Computer Supported Cooperative Work (CSCW) (2019) 28:317–355
DOI 10.1007/s10606-019-09357-w

Moral and Affective Differences in U.S. Immigration Policy Debate on Twitter

Ted Grover[1]*, Elvan Bayraktaroglu[2], Gloria Mark[1] & Eugenia Ha Rim Rho[1]
*[1]*University of California, Irvine, CA , USA (E-mail: grovere@uci.edu; E-mail: gmark@uci.edu; E-mail: hrrho@uci.edu) ;* [2]*Istanbul Technical University, Istanbul , Turkey (E-mail: ayseelvan@gmail.com)*

Abstract. Understanding ideological conflict has been a topic of interest in CSCW, for example in Value Sensitive Design research. More specifically, understanding ideological conflict is important for studying social media platforms like Twitter, which provide the ability for people to freely express their thoughts and opinions on contentious political events. In this work, we examine Twitter data to understand the moral, affective, and cognitive differences in language use between two opposing sides of the political debate over immigration related issues in the United States in the year since the 2016 presidential election. In total, we analyzed and compared the language of 45,045 pro-immigration tweets and 11,213 anti-immigration tweets spread across this period. Based on Moral Foundations Theory used to understand ideological conflict, we found pro-immigration tweets to contain more language associated with moral foundations of harm, fairness, and loyalty. Anti-immigration tweets contained more language associated with moral foundations of authority, more words associated with cognitive rigidity and more 3rd person pronouns and negative emotion. We discuss the implications of our research for political communication over social media, and for incorporating Moral Foundations Theory into other CSCW research. We discuss the potential application of this theory for Value Sensitive Design research.

Key Words: Twitter, morality, immigration, text analysis, polarization, value sensitive design

1. Introduction

We are living in a digital age where people have a means to express and spread opinions rapidly in the public sphere of social media. At the same time, people worldwide are living in challenging political times. Increasingly more, social media provides a channel for (sometimes vociferous) responses to political events in near real-time. As we have seen in multiple studies of political views in the U.S. (Conover et al. 2011; McCright and Dunlap 2011), these voices also tend to be very partisan. Partisan views are often expressed on social media with moral certainty, but what underlies these moral judgments? What affective and cognitive differences in linguistic style might be associated with the rhetoric of these judgments? How might these linguistic differences explain steadfast partisan views?

In this paper, we use Twitter as a lens to examine differences in political rhetoric. Social media platforms can serve as a public petri dish for the opinions, thoughts and activities of individuals and political groups. The affordance of Twitter, which enables users to express and propagate opinions immediately, aligns well with a focus on studying the patterns of moral rhetoric across partisan views: instinctive, intuitive reactions to political messaging and events are believed to be influenced by foundational moral beliefs (Koleva et al. 2012).

Twitter has been of interest in the CSCW community, being used to observe a variety of behavioral and emotional phenomena, such as affective desensitization (De Choudhury et al. 2014), communication during natural disasters (Vieweg et al. 2010), behavioral indicators of major life changes (De Choudhury et al. 2013). how rumors are spread (Maddock et al. 2018), political conflict (McCright and Dunlap 2011), health status updates (Park et al. 2016), and tracking happiness in communities (Quercia et al. 2012).

In the time since the 2016 U.S. presidential election, several issues have drawn a lot of partisan attention on social media. In particular, the topic of immigration has led to a voluminous discussion on Twitter and other social media where both liberals (i.e. those that primarily lean towards the left side of the political spectrum) and conservatives (i.e. those that primarily lean towards the right side of the political spectrum) express moral views on their positions of pro or anti-immigration. These debates are evident worldwide. There is, however, growing evidence that liberals and conservatives show preference for different moral foundations to inform their political views (Graham et al. 2009; Graham et al. 2012; Haidt 2012). Moral Foundations Theory (**MFT**) (Haidt and Graham 2007), which we draw on in this study, has gained recognition as a theory that can be used to understand why people adopt judgments about issues.

The goal of our study is to investigate, using computerized text methods applied to Twitter, and drawing on the case of immigration, how the rhetoric of opposing political stances employ different moral, cognitive and affective differences in linguistic style. While immigration debates on Twitter are found worldwide, we focus our analysis on debates in the U.S. We examine Tweets which reflect people's views on immigration as the events were unfolding for the year-long period since the 2016 U.S. election, as well as for two key periods where significant immigration policy was proposed: the travel ban for citizens of some Muslim countries and the DACA repeal, which is the cancellation of an act granting legal status to young immigrants brought to the U.S. as a child. Past work has investigated the relationship between psychological theories like MFT and the social divide on many policy issues including immigration (Koleva et al. 2012; Kaur and Sasahara 2016), as well as psycholinguistic differences between opposing sides of contentious social issues like abortion on Twitter (Sharma et al. 2017). However, to the best of our knowledge, our work provides the first analysis of the moral, cognitive, and affective differences in linguistic style for both sides of the ongoing debate over immigration policy in the United States. Our paper contributes to a further understanding of political rhetorical strategies surrounding immigration policy at a time when the partisan divide over this topic is growing globally (Abrajano and Hajnal 2017). However, we feel that not only can MFT lead to a deeper understanding of how large scale ideological conflict may arise and grow in online social networks, we also feel that there are opportunities to integrate MFT into Value Sensitive Design (VSD) research.

