared to Democrats they are prone to negative information (Hibbing et al., 2014) and prefer simple and straightforward communication style (Jarvis, 2004; Tetlock, 1983).

Earlier research finds that of the use of politically charged language fluctuates depending on what is happening politically among elites (Gentzkow & Shapiro, 2010; Jensen et al., 2012; Luntz, 2004), in news media (Gentzkow & Shapiro, 2010; Martin & Yurukoglu, 2017), and on social media (Colleoni et al., 2014; Conover et al., 2011). Likewise, this dissertation finds that affectively charged language on political content is connected with public opinion on polarization. Because of the limitations of the three-month aggregation used here, we do not gain insight into the causal direction of the relationships between language and polarization. It could be that messages from hyper-partisan Facebook pages influence how the public in general feels about the political parties, or the general atmosphere of the public affects the language usage of hyper-partisan pages, or that both correspond to elite political realities. What we can conclude from the aggregate analysis is that negativity and incivility expressed from partisan pages correspond with how the public thinks about the political parties.

Evidence from an experiment in Chapter 6 offers a somewhat similar, but distinct view. Exposure to negative content polarizes compared to exposure to neutral content after controlling for the effects of presence of political targets. Partisan differences exist with these patterns, however. Republicans who were exposed to negative political content affectively polarized compared to those who saw neutral political content, while there was no comparable evidence for Democrats. Other effects were also significant only among Republicans. Republicans who viewed political content targeting in-groups had less polarized partisan emotions compared to those who saw political content with non-explicit entities. The finding is consistent with evidence of depolarization by exposure to counter-attitudinal content (Barberá, 2014; Beam et al., 2018). Additional analysis of the open-ended responses about how participants feel about the posts that they encountered show that there are differences in the negativity and incivility in their responses. When participants saw negative or uncivil posts, their responses were more negative than those who view neutral posts. Likewise, when they are exposed to uncivil or in-group targets, their responses are also uncivil, compared to those are exposed to neutral or non-explicit targets content. It is clear that negative or uncivil language psychologically affects how people evaluate and feel about political entities, which may result in polarization. Overall, the evidence of polarizing effects from exposure to negative and uncivil language is limited. There are indications of a connection in the aggregate analyses, but the individual analyses only show limited effects in response to negative content. It is possible that earlier in Facebook's trajectory, there was a stronger relationship – which could have been detected by the longer time period included in the aggregate analyses – but that people are accustomed to this language now. Future research will be needed to sort through this speculation, however.

Incivility Does Not Always Have Stronger Effects than Negativity

Whether incivility has the stronger effects than negativity was tested on engagement and polarization. Findings from the aggregated level in Chapter 4 and the individual level in Chapter 6 show that incivility has stronger effects in receiving Comments, Shares, and Haha reactions on political posts than negativity while negativity has greater impacts on receiving Angry, Sad and Wow reactions and losing Love and Like reactions compared to incivility. Incivility does not always attract more engagement at the aggregated level. However, given that Comments and Shares entail more actions and cognitive resources (e.g., considering the potential audience of their friend network or typing one's thoughts in the post) than clicks, the findings suggest that uncivil content yields more effortful engagement (Kim et al., 2021; Masullo Chen & Lu, 2017; Wang & Silva, 2018). Also, because incivility invites more Haha reactions than negativity does, humorous and sarcastic components of uncivil content can appeal to Facebook users. On the other hand, there is mixed evidence in an individual level of comparison between negative and uncivil posts in Chapter 6. When each metric is examined and presence of political targets is held constant, uncivil content receives more engagement in Comments and Angry reactions when the neutral post is a reference. Evidence is consistent only for Angry reactions in that incivility has the stronger effects on social engagement than negativity, which is only marginally significant after controlling for the presence of political targets.

In Chapter 5, comparing negativity and incivility in response to affective polarization in the public show higher effects for incivility than negativity when posts aggregated by political leanings. The effects of incivility from both Democratic and Republican posts separately are significantly greater than those of negativity. In Chapter 6, however, I did not find that effects of incivility are greater than those of negativity in response to affective polarization in the individual level. Due to a lack of consistency, I conclude that uncivil posts do not always have a greater impact on polarization that negative posts do.

Presence of Political Targets (de)Polarizes

When negative and uncivil content is present in posts that also mention political targets, the dynamics show a distinct path for incivility. In Chapter 5, I found that negativity and incivility associated with political targets correlates with affective polarization. Findings show that negativity associated with in-groups by themselves reduced affective polarization, a marginally significant effect, while negativity linked to both in-/out-groups marginally increased affective polarization. For incivility, uncivil posts associated with out-groups increased affective polarization while uncivil posts linked to both in-/out groups reduced affective polarization. These findings suggest that negativity matters in the presence of in-groups while incivility matters in the presence of out-groups. It is possible that the presence of in-groups with other components of negativity such as sadness or worries influences polarization. For incivility, it is likely that presence of out-groups alone makes uncivil posts likeminded and less uncivil to the co-partisans (e.g., Muddiman & Stroud, 2017), which possibly aggravates polarization. Indeed, out-group derogation is manifestation of identity driven polarization (Iyengar et al., 2012; Mason, 2015; 2018). Perhaps the presence of in-groups and out-groups alongside uncivil content is seen favorably because incivility is not judged as harshly when both groups are present in a post.

In Chapter 6, political targets at the individual level affected polarization in some cases, especially for Republicans. Holding the effects of neutral, negative, or uncivil terms constant, in-group mentions in political content yields less polarization among Republican identifiers compared to the presence of non-explicit entities in political content although this pattern is statistically significant on one affective measure. Past research has shown that exposure to counter-attitudinal content (i.e., presence of in-group targets in my dissertation) leads to less polarization (e.g., Barberá, 2014; Beam et al., 2018). The study confirmed that the presence of political targets can influence affective polarization in an individual level.

Few People Engage with Content but Exposure is Effective

I analyzed the open-ended question of how people feel or think about the political posts in the survey in Chapter 6. Many participants responded that they did not engage with the political posts or ignored them, which is consistent that higher zero rates found in the data of how often participants interacted with the political content in the experiment. Some respondents also mentioned that they would report or block political content if it was in their timeline. This is consistent with findings from the Pew Research Center that most people do not actively engage with politics online (Mcclain, 2021) and that only a small number of people who are more politically extreme share and engage with partisan content on social media (Blazina, 2022). Due to their disinterest, many people on Facebook may not be regularly exposed to affective political content in their News Feed. Those who proactively report and block such harmful content may be able to create less polarizing News Feeds.

However, because of people's social connections and the activities of their broader network of friends and family, chances exist that people who are not active in political engagement can be exposed to this content. This is because (1) users on Facebook tend to have two friends who are opposing partisans or ideologues out of ten on average (Bakshy et al., 2015; Goel et al., 2010) and (2) politically attentive people are likely to be those who are active in engaging with political content even though they are a small group of people (Weeks, Ardèvol-Abreu, et al., 2017). Unless users actively block or report partisan content that they encounter or friends that interact with such content, there are still possibilities of exposure to hyper-partisan content. Although political content accounts for only 6% of the total content on users' News Feed, most engaging and spreading content is also political, and includes several of the hyper-partisan pages that I examined in this dissertation (Facebook, 2020). Facebook's function as a major source for political information is also supported by research on selective and incidental exposure to such content (e.g., Bode, 2016; Cacciatore et. al, 2018; Feezell, 2018; Oeldorf-Hirsch, 2018; Wells & Thorson, 2017). As people use Facebook more often, avoiding political content completely on one's personal News Feed is unlikely.

People may change their views because of something that they saw on social media, which indicates less polarization (e.g., moving from extreme positions on political issues to moderate positions) (Perrin, 2020). However, in case of negative content in this dissertation, some people who saw such political posts, particularly Republicans, switched their attitudes and became more extreme and polarized as found in Chapter 6. This confirms the findings of other research suggesting that the effects of incivility and negativity may not be uniform for all individuals (Nai & Maier, 2021).

Partisan Differences Matter

In Chapters 4 to 6, I explored a series of hypotheses and research questions about asymmetrical partisan polarization resulting from polarizing language. In Chapter 4, partisan asymmetrical engagement happens on majority of metrics (Shares, Angry, Love, Sad, Haha, and Wow), where Democratic posts are more likely to receive reactions compared to Republican posts regardless of whether the posts contain negativity or incivility. There are also partisan differences in social engagement depending on whether the posts are more negative or uncivil, but the differences are notably small.

Partisan differences in the dynamics of polarization are most evident in Chapter 5. When aggregating posts by political leanings, partisan asymmetric effects exist. The effects of negativity in the presence of political targets from Democratic posts is marginally significant in predicting affective polarization. Negativity and incivility in the presence of political targets from Republican posts correlate with affective polarization. Incivility with out-groups mentions from Republican posts in particular increase affective polarization among the public. This may indicate that Republicans who are the main audience for Republican posts are more convinced by uncivil posts with out-group targets and their responses are driving the polarization findings.

Partisan asymmetric effects of polarizing language are further supported from Chapter 6. In a similar manner, those who identified Republicans are more likely to polarize due to negative content compared to exposure to neutral content, while those who identified as Democrats do not show any differences in exposure to negative or uncivil language. In making sense of findings from Chapters 5 and 6, I conclude that Republicans are more likely to respond to negative content through affective polarization compared to Democrats.

Some Remaining Questions

Studying language in partisan pages provided more insight into the mechanisms of polarization via social media. Yet, there are some remaining questions that are worth pursuing.

Cheap Participation on Social Media as Political Engagement

Through this dissertation, I use the word "engagement" to identify the mechanical devices users can use to interact with posts and their friends on social media such as Shares, Comments and Reactions. "Social engagement" is a term that the those in academia and industry commonly use, so I also make use of the term. Yet this may not be the best word. Perhaps engagement should mean something deeper than clicking the digital content, in the vein of past research on political and civic engagement that involved more serious commitments to participating in offline environments (Zukin et al., 2006). Some may point out that the use of the word “engagement” for describing interactions with political content on social media oversimplifies the concept of engagement in the sense that activities on social media may be a mere responsiveness to the content and that (political) engagement means a status of genuine and active involvement with a matter.

Related to this argument, Lu and Myrick (2016) distinguished cheap political participation on social media (e.g., sharing, commenting on, and paying mere attention to political information) from costly political participation (e.g., donating money and volunteering time to political organizations). In their study, anger generated from cross-cutting exposure on Facebook led to the low-cost activities. Likewise, evidence from Chapter 6 supports that Angry reactions receive greater interactions from those who saw uncivil content compared to those who saw neutral content and also from those who viewed posts with an in-group presence (i.e., counter-attitudinal information) compared to those who viewed posts with a non-explicit entity presence. Those findings depict the current partisan polarization driven by affect on social media, which can be seen as emotional reactance or an affective affordance, rather than substantive engagement with politics.

Expected Outcomes from Other Platforms

Findings of my dissertation, namely how language in Facebook posts on partisan pages affect affective polarization, may have implications for Twitter and other social media. For instance, Twitter and YouTube have similar metrics that enable users to interact with other users and allow people to see social endorsement cues such as how many the content is retweeted (or viewed), commented on, or liked. The mechanical devices for engagement with political content on Facebook may work the same way for other social media platforms such as Twitter or YouTube by feeding algorithms. As a result, more negative and uncivil content may draw greater interactions from users and ultimately become more visible on people's personal Feeds on other social media platforms that algorithmically decide on what content users will see.

Also, past research finds evidence of homogeneous networks and echo chambers on Twitter (e.g., Colleoni et al., 2014; Conover et al., 2011; although see Barbera et al., 2015) and on YouTube (e.g., Bessi et al., 2016), which indicates higher chances of selective exposure and lower chances of cross-cutting incidental exposure in general. Exposure to negative and uncivil content in politics may lead to increased polarization on other social media, even though there may be factors that affect users differently on other social media such as whether users perceive negativity or incivility in political discourse as problematic (e.g., Mcclain et al., 2021).

SUGGESTIONS FOR FUTURE RESEARCH

To examine polarizing language in this dissertation, I used computational social scientific approaches to process massive amounts of text data. However, polarizing language can be found in formats other than text, such as images and videos, and computational methods can be used to study these other formats. For instance, incivility can be perceived differently depending on the modality (Sydnor, 2018). Likewise, similar patterns of language can be identified across platforms and may have different effects depending on where they appear. Future research should tackle these other formats.

As hinted at earlier, identifying how language can be used for depolarization will be an important next step. This could include investigating other forms of online political discourse, such as political slang, civility, and empathy. Furthermore, questions of how political language interventions could work to foster political tolerance on social media platforms and online community health can be raised. Depending on what features are available on a social media platform, different approaches will be needed to address political divisions in online communities. To be more specific, field experiments or longitudinal experiments will be desirable.

LIMITATIONS

As with all research, this dissertation has limitations. First, there are only about 50 observations in the time series analysis, which limited my ability to construct more complex multivariate regression models. I tried to add no more than five variables to test the effects of negative, uncivil or presence of targets as independent variables when running a regression with the residuals of ARFIMA model. The lack of data was because public opinion surveys of affective polarization are conducted irregularly, which made it impossible to look at smaller date ranges. Taking smaller gaps than quarters in a year would have been more ideal to detect possible lagged effects of polarizing language on aggregated affective polarization, but it was not possible because it generated a substantial number of missing values which would have created other problems such as less confidence about data quality created from too much interpolation. The findings may be different using a smaller lag such as bi-weekly or monthly. Related to this issue, if there were more observations, comparisons of more exogenous variables would have been possible, which could have provided a more comprehensive analysis of the language variables.

Second, in the individual analysis, there are several limitations from the experiment due to sacrificing external validity for internal validity. I tried to make the News Feed in the experiment resembles an actual Facebook setting, but it is an artificial setting that does not allow us to observe what participants would do on their own timelines. In a similar manner, the experimental stimuli were created to have consistent factors. This means that it necessarily contained posts that may differ from the actual posts that participants encounter. However, these limitations were acknowledged and purposefully chosen in favor of internal validity to allow for causal comparisons across the experimental conditions.

In addition, the sample for the experiment was not representative of the U.S. population. More representative samples would increase the generalizability of the findings. Related to this, multiple studies and/or panel studies to understand long-term effects conducted among representative samples should be considered.

CONCLUSION

Negative and uncivil language affects polarization. Negative and uncivil terms in political content attracts engagement, which enables more visibility through algorithms and social networks. These attributes of language also influence affective polarization. Incivility is more likely to influence polarization for Republicans than for Democrats. Political targets influence the effects of negativity and incivility in political content. People say that they do not want to engage with political content and would ignore it, and most people do not interact with political posts. However, the effects of exposure are compelling for individuals. Because language is the key player in polarization, future efforts to depolarize also should look to political language.

Appendix A: Coding Guideline for Political Pages on Facebook

Broken Page (1 = yes, 0 = no):

- Code 1 if the page is not available or not found, and stop coding
- Code 0, otherwise.

Admin Location & Language (E, F, or M): see the Page Transparency section and the recent three posts (any text areas),

- Code E if the page manager location is from United States and the three posts are written in English
- Code F (Stop coding here)
    - if the page manager location is not from United States OR,
    - if none of the three posts is written in English.
- Code M
    - if the page manager location is mixed of United States and other countries (e.g., U.S., and Singapore) and the three posts are written in English
    - if the page manager location is from United States and at least one of the three posts is written in English
- Code NA if the location information is not available

Verified Public Figure (1 = yes, 0 = no):

- Code 1 if the profile is verified public figures (individuals, NOT organizations) with a blue verified badge, and stop coding
- Code 0, otherwise.

Government Page (1 = yes, 0 = no):

- Code 1 if the page is an official page for government department, and stop coding
- Code 0, otherwise.

Official News Organizations (1 = yes, 0 = no):

- Code 1 if the page is an official news organization (e.g., news media, news/media company, magazine, ect) page with a blue verified badge, and stop coding
- Code 0, otherwise

Political Pages (1 = yes, 0 = no): see the recent 10 posts between <file name>, 2021 and October 2020), and (do not see the about/bio sections for this category)
if the page is not currently active, just see the recent 10 posts.

- Code 1 if majority posts (i.e., 7 out of 10) are about politics:
    - political content can be as following but not limited to: (shared) news stories about politics and issues, images or text messages that contain political entities or issues, urge political actions, invite to political events.
    - content related to topics (e.g., religion, police, animal, health, environment, gun etc) should be tied to politics)
    - do not count commercials, ads, or promotion as political content.
- Code 0, otherwise (stop coding)

Note. “Pinned Post” is not included as recent posts.

Partisan Leanings (R = Republican, D = Democratic, 0 = Other, unsure): see the About section and the recent 10 posts between <file name>, 2021 and October 2020) -- the page’s bio and/or name and the content should clearly:

- support and/or are leaning toward Republican Party, conservatives, Republican politicians etc., OR are against the opposing party (Democratic Party, liberals, Democratic politicians), code R.
- support and/or are leaning toward Democratic Party, liberals, Democratic politicians etc., OR are against the opposing party (Republican Party, conservatives, Republican politicians), code D
- Otherwise, code 0 if the page supports other parties, not clearly leaning toward either parties, or unsure (e.g., libertarian party as 0; if the about/bio clearly shows that they’re posting content about both sides, code as 0)
- Some political issues can be an indicator of partisanship, but it should be clearly demonstrated in the context of supporting/opposing partisanship or parties in any of the posts.

Appendix B: Coding Guideline for Negativity & Incivility of Political Text

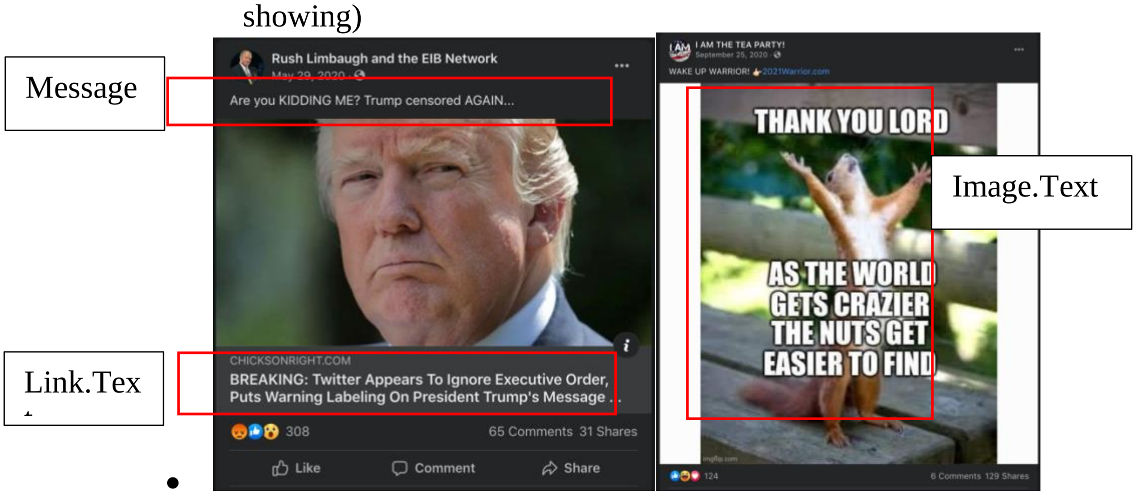
Guideline for Negative Text

Please decide whether the political posts indicate valence or emotions, using the following scale. If needed, use a sentence as a unit of sentiment to cancel out other sentence if the case is mixed of sentiment.

- Valence refers to positive and negative affectivity.
- Emotions can be positive (e.g., joy, affection, pride, nostalgia, on FB emotions, love, care, haha (funny) etc.) or negative (e.g., sadness, anger, fear, disgust, remorse, rancor, anxiety on FB emotions sad, angry etc.)

Try not to use political knowledge and bias as much as possible:

- Think about implied emotions from the writer that they want from the readers
    - e.g., 🔥 Best response to the barrage of black boxes and unproductive hashtag nonsense... The Blackout Tuesday Black Boxes On Social Media If you're on Instagram, you've probably seen the barrage of black boxes that people are posting ostensibly in support of the Black Lives Matter movement. It's a hashtag campaign called #blackouttuesday. I'm not participating, because → Negative
    - e.g., 'RIP GOP': Polling expert predicts devastating Trump 2020 electoral wipeout. Only you can make this true. #VoteBlueNoMatterWho 'RIP GOP': Polling expert predicts devastating Trump 2020 electoral wipeout Democratic pollster and strategist Stanley Greenberg went on CNN Monday to explain why he believes President Donald Trump and the Republican Party are headed toward a historic and humiliating defeat in the 2020 elections. In an interview with CNN's John Berman, Greenberg discussed his new book cal... → positive
    - Evaluate emotions from text as following order: (1) Message, (2) Image.Text AND Link.Text, and (3) Description.
        - Message: text messages under the page name
        - Image.Text: text from the shared image/photo content
        - Link.Text: (often) headlines or titles of the shared content
        - Description: some text of the first paragraph of the shared content (not showing)



- If it’s still unclear, focus on the general content (not the way the text is stated): e.g., violent or sad stories although it’s neutrally stated → negative

(1) positive (e.g., appreciation, touching stories, stories that the writer intend to evoke positive reactions although some negative emotions include):
   a. To show his wife just how much he cares for her, Greg and their five children came up with a plan... This is the best thing you'll see all week. Husband Gets Down on One Knee 37 Years After Wedding, Proposes to Terminally Ill Wife for 2nd Time
   b. e.g., Thank you, Donald Trump! Hillary Clinton Calls Donald Trump a RACIST... And Trump FIRES Back in an EPIC Way!
   c. e.g., WHOA... "Out of Control!" Trump Picks Fight with Lockheed Martin Over F-35 Cost, They Immediately Respond... Love it!