2. Related work

2.1. Moral psychology and moral foundations theory

The study of the psychology of morality has gone through a renaissance in the twenty-first century, as psychologists developed interest in how morality functions psychologically (i.e., descriptively) rather than to make normative claims about morality or how to teach people to achieve some desired stage of moral development. Modern research within moral psychology began to integrate knowledge from across social psychology, neuroscience, experimental philosophy, developmental science, evolutionary biology, and anthropology. Out of this cross-discipline integration, certain theories of moral psychology came to prominence, one of the most popular being Moral Foundations Theory (**MFT**).

MFT (Haidt and Graham 2007) claims that all cultures build morality perspectives mainly on five universally available psychological systems or foundations (Koleva et al. 2012). According to the theory, these foundations drive instinctive, immediate reactions to stimuli, which in turn lead to judgments of right or wrong (Koleva et al. 2012). Each of the five moral foundations is theorized in terms of a polarity: Care/Harm, which includes showing compassion and caring for the vulnerable, and ones in need, and avoiding causing harm; Fairness/Cheating, which includes concerns about injustice, inequality and selfishness since these create distrust among group members; Loyalty/Betrayal, which focuses responsibilities such as loyalty, patriotism, and self-sacrifice on one's in-group; Authority/Subversion, which incorporates respect and obedience to legitimate authorities, or refers to traditions, values, and institutions used by authorities to maintain stability, and fulfilling obligations in the hierarchy of society; and Sanctity/Degradation, which includes concerns about purity and sacredness, and avoidance of contamination, and defines some things as "untouchable" due to being holy or sacred, or the contrary, because it is dirty and disgusting (Graham et al. 2009; Graham et al. 2012; Haidt 2012).

Care/Harm and Fairness/Cheating have been referred to in the literature as the *individualizing* foundations, for their focus on the individual as the center of moral value (Graham et al. 2009). In contrast, Loyalty/Betrayal, Authority/Subversion, and Sanctity/Degradation have been referred to in the literature as the *binding* foundations, for their focus on the group as the center of moral value (Graham et al. 2009). Even though these foundations are universally available, every human group constructs its morality by prioritizing each of these foundations on varying levels, which in turn shows up as a divergence in cultural and social groups' moral intuitions and reasoning (Graham and Haidt 2012).

People who are intolerant to change, and who dislike uncertainty and ambiguity, have been shown to rely on the authority foundation (Koleva et al. 2012). When people are intolerant to change, they tend to respect authority more because changes in societal structure and norms are less likely to occur, thus preserving hierarchy (and certainty).

Other competing modern models exist on how moral judgements have evolved, including models that focus on how norm violations, negative affect, and perceived harm interact to explain moral judgements (Schein and Gray 2018), as well as theories that concentrate on how moral judgements evolved through universal efforts aimed at regulating and maintaining relationships (Rai and Fiske 2011). However, MFT has emerged as a powerful and popular theory through its utility in providing a common language for talking about the moral domain, which makes it ideal for interdisciplinary research purposes. However, a part of what makes MFT such a valuable theory for interdisciplinary study of morality compared to other theories is the set of methods that accompany the theory, including a Moral Foundations Questionnaire (MFQ) designed to obtain the degree to which individuals value each of the five foundations in the moral domain (Graham et al. 2011), and the Moral Foundations Dictionary (MFD), that measures the proportion of words in a sample of text associated with each foundation. In our work, we use the MFD to analyze text data from immigration related tweets from Twitter.

In doing so, we align our work with the long history of research that has used analysis of language to better understand human behavior. As Boyd and Pennebaker (2017, p.64) point out, there are 'countless patterns of attention, behaviors, and emotions deeply embedded in people's language' (p.64). Throughout our work we view moral judgements as a situated quality of language, and assume that in addition to emotion, the language we produce also contains an inherent moral valence that can be detected and analyzed through text analysis methods.

2.2. Liberal and conservative moral judgements

Empirical studies of MFT suggest that liberals and conservatives prioritize different sets of moral foundations when making moral judgements and interpreting arguments and information. While liberals draw more heavily upon the more *individualizing* foundations (Care/Harm and Fairness/Cheating), conservatives draw relatively more upon the *binding* foundations (Loyalty/Betrayal, Authority/Subversion, and Sanctity/Degradation) (Graham et al. 2009; Graham et al. 2012; Haidt 2012). Koleva et al. (2012) found that moral foundations are good predictors of opinions on controversial issues such as immigration, abortion, and euthanasia than demographics, ideology, religious attendance, political interest, and socio-political attitudes. The evidence suggests that moral foundations provide strong psychological tendencies, which can make individuals more prone to adopt certain socio-political beliefs. In fact, moral framing in line with one's political orientation can amplify one's attitude towards that issue (Day et al. 2014).

The political landscape in the U.S. historically has been characterized as one that is defined by two opposing political parties: Democrats (liberal ideology) and Republicans (conservative ideology). However, the ideological makeup of each opposing side is very diverse and multifaceted. For instance, on the liberal side, there are individuals who identify as socialists, communists, or anarchists. On the conservative side, there are some that identify strongly as social conservatives, and others that identify as libertarian capitalists, among others. Despite each side having a heterogeneous composition, there are still key ideological concepts that both liberal and conservative sides as a whole separately adhere to, that often cause ideological conflict from the opposing side (Koleva et al. 2012). We emphasize that when we use the terms 'liberal' and 'conservative' in this paper, we refer to these broad categories of political affiliation, to be consistent with past research on MFT and political ideology, where the U.S. political system has been a common area of focus (Graham et al. 2009; Haidt 2012; Koleva et al. 2012).