(2) neutral (e.g., descriptions, regular updates; is it something you can see as a headline?),
   a. Join U.S. Democratic Socialists Activists Group
   b. Leaked memo says Trump wants to allow U.S. companies to buy conflict diamonds and minerals The Intercept is reporting on a leaked memo they received that would loosen the provision in the Dodd Frank Act that forced companies to audit their supply chains when it came to collecting mineral resources like diamonds and gold and all of the things...
   c. In some cases that cause mixed feelings (it’s unclear even after considering the implied emotions from the writer and the general content), you can code it as mixed (please be cautious to code as mixed)
   d. using mild language: Did it kind of look like Ainsley Earhardt gave Brian Kilmeade an old fashion "duh" to his comment here? Even Ainsley Thinks Brian Kilmeade Is a Total Goof

(3) negative (e.g., stories that cause anger, anxiety, sadness, using profanity, punctuation marks, capital words, making some extreme cases or personal attack/name-calling, stated in unnecessary/nonessential descriptions)
   a. e.g., The evidence is there. FLASHBACK: How Many Times Did Hillary Say She Didn't Send Classified Info Over Email? - Chicks... Lying out of her lying liarhole.)
   b. personal attack: e.g., The former Superman actor called Rep. Ilhan Omar out on her really bad math and blew up the internet. 'Superman' Does Simple Math, Rips Apart Congresswoman's Minimum-Wage Example After Rep. Ilhan Omar's clip about the minimum wage went viral, Dean Cain did a little bit of math.
   c. using unnecessary descriptions: e.g., This pass-the-buck excuse doesn't work when you're the global leader in internet technology, Google. For 2nd Time This Week, Google Mislabels Republican As Hate Monger This pass-the-buck excuse doesn't work when you're the global leader in internet technology, Google.

(4) NA (not codable)
   a. Try not to code as not codable
   b. e.g., Timeline photos or their page names

Guideline for Uncivil Text

Please decide whether the political posts contain incivility using a binary scale (code 1 for incivility, 0 otherwise). The posts contain mostly negative valence/emotions. Please make sure to distinguish incivility from negative valence or emotions.

Try not to use political knowledge and bias as much as possible.

Negativity

- Valence refers to positive and negative affectivity.
- Emotions can be positive (e.g., joy, affection, pride, nostalgia, on FB emotions, love, care, haha (funny) etc.) or negative (e.g., sadness, anger, fear, disgust, remorse, rancor, anxiety on FB emotions sad, angry etc.)
- negative (e.g., stories that cause anger, anxiety, sadness)

Incivility

- code uncivil if the post clearly shows or contains any of the following:
    - with clear target (and some clear context or substance e.g., racists)
    - making fun of the target, joking
    - profanity, abusive, or insulting language (e.g., what the heck?)
    - punctuation marks or capital words (not just part but as entire messages)
    - personal attack or name-calling,
    - hashtags of uncivil/targeted words (even some parts of words)
    - stated in unnecessary/nonessential descriptions, etc.
    - cynicism/criticism AND being malicious, harmful (negativity can be somewhat mean but not necessarily uncivil)
    - you should be able to justify why it is uncivil
- For instances:
    - name-calling: The evidence is there. FLASHBACK: How Many Times Did Hillary Say She Didn't Send Classified Info Over Email? - Chicks... Lying out of her lying liarhole.
    - personal or targeted attack: The former Superman actor called Rep. Ilhan Omar out on her really bad math and blew up the internet. 'Superman' Does Simple Math, Rips Apart Congresswoman's Minimum-Wage Example After Rep. Ilhan Omar's clip about the minimum wage went viral, Dean Cain did a little bit of math.
    - using unnecessary descriptions: This pass-the-buck excuse doesn't work when you're the global leader in internet technology, Google. For 2nd Time This Week, Google Mislabels Republican As Hate Monger This pass-the-buck excuse doesn't work when you're the global leader in internet technology, Google.

Target:

if people or group(s) are targeted; write down the name/word.

Appendix C: Experimental Messages & Stimuli

Experimental Messages with Negativity and Incivility Scores

For attributes, NG = negative, UC = uncivil, TR = neutral; For targets, D = Democrats, R = Republicans, N = non-explicit

Topic	Attributes	Target	Messages	Sentiment	Incivility
gun	NG	D	Prepare for the liberal strategy. When a mass shooting happens, they immediately start throwing out gun control. Liberals use mass shootings as reasons to destroy guns! If you have a good guy with a gun on site, there would be fewer victims. Guns are for protection.	-2	0.17
gun	UC	D	Prepare for the liberal brainwashing. When a mass shooting happens, they immediately start throwing out the f*king gun control bullshit. Liberal idiots use mass shootings as reasons to destroy guns! If you have a good guy with a gun on site, there would be fewer victims. Guns are for protection!!	-2	0.91
gun	TR	D	Question the liberal argument. Every time a mass shooting happens, liberals immediately start talking about gun control. If you have a good guy with a gun on site, there would be more survivors. The liberal argument cannot fix the root cause of shootings, which isn't guns.	0	0.12
gun	NG	R	Prepare for the conservative strategy. They scream good people do not need gun control and they own a gun for protection. When a mass shooting happens, however, conservatives don't want to hear the voices of all the victims. Shooting victims have been voiceless since conservatives stole the conversation!	-2	0.21
gun	UC	R	Prepare for the conservative brainwashing. They scream good people do not need gun control and they own a gun for protection. When a mass shooting happens, however, conservative idiots don't want to hear the f*king voices of all the victims! Shooting victims have been voiceless since conservatives stole the conversation!!	-2	0.91
gun	TR	R	Question the conservative argument. They say good people do not need gun control and they own a gun for protection. When a mass shooting happens, however, conservatives don't want to hear the voice of all the survivors. The conservative argument cannot fix the problem of shooting, which is guns.	0	0.09
gun	NG	N	It's time to educate yourself! A third of Americans personally own a gun and protection is the top reason for owning a gun. At the same time, half of	-2	0.21

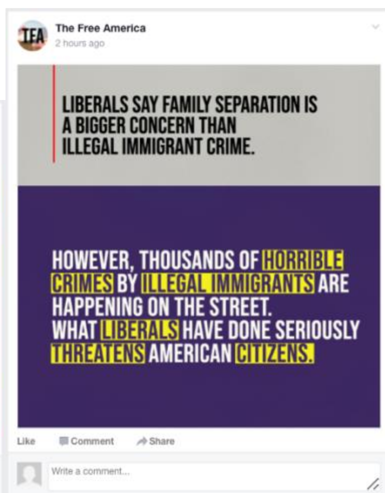
			Americans see gun violence and mass shooting victims as a very big problem in the country today and favor stricter gun control.		
gun	UC	N	A bunch of idiots do not know f*king facts about guns!! A third of Americans personally own a gun and protection is the top reason for owning a gun. At the same time, half of Americans see gun violence and mass shooting victims as a very big problem in the country today and favor stricter gun control.	-2	0.93
gun	TR	N	Here are some facts about guns in America. A third of Americans personally own a gun and protection is the top reason for owning a gun. At the same time, half of Americans see mass shooting and shooting survivors as a very big problem in the country today and favor stricter gun control.	0	0.11
immigration	NG	D	Liberals say family separation is a bigger concern than illegal immigrant crime. However, thousands of horrible crimes by illegal immigrants are happening on the street. What liberals have done seriously threatens American citizens.	-3	0.30
immigration	UC	D	Irrational liberals scream that family separation is a bigger concern than illegal immigrant crime. However, thousands of horrible criminal illegal immigrants are released on the street. What stupid liberals have done seriously threatens American citizens.	-3	0.88
immigration	TR	D	Liberals say family separation is a bigger problem than undocumented immigrant crime. However, thousands of undocumented immigrants are associated with serious crimes. The liberals' idea is having a serious impact on American citizens.	-1	0.29
immigration	NG	R	Conservatives say illegal immigrant crime is a bigger concern than family separation. However, they continued to separate families and detain people in horrible conditions. What conservatives have done is seriously harming American immigrants' children and spouses.	-3	0.24
immigration	UC	R	Evil conservative monsters scream that illegal immigrant crime is a bigger concern than family separation. However, they continued to separate families and illegally detain people in horrible and abusive conditions. What lunatic conservatives have done is seriously harming American immigrants' children and spouses.	-3	0.77
immigration	TR	R	Conservatives say undocumented immigrant crime is a bigger problem than family separation. However, they continued to separate families and detain people in poor conditions. The conservatives' act is having a serious impact on American immigrants' children and spouses.	-1	0.21
immigration	NG	N	People do not know the facts about immigrants. Seven-in-ten Americans feel sympathetic toward immigrants who are in the U.S. illegally. On the other hand,	-3	0.31

			four out of ten Americans do think immigrants without documents are more likely to commit horrible crimes.		
immigration	UC	N	Look at the facts about immigrants, dumbass!! Seven-in-ten Americans feel sympathetic toward illegal immigrants who are in the U.S. On the other hand, four out of ten Americans do think immigrants without documents are more likely to commit horrible crimes.	-3	0.91
immigration	TR	N	Consider the facts about immigrants in the U.S. Seven-in-ten Americans feel sympathetic toward immigrants who are in the U.S. illegally. On the other hand, four out of ten Americans do think immigrants without documents are more likely to commit serious crimes.	-1	0.20
partisanship	NG	D	Our nation is devastated by Democratic politicians. This is getting out of control.	-4	0.09
partisanship	UC	D	Our nation is devastated by Democratic politicians. This is getting fucking crazy!	-4	0.77
partisanship	TR	D	Our nation is affected by Democratic politicians. This is getting out of control.	0	0.07
partisanship	NG	R	Our nation is devastated by Republican politicians. This is getting out of control.	-4	0.11
partisanship	UC	R	Our nation is devastated by Republican politicians. This is getting fucking crazy!	-4	0.80
partisanship	TR	R	Our nation is affected by Republican politicians. This is getting out of control.	0	0.10
partisanship	NG	N	Our nation is devastated by politicians. This is getting out of control.	-4	0.08
partisanship	UC	N	Our nation is devastated by politicians. This is getting fucking crazy!	-4	0.78
partisanship	TR	N	Our nation is affected by politicians. This is getting out of control.	0	0.06
COVID	NG	D	Liberals are trying to repeat every failed COVID-19 policy from the last two years! They think it will magically work now.	-3	0.19
COVID	UC	D	Left-wing morons are trying to repeat every failed COVID-19 policy from the last two years! They think it will magically work now.	-3	0.95
COVID	TR	D	Liberals are trying to repeat every COVID-19 policy from the last two years. They think it will work now.	0	0.11
COVID	NG	R	Conservatives are grumbling about every COVID-19 policy and risking people's health! They think they can declare themselves 'done with it.'	-3	0.07
COVID	UC	R	Right-wing morons are grumbling about every COVID-19 policy and risking people's health! They think they can declare themselves 'done with it.'	-3	0.93
COVID	TR	R	Conservatives like to complain about every COVID-19 policy. They think they can declare themselves 'done with it.'	0	0.09

COVID	NG	N	COVID-19 continues to threaten public health! Shifting COVID-19 policy creates confusion and distrust.	-3	0.04
COVID	UC	N	COVID-19 continues to threaten public health, morons! Shifting failed COVID-19 policy creates confusion and distrust.	-3	0.93
COVID	TR	N	From a recent survey, majorities of Americans report mixed reactions to shifting COVID-19 policy and the effects of these policies on public health.	0	0.00
BLM	NG	D	Democrats think BLM protests are peaceful! They are violent!	-3	0.56
BLM	UC	D	Stupid Democrats call BLM activists peaceful protesters! They are violent rioters!	-3	0.96
BLM	TR	D	Democrats are more likely to think BLM supporters are peaceful rather than violent.	-1	0.35
BLM	NG	R	Republicans think BLM protests are violent! They are peaceful protesters!	-3	0.34
BLM	UC	R	Stupid Republicans call BLM activists violent rioters! They are peaceful protesters!	-3	0.96
BLM	TR	R	Republicans are more likely to think BLM supporters are violent rather than peaceful.	-1	0.36
BLM	NG	N	People debate over whether BLM activists are peaceful or violent. Political division is terrible, worse than ever.	-3	0.32
BLM	UC	N	Hypocrites shout over whether BLM activists are peaceful or violent. Partisan division is terrible, worse than ever.	-3	0.80
BLM	TR	N	There is ongoing debate on whether BLM supporters are peaceful or violent. The answer depends on who you ask.	-1	0.16
media	NG	D	Left-wing agendas and lies are amplified by the liberal echo chamber.	-2	0.23
media	UC	D	Left-wing propaganda and lies are spread by the pathetic liberal echo chamber.	-2	0.88
media	TR	D	Left-wing agendas and half-truths are amplified by liberal media.	1	0.14
media	NG	R	Right-wing agenda and lies are amplified by the conservative echo chamber.	-2	0.22
media	UC	R	Right-wing propaganda and lies are spread by the pathetic conservative echo chamber.	-2	0.83
media	TR	R	Right-wing agenda and half-truths are amplified by conservative media.	1	0.09
media	NG	N	Agendas and lies are amplified by the media echo chamber.	-2	0.14
media	UC	N	Propaganda and lies are spread by the pathetic media echo chamber.	-2	0.80
media	TR	N	Agendas and half-truths are amplified by the media.	1	0.07

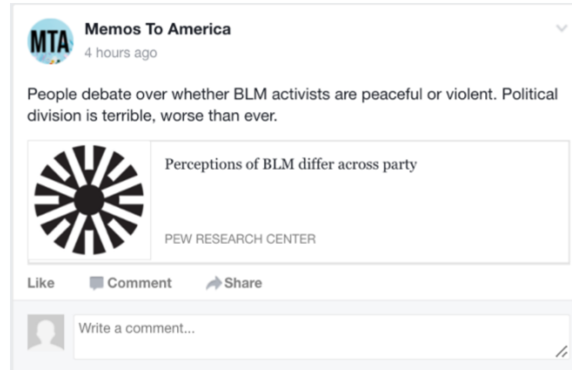
Stimuli Posts

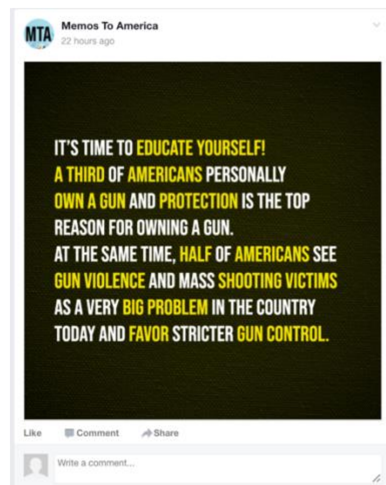
- Negative with Democrat Targets



- Negative with Republican Targets

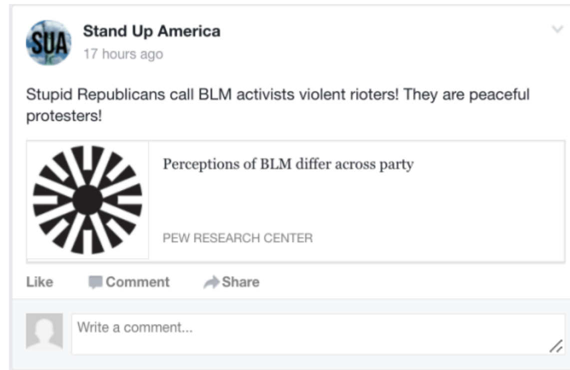
- Negative with None Targets

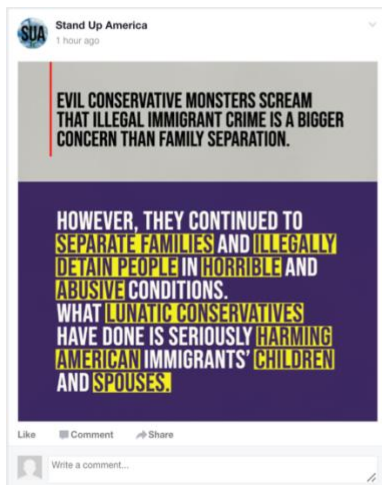




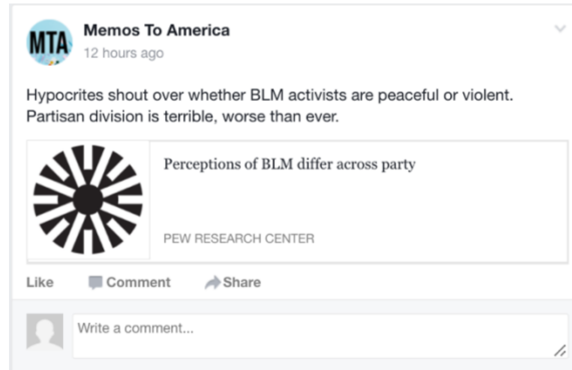
- Uncivil with Democrat Target

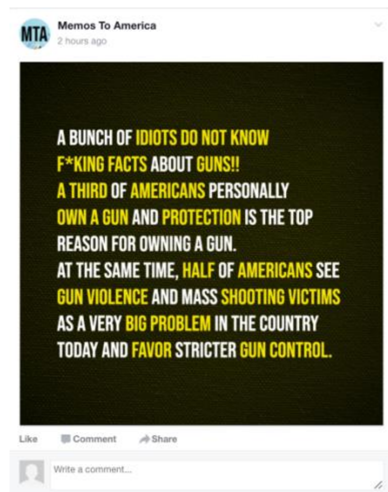
- Uncivil with Republican Targets



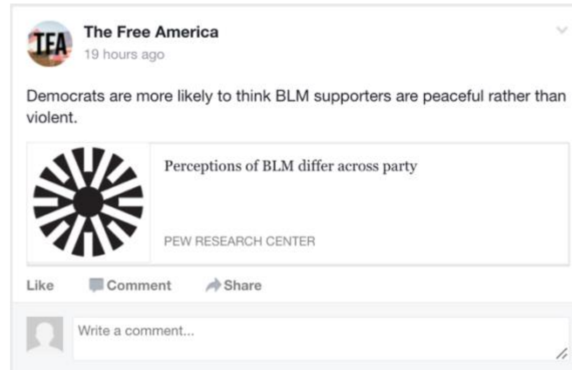


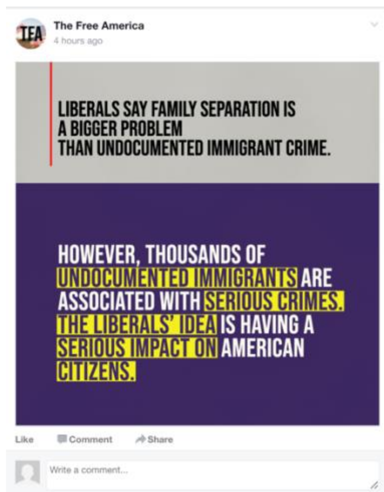
- Uncivil with None Targets

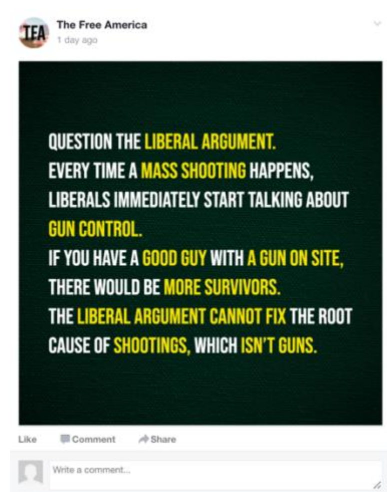




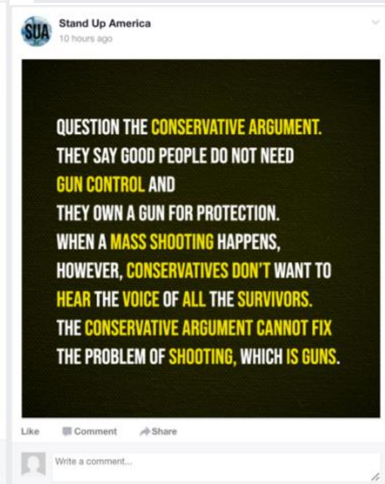
- Neutral with Democrat Targets



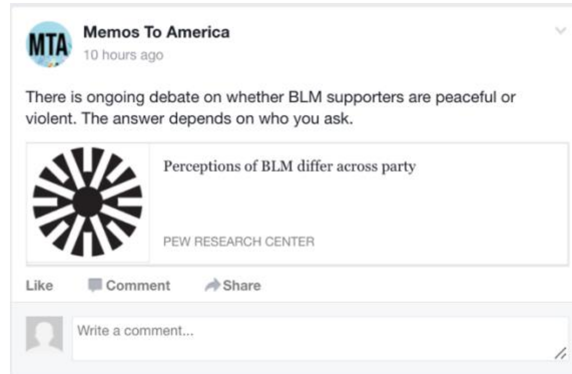


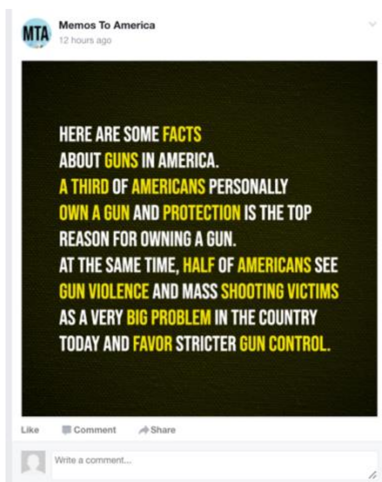


- Neutral with Republican Targets



- Neutral with None Targets





Appendix D: Survey Items for Individual Analysis

[Affect/Emotion]

Because of those who **support the Republican [Democratic] Party** in the United States or something they have done, I feel ...

- Angry
- Hostile
- Nervous
- Disgusted
- Anxious
- Afraid
- Hopeful
- Proud
- Enthusiastic

(0) a great deal, (1) somewhat, (2) very little, (3) not at all

[Feeling Thermometer]

We'd like to get your feelings toward some of our political entities on a "feeling thermometer." A rating of zero degrees means you feel as cold and negative as possible. A rating of 100 degrees means you feel as warm and positive as possible. You would rate the person at 50 degrees if you don't feel particularly positive or negative toward them.

- The Republican Party [0 – 100]
- The Democratic Party [0 – 100]

[Trait Evaluation]

Because of those who **support the Republican [Democratic] Party** in the United States or something they have done, I feel ...

- Intelligent
- Honest
- Open-minded
- Close-minded
- Lazy
- Selfish

(4) Strongly agree, (3) Agree, (2) Neither agree nor disagree, (1) Disagree, (0) Strongly disagree

[Party Identification]

Q1. Generally speaking, do you think of yourself as a/an:

- Democrat → Q1.2.
- Republican → Q1.2.
- Intendent → Q1.1.
- Other (please specify) → Q1.1.

Q1-1. Do you consider yourself closer to

- The Democratic Party
- The Republican Party
- Neither

Q1-2. Do you consider yourself

- A strong Democrat [Republican]
- Not a very strong Democrat [Republican]

References

Abramowitz, A. (2010). *The disappearing center: Engaged citizens, polarization, and American democracy*. New Haven, CT: Yale University Press.