Since its inception, MFT has been used in the U.S. and various countries to study the moral differences between groups positioned along the political spectrum, e.g. the role morality plays in the choice of political ideology (Haidt and Graham 2007), or how conservatives and liberals differ in their endorsement of each moral foundation (Graham et al. 2009). Other studies have focused on how liberals and conservatives perceive the extent of their ingroup and outgroup moral concerns (Graham et al. 2012), which are reflected in public discourse topics, like stem cell research (Clifford and Jerit 2013), climate change (Dawson and Tyson 2012; Vainio and Mäkiniemi 2016) and ethical/moral food choices (Mäkiniemi et al. 2013).

2.3. Twitter as a medium to observe ideological conflict

The high ecological validity Twitter data offers by being produced during real-life events in real-time (Hoover et al. 2018; Ji and Raney 2015) makes Twitter a convenient medium to track the moral nature of the rhetoric of groups on particular issues. Sagi and Dehghani (2014) explored the moral stance of twitter users towards the U.S. Federal Shutdown in 2013, and found that liberal tweets endorsed more moral concerns regarding sanctity and harm, and conservative tweets endorsed more moral concerns regarding fairness, authority and loyalty. Hoover et al. (2018) observed that MFT harm and loyalty foundations are related to the intention to donate after Hurricane Sandy. In the context of entertainment media, the morality expressed in tweets were consistent with the morality of characters to which viewers felt emotionally attached (Ji and Raney 2015).

Twitter has also been used to observe political participation. In a Twitter analysis of citizen engagement and partisanship during U.S. and French presidential elections in 2012, less partisan polarization was observed for the U.S. (Hanna et al. 2013). Hemphill and Roback (2014) found 16 distinct categories of tweets that citizens use to influence politics. Twitter has also been used to detect public opinion, which revealed that political affiliation of the voter, personality of the candidate, and policy issues determine the political discourse (Le et al. 2017). In addition, Twitter has been used previously as a data source to model ongoing culture wars across U.S. states over contentious topics like abortion and same-sex marriage, and to use this model to predict the outcome of policy changes (Zhang and Counts 2015, 2016).

As mentioned earlier, moral judgments tend to be instinctive and immediate reactions to a stimulus. The affordances of Twitter enable people to broadcast their opinions and propagate the opinions of others they agree with in response to events immediately as they occur. As Tweets are purposely restricted in length, this affordance encourages people to get straight to the point in their communication. Thus, Twitter can serve as a laboratory in which to understand moral rhetoric of divergent views of "hot button" societal issues.

2.4. Immigration as a moral subject

Immigration, with the power to change demographics in any developed country (Hainmueller and Hopkins 2014) has been one of the controversial subjects in most Western countries. Koleva et al. (2012), counts it as one of the constant issues between conservatives and liberals that lead to culture wars. Even though immigration may have the potential to impact countries' politics, how it will get managed depends largely on that country's natural born citizens' (ie. non-immigrant citizens) attitudes towards immigration (Hainmueller and Hopkins 2014). Citizens' opinions toward immigration can be affected by economic and/or sociopsychological reasons (O'Rourke and Sinnott 2006). While arguments focusing on economic reasons ties it to natural born self-interest regarding the competition over limited resources between immigrants and natural born citizens, other arguments tie immigration attitudes to differences based on religion, national identity, language, prejudice etc. (Hainmueller and Hopkins 2014; Abrajano and Hajnal 2017). However, moral intuition, as acquired from innate psychological mechanisms (Graham et al. 2009) and having the capacity to influence people in socio-political opinions (Koleva et al. 2012) can be the underlying factor, and can have a decisive role on one's stance towards immigration.

MFT has as well been adopted to understand the public moral stance towards immigration utilizing Twitter data. Studying Tweets generated during the Paris attacks in 2015, it was discovered that political leanings, which are related to those foundations people draw upon more in their moral judgement, determine the perceived risk of an immigrant group (Chung et al. 2016). Kaur and Sasahara (2016) found that loyalty and sanctity are the most influential moral foundations related to several controversial topics including immigration in Twitter conversations. These results of political ideology, MFT and attitude towards immigration, were congruent with results using other sources of data (Koleva et al. 2012; Van de Vyver et al. 2016). Altogether, these results suggest that Twitter can be a medium to observe morality in judgements of public issues.

2.5. Ideological conflict in CSCW and value sensitive design research

Returning to CSCW research, understanding and managing ideological conflict has been investigated in studies of collaboration in online spaces. Researchers have previously studied the effects of ideological conflict within online platforms focusing on how users frame their communication within their social network, as well as how they might facilitate collaboration. Kou et al. (2017), investigated how the rhetorical framing of the Hong Kong Umbrella Revolution was strikingly different on Facebook (primarily used by Hong Kong residents) compared to Weibo (primarily used by residents of mainland China) due to socio-political-cultural differences between the two regions. Pal et al. (2018) showed how populist leaders on Twitter like Donald Trump, Nigel Farage, Geert Wilders, and Narenda Modi benefit from using antagonistic rhetoric appealing to in-group homogenization and disparaging the out-group, within the specific cultural and political landscape each leader resides in. Other work from Boulus-Rødje et al. (2015) described the challenges Israeli and Palestinian software developers faced in collaboration, as politics were extremely difficult to keep outside of collaborative practice. In previous studies examining ideological conflict in these contexts, there has been a strong emphasis on examining how socio-cultural-political contexts contribute to differences in rhetorical framing on online platforms. However, there has been little work on examining how rhetorical differences in moral framing of issues between groups in online platforms may be an important covariate in ideological polarization and conflict.