Abramowitz, A. I., & Saunders, K. L. (2008). Is polarization a myth? *The Journal of Politics*, *70*(2), 542-555. doi: 10.1017/S0022381608080493

Allport, G. W., Clark, K., & Pettigrew, T. (1954). The nature of prejudice. Reading, Massachusetts: Addison-Wesley Publishing Company.

Amira, K., Wright, J. C., & Goya-Tocchetto, D. (2021). In-group love versus out-group hate: Which is more important to partisans and when? *Political Behavior*, *43*(2), 473-494. doi: 10.1007/s11109-019-09557-6

An, J., Quercia, D., & Crowcroft, J. (2014, October). Partisan sharing: Facebook evidence and societal consequences. In *Proceedings of the second ACM conference on Online social networks* (pp. 13-24).

Ansolabehere, S., Rodden, J., & Snyder Jr, J. M. (2006). Purple America. *Journal of Economic Perspectives*, 20(2), 97-118. doi: 10.1257/jep.20.2.97

Anspach, N. M. (2017). The new personal influence: How our Facebook friends influence the news we read. *Political Communication*, *34*(4), 590-606. doi: 10.1080/10584609.2017.1316329

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216-9221. doi: 10.1073/pnas.1804840115

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*(6239), 1130-1132. doi: 10.1126/science.aaa1160

Barberá, P. (2014). How social media reduces mass political polarization: Evidence from Germany, Spain, and the US. *Job Market Paper, New York University*, *46*. Retrieved from: http://rubenson.org/wp-content/uploads/2015/10/barbera-tpbw.pdf

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political analysis*, *23*(1), 76-91. doi: 10.1093/pan/mpu011

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber?. *Psychological science*, *26*(10), 1531-1542. doi: 10.1177/0956797615594620

Basile, A., Caselli, T., & Nissim, M. (2017). Predicting controversial news using Facebook reactions. *Proceedings of the Fourth Italian Conference on Computational Linguistics* (*CliC-it 2017*), 11-13. doi: 10.4000/books.aaccademia.2370

Baum, M. A., & Groeling, T. (2008). New media and the polarization of American political discourse. *Political Communication*, *25*(4), 345-365. doi: 10.1080/10584600802426965

Beam, M. A., Hutchens, M. J., & Hmielowski, J. D. (2018). Facebook news and (de) polarization: Reinforcing pirals in the 2016 US election. *Information, Communication & Society*, *21*(7), 940-958. doi: 10.1080/1369118X.2018.1444783

Bennett, W. L., & Iyengar, S. (2008). A new era of minimal effects? The changing foundations of political communication. *Journal of Communication*, *58*(4), 707-731. doi: 10.1111/j.1460-2466.2008.00410.x

Berger, J. (2013). *Contagious: Why things catch on*. New York, NY: Simon & Schuster.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political analysis*, *20*(3), 351-368. doi: 10.1093/pan/mpr057

Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. *First Monday*, *21*(11-7). Retrieved from https://ssrn.com/abstract=2982233

Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., Uzzi, B., & Quattrociocchi, W. (2016). Users polarization on Facebook and Youtube. *PloS one*, *11*(8). doi: 10.1371/journal.pone.0159641

Bishop, B. (2009). *The big sort: Why the clustering of like-minded America is tearing us apart*. New York: Houghton Mifflin Harcourt.

Blazina, C. (2022, January 5). Americans at the ends of the ideological spectrum are the most active in national politics, *Pew Research Center*. Retrieved from https://www.pewresearch.org/fact-tank/2022/01/05/americans-at-the-ends-of-the-ideological-spectrum-are-the-most-active-in-national-politics

Bode, L. (2012). Facebooking it to the polls: A study in online social networking and political behavior. *Journal of Information Technology & Politics*, *9*(4), 352-369.doi: 10.1080/19331681.2012.709045

Bode, L. (2016). Political news in the news feed: Learning politics from social media. *Mass Communication and Society*, *19*(1), 24-48. doi: 10.1080/15205436.2015.1045149

Brandt, M. J., Wetherell, G., & Reyna, C. (2014). Liberals and conservatives can show similarities in negativity bias. *Behavioral and Brain Sciences*, *37*(3), 307. doi: 10.1017/S0140525X13002513

Brooks, D. J., & Geer, J. G. (2007). Beyond negativity: The effects of incivility on the electorate. *American Journal of Political Science*, *51*(1), 1-16. doi: 10.1111/j.1540-5907.2007.00233.x

Brundidge, J. (2010). Encountering "difference" in the contemporary public sphere: The contribution of the Internet to the heterogeneity of political discussion networks. *Journal of Communication*, *60*(4), 680-700. doi: 10.1111/j.1460-2466.2010.01509.x

Budak, C., Goel, S., & Rao, J. M. (2016). Fair and balanced? Quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, *80*(S1), 250-271. doi: 10.1093/poq/nfw007

Cacciatore, M. A., Yeo, S. K., Scheufele, D. A., Xenos, M. A., Brossard, D., & Corley, E. A. (2018). Is Facebook making us dumber? Exploring social media use as a predictor of political knowledge. *Journalism & Mass Communication Quarterly*, *95*(2), 404-424. doi: 10.1177/1077699018770447

Cambridge Dictionary. (n.d.). Libtard. in *Cambridge Dictionary.* Retrieved from https://dictionary.cambridge.org/us/dictionary/english/libtard

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). The American Voter: Chicago: University of Chicago Press.

Campbell, J. E. (2008). An exceptional election: Performance, values, and crisis in the 2008 presidential election. In *The Forum*, 6(4), De Gruyter. doi: 10.2202/1540-8884.1276

Carnaghi, A., & Maass, A. (2007). In-group and out-group perspectives in the use of derogatory group labels: Gay versus fag. *Journal of Language and social Psychology*, *26*(2), 142-156. doi: 10.1177/0261927X07300077

Carraro, L., Castelli, L., & Macchiella, C. (2011). The automatic conservative: Ideology-based attentional asymmetries in the processing of valenced information. *PLoS One*, *6*(11), e26456. doi: 10.1371/journal.pone.0026456

Carsey, T. M., & Layman, G. C. (2006). Changing sides or changing minds? Party identification and policy preferences in the American electorate. *American Journal of Political Science*, *50*(2), 464-477. doi: 10.1111/j.1540-5907.2006.00196.x

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658-679. doi: 10.1111/jcom.12104

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, *64*(2), 317-332. doi: 10.1111/jcom.12084

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on twitter. In *Fifth International AAAI Conference on Weblogs and Social Media*. doi: 10.1109/PASSAT/SocialCom.2011.34

Conover, P. J., & Feldman, S. (1986). Emotional reactions to the economy: I'm mad as hell and I'm not going to take it anymore. *American Journal of Political Science*, *30*(1), 50-78. doi: 10.2307/2111294

Conroy, M., Feezell, J. T., & Guerrero, M. (2012). Facebook and political engagement: A study of online political group membership and offline political engagement. *Computers in Human Behavior*, *28*(5), 1535-1546. doi: 10.1016/j.chb.2012.03.012

Conversation AI (2017). Perspective Comment Analyzer API Documentation, GitHub Repository, https://github.com/conversationai/perspectiveapi

Coppock, A. (2019). Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Political Science Research and Methods*, *7*(3), 613-628. doi: 10.1017/psrm.2018.10

Cormode, G., & Krishnamurthy, B. (2008). Key differences between Web 1.0 and Web 2.0. *First Monday*, 13(6). doi: 10.5210/fm.v13i6.2125

CrowdTangle Team (2020a). Partisan Pages (D), CrowdTangle. *Facebook*, Menlo Park, California, United States. URLs: https://apps.crowdtangle.com/politicalpages/lists/1261361

CrowdTangle Team (2020b). Partisan Pages (R), CrowdTangle. *Facebook*, Menlo Park, California, United States. URLs: https://apps.crowdtangle.com/politicalpages/lists/1253992

De Boef, S., & Kellstedt, P. M. (2004). The political (and economic) origins of consumer confidence. *American Journal of Political Science*, *48*(4), 633-649. doi: 10.1111/j.0092-5853.2004.00092.x

De León, E., & Trilling, D. (2021). A sadness bias in political news sharing? The role of discrete emotions in the engagement and dissemination of political news on Facebook. *Social Media+ Society*, *7*(4), 20563051211059710. doi: 10.1177/20563051211059710

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Echo chambers: Emotional contagion and group polarization on Facebook. *Scientific Reports*, *6*, 37825. doi: 10.1038/srep37825

DeSteno, D., Dasgupta, N., Bartlett, M. Y., & Cajdric, A. (2004). Prejudice from thin air: The effect of emotion on automatic intergroup attitudes. *Psychological Science*, *15*(5), 319-324. doi: 10.1111/j.0956-7976.2004.00676.x

DeVito, M. A. (2017). From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism*, *5*(6), 753-773. doi: 10.1080/21670811.2016.1178592

DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American's social attitudes become more polarized? *American journal of Sociology*, *102*(3), 690-755. doi: 10.1086/230995

Dodd, M. D., Balzer, A., Jacobs, C. M., Gruszczynski, M. W., Smith, K. B., & Hibbing, J. R. (2012). The political left rolls with the good and the political right confronts the bad: connecting physiology and cognition to preferences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1589), 640-649. doi: 10.1098/rstb.2011.0268

Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. *American Political Science Review*, *107*(1), 57-79. doi: 10.1017/S0003055412000500

Duggan, M., & Smith, A. (2016, October 25). The political environment on social media, *Pew Research Center*, Retrieved from https://www.pewresearch.org/internet/2016/10/25/the-political-environment-on-social-media

Eberl, J. M., Tolochko, P., Jost, P., Heidenreich, T., & Boomgaarden, H. G. (2020). What's in a post? How sentiment and issue salience affect users' emotional reactions on Facebook. *Journal of Information Technology & Politics*, *17*(1), 48-65. doi: 10.1080/19331681.2019.1710318

Facebook. (2020, November 10). What Do People Actually See on Facebook in the US? Retrieved from https://about.fb.com/news/2020/11/what-do-people-actually-see-on-facebook-in-the-us/

Facebook. (2021). Reducing Political Content in News Feed. Retrieved from https://about.fb.com/news/2021/02/reducing-political-content-in-news-feed/

Facebook. (n.d.). How Feed Works. Retrieved from https://www.facebook.com/help/1155510281178725/?helpref=hc_fnav

Feddersen, T., & Gul, F. (2014). *Polarization and income inequality: A dynamic model of unequal democracy*. Working Paper. Retrieved from http://www.cireqmontreal.com/wp-content/uploads/2015/04/feddersen.pdf

Feezell, J. T. (2018). Agenda setting through social media: The importance of incidental news exposure and social filtering in the digital era. *Political Research Quarterly*, *71*(2), 482-494. doi: 10.1177/1065912917744895

Fine, J. A., & Hunt, M. F. (2021). Negativity and elite message diffusion on social media. *Political Behavior,* 1-19. doi: doi.org/10.1007/s11109-021-09740-8

Fiorina, M. P., & Abrams, S. J. (2008). Political polarization in the American public. *Annual Review of Political Science*, *11*, 563-588. doi: 10.1146/annurev.polisci.11.053106.153836

Fiorina, M., Abrams, S., & Pope, J. (2006). *Culture war? The myth of a polarized America*. New York: Pearson and Longman.

Fuller, W. A. (1996). *Introduction to statistical time series* (2nd ed.). New York: Wiley.

Garmur, M., King, G., Mukerjee, Z., Persily, N., & Silverman, B. (2019). CrowdTangle Platform and API, Harvard Dataverse, V2. doi: 10.7910/DVN/SCCQYD.

Garrett, R. K., Gvirsman, S. D., Johnson, B. K., Tsfati, Y., Neo, R., & Dal, A. (2014). Implications of pro- and counter-attitudinal information exposure for affective polarization. *Human Communication Research*, *40*(3), 309-332. doi: 10.1111/hcre.12028

Garz, M., Sörensen, J., & Stone, D. F. (2020). Partisan selective engagement: Evidence from Facebook. *Journal of Economic Behavior & Organization*, *177*, 91-108. doi: 10.1016/j.jebo.2020.06.016

Gearhart, S., & Zhang, W. (2015). "Was it something I said?""No, it was something you posted!" A study of the spiral of silence theory in social media contexts. *Cyberpsychology, Behavior, and Social Networking*, *18*(4), 208-213. doi: 10.1089/cyber.2014.0443

Geis, M. L. (1987). *The Language of politics*. New York: Springer-Verlag.

Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician*, *60*(4), 328-331. doi: 10.1198/000313006X152649

Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica*, *78*(1), 35-71. doi: 10.3982/ECTA7195

Gentzkow, M., Shapiro, J. M., & Taddy, M. (2016). *Measuring polarization in high-dimensional data: Method and application to congressional speech* (No. w22423). National Bureau of Economic Research. doi: 10.3386/w22423

Georgakopoulos, S. V., Tasoulis, S. K., Vrahatis, A. G., & Plagianakos, V. P. (2019). Convolutional Neural Networks for Twitter Text Toxicity Analysis. In *INNS Big Data and Deep Learning conference* (pp. 370-379). Springer, Cham. doi: 10.1007/978-3-030-16841-4_38

Gervais, B. T. (2014). Following the news? Reception of uncivil partisan media and the use of incivility in political expression. *Political Communication*, *31*(4), 564-583. doi: 10.1080/10584609.2013.852640

Gervais, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics*, *12*(2), 167-185. doi: 10.1080/19331681.2014.997416

Glaeser, E. L., & Ward, B. A. (2006). Myths and realities of American political geography. *Journal of Economic Perspectives*, *20*(2), 119-144. doi: 10.1257/jep.20.2.119

Goel, S., Mason, W., & Watts, D. J. (2010). Real and perceived attitude agreement in social networks. *Journal of personality and social psychology*, *99*(4), 611. doi: 10.1037/a0020697

Gramlich, J. (2019, May 16). 10 facts about Americans and Facebook, *Pew Research Center*. Retrieved from https://www.pewresearch.org/fact-tank/2019/05/16/facts-about-americans-and-facebook

Green, D. P., Palmquist, B., & Schickler, E. (2002). *Partisan hearts and minds: Political parties and the social identities of voters*. New Haven: Yale University Press.

Green, J. C., Guth, J. L., & Smidt, C. E. (1996). *Religion and the culture wars: Dispatches from the front*. Rowman & Littlefield.

Greene, S. (1999). Understanding party identification: A social identity approach. *Political Psychology*, *20*(2), 393-403. doi: 10.1111/0162-895X.00150

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267-297. doi: 10.1093/pan/mps028

Gutmann, A., & Thompson, D. F. (2012). *The Spirit of Compromise: Why Governing Demands It and Campaigning Undermines It*. Princeton, NJ: Princeton University Press.

Haidt, J. and Hetherington, M. J. (2012, Sep 17). Look how far we've come apart. *The New York Times*. Retrieved from: https://campaignstops.blogs.nytimes.com/2012/09/17

Hamilton, J. (2004). *All the news that's fit to sell: How the market transforms information into news*. Princeton, NJ: Princeton University Press.

Hasell, A., & Weeks, B. E. (2016). Partisan provocation: The role of partisan news use and emotional responses in political information sharing in social media. *Human Communication Research*, *42*(4), 641-661. doi: 10.1111/hcre.12092

Haslett, J., & Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *38*(1), 1-21.

Heiss, R., Schmuck, D., & Matthes, J. (2019). What drives interaction in political actors' Facebook posts? Profile and content predictors of user engagement and political actors' reactions. *Information, Communication & Society*, *22*(10), 1497-1513. doi: 10.1080/1369118X.2018.1445273

Hester, J. B., & Gibson, R. (2003). The economy and second-level agenda setting: A time-series analysis of economic news and public opinion about the economy. *Journal of Mass Communication Quarterly*, *80*(1), 73-90. doi: 10.1177/107769900308000106

Hetherington, M. J. (2009). Putting polarization in perspective. *British Journal of Political Science*, *39*(2), 413-448. doi: 10.1017/S0007123408000501

Hetherington, M., & Weiler, J. (2018) *Prius or Pickup?: How the Answers to Four Simple Questions Explain America's Great Divide*. New York: Houghton Mifflin Harcourt.

Hibbing, J. R., Smith, K. B., & Alford, J. R. (2014). Differences in negativity bias underlie variations in political ideology. *Behavioral and brain sciences*, *37,* 297-307. doi: 10.1017/S0140525X13001192

Hill, S. J., & Tausanovitch, C. (2015). A disconnect in representation? Comparison of trends in congressional and public polarization. *The Journal of Politics*, *77*(4), 1058-1075. doi: 10.1086/682398

Hong, S., & Kim, S. H. (2016). Political polarization on twitter: Implications for the use of social media in digital governments. *Government Information Quarterly*, *33*(4), 777-782. doi: 10.1016/j.giq.2016.04.007

Hossain, N., Tran, T. T. T., & Kautz, H. (2018). Discovering Political Slang in Readers' Comments. In *Twelfth International AAAI Conference on Web and Social Media*.

Hu, M. C., Pavlicova, M., & Nunes, E. V. (2011). Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *The American journal of drug and alcohol abuse*, *37*(5), 367-375. doi: 10.3109/00952990.2011.597280

Huitt, W. (2003). The affective system: Educational psychology interactive. Valdosta, GA: Valdosta State University.

Hunter, J. D. (1994). *Before the Shooting Begins: Searching for Democracy in America's Culture War*. New York: Free Press.

Hutchens, M. J., Hmielowski, J. D., & Beam, M. A. (2019). Reinforcing spirals of political discussion and affective polarization. *Communication Monographs*, *86*(3), 357-376.

Hutchinson, A. (2020, April 18). Facebook Unveils New 'Care' Reactions to Help Express Responses to COVID-19. *Social Media Today*. Retrieved from https://www.socialmediatoday.com/news/facebook-unveils-new-care-reactions-to-help-express-responses-to-covid-19/576318/

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216-225. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/14550

Hwang, H., Kim, Y., & Huh, C. U. (2014). Seeing is believing: Effects of uncivil online debate on political polarization and expectations of deliberation. *Journal of Broadcasting & Electronic Media*, *58*(4), 621-633. doi: 10.1080/08838151.2014.966365

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). United States: OTexts. Retrieved from https://otexts.com/fpp2/

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of statistical software*, *27*, 1-22. doi :10.18637/jss.v027.i03

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of forecasting*, *18*(3), 439-454. doi: 10.1016/S0169-2070(01)00110-8

Ito, Tiffany A., Jeff T. Larsen, N. Kyle Smith, and John T. Cacioppo. 1998. “Negative Information Weighs More Heavily on the Brain: the Negativity Bias in Evaluative Categorizations.” *Journal of Personality and Social Psychology* 75 (4): 887–900. doi: 10.1037/0022-3514.75.4.887

Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication, 59*(1), 19-39. doi: 10.1111/j.1460-2466.2008.01402.x

Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, *59*(3), 690-707. doi: 10.1111/ajps.12152

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2018). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*. doi: 0.1146/annurev-polisci-051117-073034

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public opinion quarterly*, *76*(3), 405-431. doi:10.1093/poq/nfs038

Jacobson, G. C. (2000). Party polarization in national politics: The electoral connection. In *Polarized politics: Congress and the president in a partisan era* (Vol. 5, pp. 17-18).

Jarvis, S. E. (2004). Partisan patterns in presidential campaign speeches, 1948–2000. *Communication Quarterly*, *52*(4), 403-419. doi: 10.1080/01463370409370209

Jarvis, S. E., & Jennings, J. T. (2017). Republicans should vote: Partisan conceptions of electoral participation in campaign 2016. *American Behavioral Scientist*, *61*(6), 633-644. doi: 10.1177/0002764217720481

Jensen, J., Naidu, S., Kaplan, E., Wilse-Samson, L., Gergen, D., Zuckerman, M., & Spirling, A. (2012). Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech [with Comments and Discussion]. *Brookings Papers on Economic Activity*, 1-81. Retrieved from: http://www.jstor.org/stable/41825364

Jing-Schmidt, Z. (2007). Negativity bias in language: A cognitive-affective model of emotive intensifiers. *Cognitive Linguistics*, *18*(3), 417-443. doi:10.1515/COG.2007.023

Johnson, B. K., Neo, R. L., Heijnen, M. E., Smits, L., & van Veen, C. (2020). Issues, involvement, and influence: Effects of selective exposure and sharing on polarization and participation. *Computers in Human Behavior*, *104*, 106-155. doi: 10.1016/j.chb.2019.09.031.

Jones, D. R. (2001). Party polarization and legislative gridlock. *Political Research Quarterly, 54*(1), 125-141. doi :10.1177/106591290105400107

Jost, P., Maurer, M., & Hassler, J. (2020). Populism fuels love and anger: The impact of message features on users' reactions on Facebook. *International Journal of Communication*, *14*, 2081-2102. Retrieved from https://ijoc.org/index.php/ijoc/article/view/13400

Jurkowitz, M. & Mitchell, A. (2020, February 5). A sore subject: Almost half of Americans have stopped talking politics with someone, *Pew Research Center*. Retrieved from https://www.pewresearch.org/journalism/2020/02/05/a-sore-subject-almost-half-of-americans-have-stopped-talking-politics-with-someone/

Kahne, J., & Bowyer, B. (2018). The political significance of social media activity and social networks. *Political Communication*, *35*(3), 470-493. doi: 10.1080/10584609.2018.1426662

Kenski, K., Coe, K., & Rains, S. A. (2017). Perceptions of uncivil discourse online: An examination of types and predictors. *Communication Research*, doi: 10.1177/0093650217699933.

Kim, J. W. (2018). Online incivility in comment boards: Partisanship matters-But what I think matters more. *Computers in Human Behavior*, *85*, 405-412. doi: 10.1016/j.chb.2018.04.015

Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity. *Journal of Communication*, *71*(6), 922-946. doi :10.1093/joc/jqab034

Kim, Y., & Kim, Y. (2019). Incivility on Facebook and political polarization: The mediating role of seeking further comments and negative emotion. *Computers in Human Behavior*, *99*, 219-227. doi: 10.1016/j.chb.2019.05.022

Kingzette, J., Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2021). How affective polarization undermines support for democratic norms. *Public Opinion Quarterly*, *85*(2), 663-677. doi :10.1093/poq/nfab029

Knobloch-Westerwick, S., & Lavis, S. M. (2017). Selecting serious or satirical, supporting or stirring news? Selective exposure to partisan versus mockery news online videos. *Journal of Communication*, *67*(1), 54-81. doi :10.1111/jcom.12271

Kobayashi, T., Hoshino, T., & Suzuki, T. (2020). Inadvertent learning on a portal site: A longitudinal field experiment. *Communication Research*, *47*(5), 729-749. doi: 10.1177/0093650217732208

Krug, S. (2016, February 24). Reactions Now Available Globally. *Facebook Newsroom*. Retrieved from https://about.fb.com/news/2016/02/reactions-now-available-globally/

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1-3), 159-178. doi :10.1016/0304-4076(92)90104-Y

Larsson, A. O. (2018). Diversifying Likes: Relating reactions to commenting and sharing on newspaper Facebook pages. *Journalism Practice*, *12*(3), 326-343. doi: 10.1080/17512786.2017.1285244

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, *97*(2), 311-331. doi: 10.1017/S0003055403000698

Layman, G. C. (1999). “Culture Wars” in the American Party System: Religious and Cultural Change among Partisan Activists Since 1972. *American Politics Quarterly*, *27*(1), 89-121. doi: 10.1177/1532673X99027001006

Layman, G. C. (2001). *The great divide: Religious and cultural conflict in American party politics*. New York: Columbia University Press.