Ideological conflict has also been studied in Value Sensitive Design (VSD). For instance, Abokhodair and Vieweg (2016), showed how religious and cultural differences in the Arab Gulf region in comparison to western culture lead to different normative conceptions about privacy in technologically mediated environments. VSD historically describes a three-part methodological framework that includes conceptual, empirical, and technical investigations that aim to investigate the interactions between human values, system users, and system non-users (Le Dantec et al. 2009). Typically, the investigation of values has focused on the conceptual phase, where purposefully broad and expansive prescribed 'values' such as privacy, universal usability, and accountability are implicated and discussed in the context of design (Friedman et al. 2013). However, this approach has come under criticism (Borning and Muller 2012; Le Le Dantec et al. 2009), as scholars have noted how such predefined lists of values cultivate "a dogmatic response with respect to which values are worthy of consideration" and disengage from a "commitment to understanding the nuanced manifestation of a plurality of values." (Le Le Dantec et al. 2009, p.1142). As a result, scholars have called for more of an emphasis on the empirical investigation of values, moving beyond a rigid commitment to a set of universal values (Borning and Muller 2012; Le Le Dantec et al. 2009, Manders-Huits 2011). To the best of our knowledge, theories like MFT have rarely if ever been integrated into the empirical aspects of VSD research. We see MFT as a potentially powerful theoretical construct to better understand how values that conflict between groups may play out and mediate itself in online spaces, especially when those values are made less explicit and not dictated by clear geographical and cultural boundaries. Our intent is to bring MFT into a broader discussion of how knowledge from moral psychology can be integrated into future research for understanding conflict and collaboration in online environments.

3. Background and hypotheses

3.1. Background: U.S. immigration policies since the 2016 election

Since President Trump has taken office, his administration's immigration policy actions about restricting immigration to the U.S. have generated an ongoing controversy between groups with different perspectives. During the presidential campaign, the proposal of building a wall on the Mexican border to block unauthorized border crossings from Mexico was a main topic. In January 27, 2017 an executive order titled "Protecting the Nation from Terrorist Attacks by Foreign Nationals" was signed to suspend seven Muslim countries' citizens' entry to the U.S. for the next 90 days, which also blocked the admission of Syrian refugees to the country. The aim of the executive order has been explained to protect the country from the admission of radical Islamist terrorists (Vergani 2018). The executive order was met with large-scale protest and was blocked by the courts in February. However, since then, the government has signed revised versions of the executive order (March 6 and September 24), and the latest one has been allowed by the U.S. Supreme Court to go into effect pending appeal.

On September 5, 2017, the Trump administration announced that the Deferred Action for Childhood Arrivals (DACA) program, which is affecting 800,000 young adults, will be cancelled (Venkataramani and Tsai 2017). The program was signed in 2012 during the Obama administration and granted immigrants, who are called 'Dreamers' and were brought to the country illegally as children, a legal status to remain and to work in the U.S. Again, the announcement was met with protests, both in demonstrations and on social media.

3.2. Hypotheses

Based on the literature on MFT, combined with the described affordances of Twitter, we developed the following hypotheses that we test in this study. We emphasize again that when we refer to 'liberal' and 'conservative' we are referring to the broad heterogeneous categorization of political ideologies that characterize the most macro-level divisions in the U.S. political landscape.

> *H1: During the period of interest, we expect that tweets that are explicitly pro-immigration should contain more language associated with the Care/Harm and Fairness/Cheating foundations, compared to anti-immigration tweets.*

Based on the findings of the literature on MFT, and the literature on liberal and conservative attitudes toward immigration, the moral foundations categories that resonate most with liberals are the *individualizing* foundations: Care/Harm and Fairness/Cheating (Graham et al. 2009; Graham et al. 2012; Haidt 2012). It can be expected that liberal thought is more associated with support for open borders than conservative thought (Graham et al. 2012; Haidt 2012), i.e. to be pro-immigration. From a liberal point of view, the Care/Harm foundation could be portrayed through appeals to the unfortunate circumstances of most immigrants just trying to find a better life. Similarly, the Fairness/Cheating foundation could be portrayed through appeals to treat everyone equally regardless of their religion (e.g. Muslims) or how they entered the country (e.g. unauthorized immigrant status).

H2a. During the period of interest, we expect that tweets that are explicitly anti-immigration should contain more language associated with the Loyalty/Betrayal, Authority/Subversion and Sanctity/Degradation foundations, compared to pro-immigration tweets.

Conservative thought is usually associated with more support for closed borders, and cultural homogeneity than liberal thought, and as discussed, the moral foundations categories that resonate most with conservatives are the *binding* foundations of Loyalty/Betrayal, Authority/Subversion, and Sanctity/Degradation (Graham et al. 2009; Graham et al. 2012; Haidt 2012). The Loyalty/Betrayal foundation could be portrayed through appeals to patriotism and to maintain a certain conception of an American population that is culturally homogenous. Appeals to the Authority/Subversion foundation may be present through clear support for the President as an authority figure and the legislation he sets forward, while appeals to the Sanctity/Degradation foundation may often be present through framing immigrants as very culturally different, disrupting, tainting and ultimately contaminating American culture.

H2b: During the period of interest, we should expect that tweets that are explicitly anti-immigration should contain more third person plural words, compared to pro-immigration tweets.

Such a preoccupation with an outgroup (i.e. immigrants) can be observed by the frequent usage of third person plural nouns within a group (Kaati et al. 2016). In the literature, frequent usage of third person plural noun has been found to be the single best predictor of extremism, indicating that this group delineates itself by opposing the other group (Chung and Pennebaker 2008).

H2c: During the period of interest, we expect tweets that are explicitly anti-immigration to exhibit more language associated with cognitive rigidity and certainty, compared to pro-immigration tweets.

Past research has shown that an intolerance of ambiguity is a strong predictor of political conservatism (Jost et al. 2003). Intolerance for ambiguity is frequently referred to as 'cognitive rigidity' in the psychological literature (Cohen 2012). Those with cognitively rigid thinking styles tend to over-generalize, be over-confident in their convictions and often think in extreme terms (Cohen 2012). In the context of immigration, there is ample ambiguity about costs and benefits of letting immigrants into the country, including concerns about security, financial cost, and cultural impact (Esses et al. 2013). Politically, this dislike of ambiguity might manifest itself with a general mistrust of open-border policies. We thus expect that those who support legislation that closes borders may also use language that is less ambiguous, and more certain. Also, in relation to MFT, the Authority/Subversion foundation is expected to fulfill this need for certainty and stability of conservatives, by offering stability through the maintenance of hierarchy (Koleva et al. 2012). Consistent with the notion of certainty in language use with anti-immigration expression, and consistent with the Authority/Subversion foundation in anti-immigration tweets, we expect that we will also observe more language associated with cognitive rigidity within anti-immigration tweets as well.

> *H3: During the period of interest, we expect to observe that pro-immigration tweets will show more negative emotion in language (anger, anxiety, sadness) than anti-immigration tweets.*

Despite the protests and blockings by the courts against its executive actions, the U.S. government maintained its immigration policy. Pro-immigration tweets, which we hypothesize (H1) to gravitate more on the Fairness/Cheating and Care/Harm foundations should use a much more negative tone as these immigration policies are in direct opposition to the tenets of these foundations. This could result in more negative emotion expressed in pro-immigration tweets compared to anti-immigration tweets, which express perspectives supported by government policy. Thus, we expect the language used in pro-immigration tweets to include more negative emotions compared to anti-immigration tweets.

4. Data and methods

4.1. Tweet extraction

To examine our hypotheses, we performed analyses on a sample of tweets related to immigration policies and events between November 8th, 2016 (the date of the most recent U.S. presidential election), and November 4th, 2017. We chose a period of approximately one year to examine whether the tweet content would show persistent effects over time. We only sampled tweets from Twitter users with profile locations within the United States. Using the Twitter API, we collected two separate random samples of tweets over this period: (1) tweets with explicitly pro-immigration hashtags, and (2) tweets with explicitly anti-immigration hashtags. The hashtags collected for each sample group are presented in Table 1. To select these hashtags, we used Hashtagify (hashtagify.me), an online hashtag tracking tool. Hashtagify allows one to search individual hashtags and find other related hashtags based on pure correlation as well as overall popularity. Our hashtag selection process consisted of two steps. First, we used a set of neutral seed hashtags (#immigration, #immigrant, #muslimban, #daca) to identify hashtags that we judged to be strictly about immigration policies and that were highly popular and correlated hashtags with these seed hashtags. We then chose the most frequently used hashtags that were representative of both sides of the debate, for both the Muslim ban period and announcement of the DACA repeal. For the Muslim ban period, the most popular pro and anti-immigration hashtags were #nobannowall and #buildthewall respectively. For the DACA repeal period the most popular pro and anti-immigration hashtags were #defendDACA and #endDACA respectively. Second, we then used hashtagify to find five or six more hashtags that we both judged to be strictly about immigration policies and were the most popular correlated hashtags to these two hashtags found from the previous step, for each side of the debate. This resulted in a final set of seven pro-immigration and eight anti-immigration hashtags. Using this hashtag filtering process, the final set of hashtags in Table 1 were the most popular hashtags which we judged to be strictly related to immigration policies. Although some hashtags, like #MAGA or #resist, were often used in immigration related tweets, hashtags like this were too ambiguous and could be used in contexts outside of discussing immigration policy. For this reason, we did not include these tweets in our final set.