Layman, G. C., Carsey, T. M., & Horowitz, J. M. (2006). Party polarization in American politics: Characteristics, causes, and consequences. *Annual Review of Political Science, 9*, 83-110. doi: 10.1146/annurev.polisci.9.070204.105138

Lee, J. K., Choi, J., Kim, C., & Kim, Y. (2014). Social media, network heterogeneity, and opinion polarization. *Journal of Communication*, *64*(4), 702-722. doi :10.1111/jcom.12077

Lee, T. (2002). *Mobilizing public opinion: Black insurgency and racial attitudes in the civil rights era*. Chicago: University of Chicago Press.

Lelkes, Y. (2016). Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, *80*(S1), 392-410. doi: 10.1093/poq/nfw005

Lelkes, Y., & Westwood, S. J. (2017). The limits of partisan prejudice. *The Journal of Politics*, *79*(2), 485-501. doi: 10.1086/688223

Lelkes, Y., Sood, G., & Iyengar, S. (2017). The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science*, *61*(1), 5-20. doi: 10.1111/ajps.12237

Levendusky, M. (2009). *The partisan sort: How liberals became Democrats and conservatives became Republicans*. Chicago: The University of Chicago Press.

Levendusky, M. S. (2010). Clearer cues, more consistent voters: A benefit of elite polarization. *Political Behavior*, *32*(1), 111-131. doi: 10.1007/s11109-009-9094-0

Levendusky, M. S. (2013). Partisan media exposure and attitudes toward the opposition. *Political Communication*, *30*(4), 565-581. doi: 10.1080/10584609.2012.737435

Lexico. (n.d.). Libtard. in *lexico.com dictionary.* Retrieved from https://www.lexico.com/en/definition/libtard

Leyens, J. P., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups. *Personality and Social Psychology Review*, *4*(2), 186-197. doi: 10.1207/S15327957PSPR0402_06

Lu, Y. (2019). Incidental exposure to political disagreement on Facebook and corrective participation: Unraveling the effects of emotional responses and issue relevance. *International Journal of Communication*, *13*, 23. Retreived from https://ijoc.org/index.php/ijoc/article/view/9235

Lu, Y., & Myrick, J. G. (2016). Cross-cutting exposure on Facebook and political participation. *Journal of Media Psychology*, *28*(3), 100-110. doi: 10.1027/1864-1105/a000203

Luntz, F. I. (2004, November 9). Interview Frank Luntz. *Frontline*, Retrieved from http://www.pbs.org/wgbh/pages/frontline/shows/persuaders/interviews/luntz.html

Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology*, *57*(6), 981. doi: 10.1037/0022-3514.57.6.981

MacKuen, M., Wolak, J., Keele, L., & Marcus, G. E. (2010). Civic engagements: Resolute partisanship or reflective deliberation. *American Journal of Political Science*, *54*(2), 440-458. doi :10.1111/j.1540-5907.2010.00440.x

Marcus, G. E., Neuman, W. R., & MacKuen, M. (2000). *Affective intelligence and political judgment*. University of Chicago Press.

Martin, G. J., & Yurukoglu, A. (2017). Bias in cable news: Persuasion and polarization. *American Economic Review*, *107*(9), 2565-99. doi: 10.1257/aer.20160812

Marwick, A. E., & boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114-133. doi: 10.1177/1461444810365313

Mason, L. (2015). "I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, *59*(1), 128-145. doi: 10.1111/ajps.12089

Mason, L. (2016). A cross-cutting calm: How social sorting drives affective polarization. *Public Opinion Quarterly*, *80*(S1), 351-377. doi: 10.1093/poq/nfw001

Mason, L. (2018). Ideologues without Issues: The Polarizing Consequences of Ideological Identities. *Public Opinion Quarterly*, *82*(S1), 280-301. doi: 10.1093/poq/nfy005

Mason, L., & Wronski, J. (2018). One tribe to bind them all: How our social group attachments strengthen partisanship. *Political Psychology*, *39*, 257-277. doi: 10.1111/pops.12485

Massaro, T. M., & Stryker, R. (2012). Freedom of speech, liberal democracy, and emerging evidence on civility and effective democratic engagement. *Arizona Law Review*, *54*(2), 375-411.

Masullo Chen, G., & Lu, S. (2017). Online political discourse: Exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal of Broadcasting & Electronic Media*, *61*(1), 108-125. doi :10.1080/08838151.2016.1273922

Matsumoto, D., Hwang, H. C., & Frank, M. G. (2016). The effects of incidental anger, contempt, and disgust on hostile language and implicit behaviors. *Journal of Applied Social Psychology*, 46(8), 437-452. doi: 10.1111/jasp.12374

Matthes, J., Knoll, J., Valenzuela, S., Hopmann, D. N., & Von Sikorski, C. (2019). A meta-analysis of the effects of cross-cutting exposure on political participation. *Political Communication*, *36*(4), 523-542. doi: 10.1080/10584609.2019.1619638

McCarty, N., Poole, K. T., & Rosenthal, H. (2006). *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge: MIT Press.

Mcclain, C. (2021, May 4). 70% of U.S. social media users never or rarely post or share about political, social issues, *Pew Research Center*, Retrieved from https://www.pewresearch.org/fact-tank/2021/05/04/70-of-u-s-social-media-users-never-or-rarely-post-or-share-about-political-social-issues

Mcclain, C., Widjaya, R., Rivero, G., & Smith, A. (2021, November 15). The Behaviors and Attitudes of U.S. Adults on Twitter, *Pew Research Center*, Retrieved from https://www.pewresearch.org/internet/2021/11/15/the-behaviors-and-attitudes-of-u-s-adults-on-twitter

McLaughlin, C., & Vitak, J. (2012). Norm evolution and violation on Facebook. *New media & society*, *14*(2), 299-315. doi :10.1177/1461444811412712

Merrill, J. & Oremus, W. (2021, October 26). Five points for anger, one for a ‘like’: How Facebook’s formula fostered rage and misinformation, *The Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/

Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, *41*(8), 1042-1063. doi: 10.1177/0093650212466406

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119). Retrieved from https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html

Mohammad, S. M. (2016). Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text. In *Emotion Measurement* (pp. 201-237). Woodhead Publishing. doi: 10.1016/B978-0-08-100508-8.00009-6

Monroe, B. L., Colaresi, M. P., & Quinn, K. M. (2008). Fightin’words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, *16*(4), 372-403. doi: 10.1093/pan/mpn018

Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: time series missing value imputation in R. *R J.*, *9*(1), 207.

Morris, J. S. (2007). Slanted objectivity? Perceived media bias, cable news exposure, and political attitudes. *Social Science Quarterly*, *88*(3), 707-728. doi: 10.1111/j.1540-6237.2007.00479.x

Muddiman, A., & Stroud, N. J. (2017). News values, cognitive biases, and partisan incivility in comment sections. *Journal of Communication*, *67*(4), 586-609. doi: 10.1111/jcom.12312

Muddiman, A., Budak, C., Romas, B., Kim, Y., Murray, C., Burniston, M. M., Geiger, J., Purcell, A., Ludzenski, J., Turner, M., Duchovany, M., & Stroud, N. J. (December, 2020). Cable and nightly network news coverage of coronavirus. Center for Media Engagement. Retrieved from https://mediaengagement.org/research/coronavirus-network-coverage

Muddiman, A., McGregor, S. C., & Stroud, N. J. (2019). (Re) claiming our expertise: Parsing large text corpora with manually validated and organic dictionaries. *Political Communication*, *36*(2), 214-226. doi: 10.1080/10584609.2018.1517843

Muddiman, A., Pond-Cobb, J., & Matson, J. E. (2020). Negativity bias or backlash: Interaction with civil and uncivil online political news content. *Communication Research*, *47*(6), 815-837. doi: 10.1177/0093650216685625

Murthy, D. (2015). Twitter and elections: are tweets, predictive, reactive, or a form of buzz? *Information, Communication & Society*, *18*(7), 816-831. doi: 10.1080/1369118X.2015.1006659

Mutz, D. C. (2006a). Hearing the other side: Deliberative versus participatory democracy. New York: Cambridge University Press.

Mutz, D. C. (2006b). How the mass media divide us. In P. S. Nivola & D. W. Brady (Eds.) *Red and blue nation? Characteristics and causes of America's polarized politics* (Vol. 1, pp. 223-242). Washington: Brookings Institution Press.

Mutz, D. C. (2015). *In-your-face politics: The consequences of uncivil media*. Princeton University Press.

Nai, A., & Maier, J. (2021). Is negative campaigning a matter of taste? Political attacks, incivility, and the moderating role of individual differences. *American Politics Research*, *49*(3), 269-281. doi: 10.1177/1532673X20965548

Obadimu, A., Mead, E., Hussain, M. N., & Agarwal, N. (2019). Identifying Toxicity Within YouTube Video Comment. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation* (pp. 214-223). Springer, Cham.

OECD (2021), Consumer confidence index (CCI) (indicator). doi: 10.1787/46434d78-en

Oeldorf-Hirsch, A. (2018). The role of engagement in learning from active and incidental news exposure on social media. *Mass Communication and Society*, *21*(2), 225-247. doi: 10.1080/15205436.2017.1384022

Orwell, G. (1946). *Politics and the English language*. London: Horizon.

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, *6*(2), 259-283. doi: 10.1177/1461444804041444

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. London: Penguin UK.

Pasek, J., Yan, H. Y., Conrad, F. G., Newport, F., & Marken, S. (2018). The stability of economic correlations over time: identifying conditions under which survey tracking polls and Twitter sentiment yield similar conclusions. *Public Opinion Quarterly*, *82*(3), 470-492. doi: 10.1093/poq/nfy030

Patterson, T. E. (1993). *Out of order*. New York: Vintage Press.

Pedersen, R. T., Anspach, N. M., Hansen, K. M., & Arceneaux, K. (2021). Political predispositions, not popularity: people's propensity to interact with political content on Facebook. *Journal of Elections, Public Opinion and Parties*, 1-17. doi: 10.1080/17457289.2021.1952209

Pennacchiotti, M., & Popescu, A. M. (2011, July). A machine learning approach to twitter user classification. In *Proceedings of the international AAAI conference on web and social media* (Vol. 5, No. 1). Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/14139

Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC 2015.* Austin, TX: University of Texas at Austin.