Using the Twitter streaming API, we continually collected tweets containing these hashtags over a year long period. We filtered this stream to also only include tweets with shared geolocations within the U.S. to ensure that our sample was localized appropriately in order to limit potential noise from foreign influence. In total we extracted 14,126 tweets containing one or more anti-immigration hashtags, and 48,588 tweets containing one or more pro-immigration hashtags over this year-long period. These separate tweet samples were filtered to include only tweets in English, to delete any duplicated tweets, and to exclude tweets containing only hashtags and no additional text. This filtering reduced our samples to 11,213 anti-immigration tweets, and 45,045 pro-immigration tweets used in the analyses. Figure 1 shows the distribution of tweets per week in each sample. From the figures, it is apparent, especially in the pro-immigration sample, that most of the tweets sampled were produced within multiple week periods starting from both the announcement of the Muslim ban

Table 1. Hashtags filtered for in each sample.

Pro-Immigration	Anti -Immigration
#nomuslimban	#buildthewall/#buildthatwall
#nobannowall	#norefugees
#refugeeswelcome	#securetheborder
#heretostay	#deportillegals
#defenddaca	#deportthemall
#dreamactnow	#enddaca/#nodaca
#defenddreamers	#noamnesty

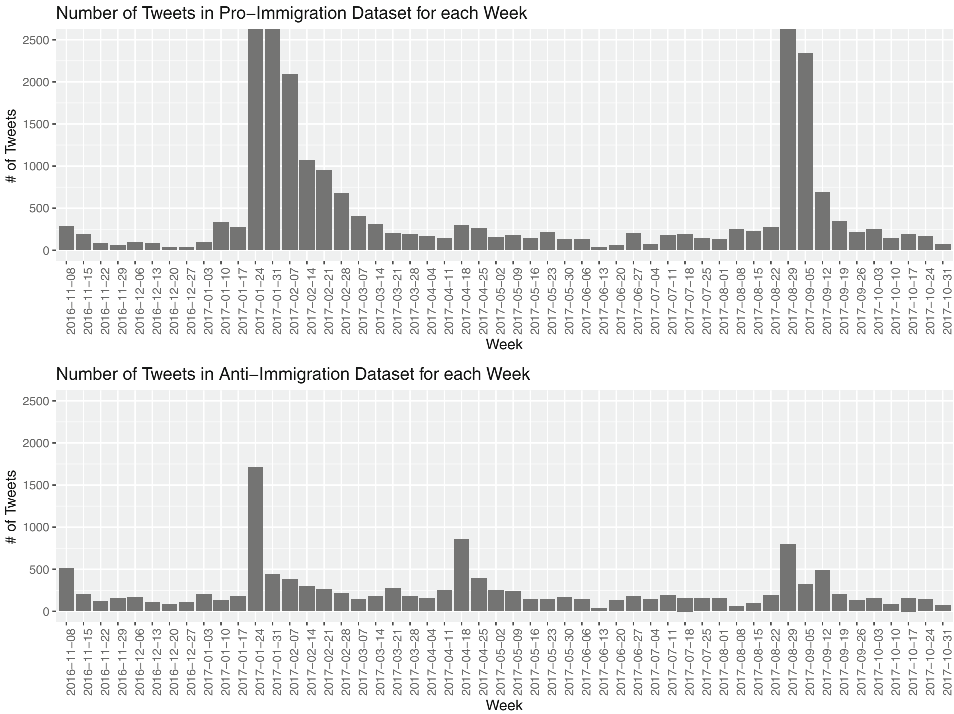

Figure 1. Distribution of Tweets per week for Pro-Immigration and Anti-Immigration hashtags. On the Pro-Immigration graph the y-axis scale is cut at 2500 to maintain symmetry between the graphs. For the weeks of 2017-01-24, 2017-01-31, and 2017-08-29, the number of tweets are 23,650, 5776, and 5216 respectively.

(January 24th, 2017), and the announcement of the DACA repeal (September 5th, 2017). Because of this, we extend our focus to these two events, in addition to a year-long analysis of the data. Table 2 provides an overview of the aggregated wordcount for each sample period.

Twitter users vary greatly, with some users generally observing the activity of their followers passively, while others take a very active role, creating or retweeting new content multiple times a day (DeChoudhury et al. 2012). In recent years, the presence of automated accounts (bots) on Twitter that create or propagate exceptionally large amounts of content has been suspected to have increased substantially, especially within arenas of political debate (Pew Research Center, 2018; Ferrara et al. 2016). These Twitter bot accounts have become progressively more indistinguishable from regular accounts of passionate users, and as a result, literature suggests that their presence on Twitter has played a crucial role in shaping the nature and direction of political discourse (Pew Research Center 2018; Kollanyi et al. 2016; Howard and Kollanyi 2016). In our work, we made a conscious decision to include the potential effect of bots in our analysis, as our focus is on understanding the differences in linguistic characteristics of the rhetoric circulating across both sides of the debate over US immigration policy.

Whereas it is possible that some content was generated by bots, this content nevertheless is playing a role in shaping the rhetoric of pro and anti-immigration conversations.

One characteristic of bots is very active posting of tweets. Table 3 below provides a brief overview of the role that 'extra active' accounts played in our pro and anti-immigration tweet samples:

Our analysis suggests that the potential effect of bot accounts is more pronounced in the anti-immigration sample, with 11.4% of the tweet sample coming from just 10 accounts. This observation aligns with past work showing that many pro Donald Trump tweets were partly driven by bots in the 2016 election (Kollanyi et al. 2016). Ultimately, it is very difficult to determine if these users are indeed bots, or are just passionate or fervent Twitter users. Regardless, we made a deliberate decision to include tweets from these suspected bot accounts given that linguistic style and affect is known to influence the way people perceive political content, regardless of the initial source (Rho et al. 2018). In the limitations section, we explore the role that bots could have played in our sample further, and implications for future research.

4.2. Tweet validation

We expected that the two hashtag categories may not necessarily accurately bifurcate the data as it is possible for the meaning of the tweet to be misaligned with the hashtag (e.g. using an anti-immigration hashtag to espouse a pro-immigration message or vice versa). To determine the degree to which this was a problem, we first had one coder review 2000 randomly sampled tweets from our final sets, and found that this sort of misalignment occurred in only 0.1% percent of tweets in the pro-immigration sample, and 0.8% of the anti-immigration sample.

Next, we also checked to see if hashtag misalignment might be more prevalent during the weeks with more overall tweet traffic in our samples. To check this, we again had the same coder review 2000 randomly sampled tweets from the ten days with the greatest number of tweets in each sample. We found a slight increase in misalignment in the anti-immigration sample (0.1% for pro-immigration tweets and 2.5% for anti-immigration tweets). However, we concluded that the overall effect of this misalignment would be negligible on our overall results.

4.3. Text analysis

The primary tool we used to analyze the text data and address our hypotheses is LIWC (Linguistic Word County and Inquiry) (Pennebaker et al. 2015). LIWC calculates a rate,

Table 2. Aggregated word counts for each sample period.

	Muslim Ban Period	DACA Repeal Period	Post-Election Period
Pro-Immigration	301,083	117,657	423,977
Anti-Immigration	32,678	22,219	96,262

Table 3. Summary of most active accounts in each sample.

Sample	Sample Size	# Distinct Accounts	% of Tweets from top 10	Min/Max avg. per day for top 10	Min/Max # of days for top 10
Pro-Immigration	45,045	31,066	1.3%	Min = 1.05, Max = 2.67	Min = 36, Max = 102
Anti-Immigration	11,213	5233	11.4%	Min = 1.4, Max = 3.53	Min = 50, Max = 139

This table presents the sample size of each sample (Sample Size), the number of distinct usernames in the sample (# Distinct Accounts), the percentage of the sample that comes from the top 10 most active accounts (% of Tweets from top 10), the minimum and maximum average posting rate per day for the top 10 most active accounts (Min/Max avg. per day for top 10), and the minimum and maximum number of days with posting history for the top 10 most active accounts (Min/Max # of days for top 10).

or percentage, of words in a sample of text belonging to several predefined categories. LIWC produces a percentage score for any size text sample, where all words in the target category are given the same weight (i.e. all contribute to the overall score equally). Past empirical studies on the validity of LIWC have found that it is reliably able to detect meaning from text in a wide variety of contexts, as well as detect emotional states, intentions, motivations, thinking styles, and individual differences (Pennebaker et al. 2015). In our work, we use the latest available LIWC dictionary (LIWC 2015), which provides a broader range of word categories relating to different social and psychological processes compared to earlier LIWC versions (Pennebaker et al. 2015; Tausczik and Pennebaker 2010). We chose to use LIWC as our text analysis tool for two primary reasons. First, LIWC has been used extensively across social science research, including in research contexts quite similar to our own working on social media data (De Choudhury et al. 2013, 2014). Second, we perceived many of LIWC's built in dictionary categories to be the best proxies available to address our research questions. Although other dictionaries exist, like the National Research Council of Canada (NRC) Sentiment and Emotion Lexicons (Mohammad and Turney 2010), these dictionaries categories are developed through crowdsourcing, and are therefore less theoretically refined compared to LIWC. To keep methodological consistency to past work in similar domains, and to have as much confidence in our research proxies as possible, we chose LIWC as our research tool.

To address our first two hypotheses (H1 and H2a) related to Moral Foundations Theory, we use the Moral Foundations Dictionary (**MFD**), which is an extension to the base LIWC dictionary that adds additional word categories for each of the five moral foundations (available at www.moralfoundations.org/othermaterials). The MFD dictionary categories are used in the same manner as all other built-in LIWC categories, where all words in each category are weighted equally, and an overall percentage rate score is calculated from a sample of text. The MFD was created, tested, and validated by Graham et al. (2009). To create this dictionary, they first began by getting research assistants to create a large set of synonyms and antonyms to the names of each of the five moral foundations, and then this set was further refined by the authors to exclude words that seemed to be ambiguous or not very morally salient. Using this refined set, they investigated hundreds of transcripts of both liberal and conservative sermons over a period of 12 years (1994 to 2006), and chose words that had the largest differences in relative frequency between the liberal and conservative sermon transcripts. The final set consisted of 295 unique word stems. The MFD has been used previously to analyze blogs, analyze the content of blogs discussing the 'Ground Zero Mosque' (Dehghani et al. 2014), and to study the public debate over stem-cell research (Clifford and Jerit 2013).

To address our next two hypotheses (H2b and H2c) that anti-immigration tweets should use more third-person plural words, as well as language associated with cognitive rigidity than pro-immigration tweets, we apply the LIWC '3rd Person Plural' and 'Certainty' lexical categories to the text (Pennebaker et al. 2015). The '3rd person plural' LIWC category counts all third person plural references and maps directly to our hypothesis. The LIWC 'Certainty' category is a sub-category under the 'Cognitive Processes' super-category (Pennebaker et al. 2015). We propose it as the best publicly available proxy for studying cognitive rigidity in language, as all words within this category are associated with minimal ambiguity (e.g. *always, never*), and it has a strong correlation ($r = 0.39$) with Cohen's specialized dictionary designed specifically to detect cognitive rigidity within text (Cohen 2012)[1].