Perrin, A. (2020, October 15). 23% of users in U.S. say social media led them to change views on an issue; some cite Black Lives Matter, *Pew Research Center*. Retrieved from https://www.pewresearch.org/fact-tank/2020/10/15/23-of-users-in-us-say-social-media-led-them-to-change-views-on-issue-some-cite-black-lives-matter

Pew Research Center (2018), July 30 – August 12, 2018 – American Trends Panel Wave 37 [Data file]. Retrieved https://www.pewresearch.org/journalism/dataset/american-trends-panel-wave-37

Pew Research Center (2019). January 8 – February 7, 2019 – Core Trends Survey [Data file]. Retrieved from https://www.pewresearch.org/internet/dataset/core-trends-survey

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*(2), 335-346. doi :10.1093/biomet/75.2.335

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. New York: Cambridge University Press.

Rains, S. A., Kenski, K., Coe, K., & Harwood, J. (2017). Incivility and political identity on the internet: Intergroup factors as predictors of incivility in discussions of news online. *Journal of Computer-Mediated Communication*, 22(4), 163-178. doi: 10.1111/jcc4.12191

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, *118*(26). doi: 10.1073/pnas.2024292118

Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., & Benevenuto, F. (2016). Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, *5*(1), 1-29. doi: 10.1140/epjds/s13688-016-0085-1

Rizopoulos, D. (2021). GLMMadaptive: generalized linear mixed models using adaptive Gaussian quadrature. *R package version 0.8–2*.

Salgado, S., & Bobba, G. (2019). News on events and social media: A comparative analysis of Facebook users' reactions. *Journalism Studies*, *20*(15), 2258-2276. doi: 10.1080/1461670X.2019.1586566

Schmidt, A. L., Zollo, F., Del Vicario, M., Bessi, A., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2017). Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences, 114*(12), 3035-3039. doi: 10.1073/pnas.1617052114

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M.E., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one, 8*(9), e73791. doi: 10.1371/journal.pone.0073791

Settle, J. E. (2018). *Frenemies: How social media polarizes America*. New York: Cambridge University Press.

Shearer, E. (2021, January 12). More than eight-in-ten Americans get news from digital devices, *Pew Research Center*. Retrieved from https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices

Shearer, E. & Mitchell, A. (2021, January 12). News Use Across Social Media Platforms in 2020, *Pew Research Center,* Retrieved from https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021

Shearer, E., & Matsa, K. E. (2018, September 10). News Use Across Social Media Platforms 2018, *Pew Research Center,* Retrieved from https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018

Shin, J., & Thorson, K. (2017). Partisan selective sharing: The biased diffusion of fact-checking messages on social media. *Journal of Communication, 67*(2), 233-255. doi: 10.1111/jcom.12284

Shugars, S., & Beauchamp, N. (2019). Why keep arguing? Predicting engagement in political conversations online. *Sage Open*, *9*(1), doi: 10.1177/2158244019828850

Sobieraj, S., & Berry, J. M. (2011). From incivility to outrage: Political discourse in blogs, talk radio, and cable news. *Political Communication*, *28*(1), 19-41. doi: 10.1080/10584609.2010.542360

Soroka, S. N. (2014). *Negativity in democratic politics: Causes and consequences*. Cambridge, UK: Cambridge University Press.

Soroka, S., Daku, M., Hiaeshutter-Rice, D., Guggenheim, L., & Pasek, J. (2018). Negativity and positivity biases in economic news coverage: Traditional versus social media. *Communication Research*, *45*(7), 1078-1098. doi: 10.1177/0093650217725870

Soroka, S., Fournier, P., & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences*, *116*(38), 18888-18892. doi: 10.1073/pnas.1908369116

Soroka, S., Young, L., & Balmas, M. (2015). Bad news or mad news? Sentiment scoring of negativity, fear, and anger in news content. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 108-121. doi: 10.1177/0002716215569217

Spottswood, E., & Wohn, D. Y. (2019). Beyond the "like": How people respond to negative posts on Facebook. *Journal of broadcasting & electronic media*, *63*(2), 250-267. doi :10.1080/08838151.2019.1622936

Stonecash, J. M., Brewer, M. D., & Mariani, M. D. (2003). *Diverging Parties: Social Change, Realignment, and Party Polarization*. Boulder: Westview Press.

Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior*, *30*(3), 341-366. doi: 10.1007/s11109-007-9050-9

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, *60*(3), 556-576. doi: 10.1111/j.1460-2466.2010.01497.x

Stroud, N. J. (2011). *Niche news: The politics of news choice*. New York: Oxford University Press.

Stroud, N. J., & Muddiman, A. (2013). Selective exposure, tolerance, and satirical news. *International Journal of Public Opinion Research*, *25*(3), 271-290. doi: 10.1093/ijpor/edt013

Stroud, N. J., & Muddiman, A. (2019). Social media engagement with strategy-and issue-framed political news. *Journal of Communication*, *69*(5), 443-466. doi :10.1093/joc/jqz029

Stroud, N. J., Muddiman, A., & Scacco, J. M. (2017). Like, recommend, or respect? Altering political behavior in news comment sections. *New media & society*, *19*(11), 1727-1743. doi: 10.1177/1461444816642420

Sturm Wilkerson, H., Riedl, M. J., & Whipple, K. N. (2021). Affective affordances: Exploring Facebook reactions as emotional responses to hyperpartisan political news. *Digital Journalism*, *9*(8), 1040-1061. doi :10.1080/21670811.2021.1899011

Su, L. Y. F., Xenos, M. A., Rose, K. M., Wirz, C., Scheufele, D. A., & Brossard, D. (2018). Uncivil and personal? Comparing patterns of incivility in comments on the Facebook pages of news outlets. *New Media & Society*, *20*(10), 3678-3699. doi: 10.1177/1461444818757205

Suhay, E., Bello-Pardo, E., & Maurer, B. (2018). The polarizing effects of online partisan criticism: Evidence from two experiments. *The International Journal of Press/Politics*, *23*(1), 95-115. doi: 10.1177/1940161217740697

Sumner, E. M., Ruge-Jones, L., & Alcorn, D. (2018). A functional approach to the Facebook Like button: An exploration of meaning, interpersonal functionality, and potential alternative response buttons. *New Media & Society*, *20*(4), 1451-1469. doi: 10.1177/1461444817697917

Sunstein, C. R. (2007). *Republic.com 2.0.* Princeton, NJ: Princeton University Press.

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

Sydnor, E. (2018). Platforms for incivility: Examining perceptions across different media formats. *Political Communication*, *35*(1), 97-116. doi: 10.1080/10584609.2017.1355857

Tajfel, H. (1978). *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations*, New York/London: Academic Press.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Political of intergroup relations* (pp. 7-24). Chicago: Nelson-Hall.

Tetlock, P. E. (1983). Cognitive style and political ideology. *Journal of Personality and social Psychology*, *45*(1), 118. doi: 10.1037/0022-3514.45.1.118

Thelwall, M. (2017). The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In J. Holyst (Eds.), *Cyberemotions* (pp. 119-134). Springer, Cham. doi: 10.1007/978-3-319-43639-5_7

Thorson, K., & Wells, C. (2016). Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*, *26*(3), 309-328. doi: 10.1111/comt.12087

Trester, A. M., & West, L. (2013). Facework on Facebook: Conversations on social media. In D. Tannen and A. M. Trester (Eds.), *Discourse 2.0. Language and new media.* (pp. 133–154). Washington, DC: Georgetown University Press.

Trussler, M., & Soroka, S. (2014). Consumer demand for cynical and negative news frames. *The International Journal of Press/Politics*, *19*(3), 360-379. doi: 10.1177/1940161214524832

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Election Forecasts with Twitter: How 140 characters reflect the political landscape. *Social Science Computer Review*, *29*(4), 402-418. doi: 10.1177/0894439310386557

Tynan, D. (2018, October 11). Facebook removes hundreds of US political pages for 'inauthentic activity', *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018 /oct/11/facebook-purge-page-removal-spam

Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J. T., Nagler, J., & Tucker, J. (2013). Social media and political communication. A survey of Twitter users during the 2013 Italian general election. *Rivista Italiana di Scienza Politica*, *43*(3), 381-410. doi: 10.1426/75245

Vaes, J., Paladino, M. P., & Leyens, J. P. (2002). The lost e-mail: Prosocial reactions induced by uniquely human emotions. *British Journal of Social Psychology*, *41*(4), 521-534. doi: 10.1348/014466602321149867

Valenzuela, S., Park, N., & Kee, K. F. (2009). Is there social capital in a social network site? Facebook use and college students' life satisfaction, trust, and participation. *Journal of*

Computer-Mediated Communication, *14*(4), 875-901. doi: 10.1111/j.1083-6101.2009.01474.x

Van Dijk, T. A. (1987). *Communicating racism: Ethnic prejudice in thought and talk*. Sage Publications, Inc.

Wagner, K. (2021, July 14). Twitter Penalizes Record Number of Accounts for Posting Hate Speech, *Time*, Retrieved from https://time.com/6080324/twitter-hate-speech-penalties

Wang, M. Y., & Silva, D. E. (2018). A slap or a jab: An experiment on viewing uncivil political discussions on Facebook. *Computers in Human Behavior*, *81*, 73-83. doi: 10.1016/j.chb.2017.11.041

Webster, S. W., & Abramowitz, A. I. (2017). The ideological foundations of affective polarization in the US electorate. *American Politics Research*, *45*(4), 621-647. doi:10.1177/1532673X17703132

Weeks, B. E., Ardèvol-Abreu, A., & Gil de Zúñiga, H. (2017). Online influence? Social media use, opinion leadership, and political persuasion. *International Journal of Public Opinion Research*, *29*(2), 214-239. doi: 10.1093/ijpor/edv050

Weeks, B. E., Lane, D. S., Kim, D. H., Lee, S. S., & Kwak, N. (2017). Incidental Exposure, Selective Exposure, and Political Information Sharing: Integrating Online Exposure Patterns and Expression on Social Media. *Journal of Computer-Mediated Communication, 22*(6), 363-379. doi: 10.1111/jcc4.12199

Wells, C., & Thorson, K. (2017). Combining big data and survey techniques to model effects of political content flows in Facebook. *Social Science Computer Review*, *35*(1), 33-52. doi: 10.1177/0894439315609528

Winkler, H. (2019). The effect of income inequality on political polarization: Evidence from European regions, 2002–2014. *Economics & Politics*, *31*(2), 137-162. doi: 10.1111/ecpo.12129

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67-82. doi: 10.1109/4235.585893

Young, L., & Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. *Political Communication*, *29*(2), 205-231. doi: 10.1080/10584609.2012.671234

Yu, B., Kaufmann, S., & Diermeier, D. (2008). Classifying party affiliation from political speech. *Journal of Information Technology & Politics*, *5*(1), 33-48. doi: 10.1080/19331680802149608

Zaller, J. (1992). *The nature and origins of mass opinion*. New York: Cambridge University Press.

Zhang, X., & Yi, N. (2020). NBZIMM: negative binomial and zero-inflated mixed models, with application to microbiome/metagenomics data analysis. *BMC bioinformatics*, *21*(1), 1-19. doi :10.1186/s12859-020-03803-z

Zukin, C., Keeter, S., Andolina, M., Jenkins, K., & Carpini, M. X. D. (2006). *A new engagement? Political participation, civic life, and the changing American citizen*. New York: Oxford University Press.