To address our final hypothesis (H3) where we expect more negative emotive language over the entire post-election period in pro-immigration tweets compared to anti-immigration tweets, we apply the LIWC 'Negative Emotion' word super-category, which is composed of separate sub-categories for 'Anger', 'Anxiety', and 'Sadness' (Pennebaker et al. 2015).

5. Results

5.1. Global tweet sample summaries

Table 4 presents a high-level descriptive account of the differences within each Post-Election period sample. Comparing basic tweet statistics using independent t-tests, we found that anti-immigration tweets used more hashtags ($t(18185) = 25.28$, $p < .0001$, difference in means = .50), and used slightly fewer words ($t(21767) = -10.50$, $p < .0001$, difference in means = .73) (excluding hashtags, @ symbols, and URLs).

In addition to basic descriptive statistics, we were also interested to see if there existed qualitative difference in the type of words associated with specific terms of interest within each tweet sample (pro-immigration vs. anti-immigration). To

[1] We were not able to obtain a copy of Cohen's cognitive rigidity dictionary to use in the analysis.

perform this analysis, we looked at the co-occurrence of terms in each sample, or in other words, what words seem to appear together across different contexts within each sample. To, capture and analyze word co-occurrence in each sample we applied GloVe (Pennington et al. 2014), an unsupervised learning algorithm that creates vector representations of all words in a sample from their aggregated word-word co-occurrence statistics. GloVe first builds a term co-occurrence matrix, and then constructs the word vectors from this matrix (Pennington et al. 2014). GloVe is known as a word-embedding algorithm that allows one to examine how semantically related two words are in a corpus of text, based on how terms appear to co-occur within a given window of text before or after a target word. GloVe vector representations have been used to develop intelligent systems for word associations (e.g. man is to king, like woman is to queen), amongst other applications (Pennington et al. 2014). We built a model with word vectors of 24 dimensions (we tried dimension lengths of 8, 16, and 24 and found 24 to give the most interpretable results), and a possible co-occurrence window of up to 4 words to the left and right of each word for each sample (as on average most tweets contained 8 to 9 words ($\overline{x}$ = 8.623 for anti-immigration tweets, $\overline{x}$ = 9.354 for pro-immigration tweets)). We only included terms in the model that occurred at least 20 times in each sample to avoid including uncommon and outlier words in our model (Pennington et al. 2014). By obtaining cosine similarity measures between the vector representation of a word of interest and all other words in a sample, we can get an idea of which terms are most semantically like the word of interest within the sample. We chose to investigate key terms that often act as the subjects for political division within US immigration policy (“Immigrant”, “Muslim”, “Refugee”), as we thought these terms might illuminate immediate key differences in overall sentiment between the samples. Table 5 presents the top 10 words with the highest similarity to these chosen terms. The range for cosine similarity is from −1 (most dissimilar) to 1 (most similar).

The results show noticeable differences in terms of themes that appear between the patterns of word co-occurrence. For pro-immigration tweets, we observe that there are several words related to the idea of family and community that co-occur with these chosen terms (e.g. *family, community, families, sisters, neighbors*). In contrast, for anti-immigration tweets, we observe that there is an abundance of words often associated with law, policy and national security that co-occur with these chosen terms (e.g. *policy, aliens, criminal, laws, terrorists, legal, illegal(s)*). We also observe a stark difference in the terminology that commonly co-occurs with “Immigrant”

Table 4. Differences in basic tweet statistics between pro-immigration (PRO) and anti-immigration (ANTI) tweets.

	Samp. mean No. of hashtags	Samp. mean words per tweet
PRO	2.39	9.35
ANTI	2.89	8.62

(*alien* for anti-immigration tweets vs. *undocumented* for pro-immigration tweets). In this work, we aim to use these co-occurrence statistics to supplement the results of our outlined hypotheses by comparing the differences in word co-occurrence with the global differences we find in language use within each sample.

5.2. Lexical hypotheses

To address our lexical hypotheses (H1 and H2a) related to moral foundations theory (that pro-immigration tweets should use more Care/Harm and Fairness cheating words, and anti-immigration tweets should use more Loyalty/Betrayal, Authority/Subversion, and Sanctity/Degradation words), we added the categories from the MFD as linguistic categories to be used by the LIWC engine. Before adding the MFD categories to LIWC engine, we excluded the word root 'immigra*' from the list of words in MFD Loyalty/Betrayal category to avoid skewed results due to our research topic being immigration. For our other lexical hypotheses (H2b, H2c, and H3), we used linguistic categories already available with LIWC (as outlined in the data and methods section). As a few of the weeks of our tweet samples had an inordinate number of tweets (e.g. the week immediately preceding the announcement of the Muslim ban), we did not want to let this sample imbalance create bias in our analysis. To counteract the effect of imbalanced word counts between weeks, we apply LIWC to the tweet data within each week in our samples to obtain a single

Table 5. Top 10 words with highest cosine similarity to controversial terms for pro-immigration (PRO) and anti-immigration (ANTI) tweets.

	Immigrant	Muslim	Refugee	DACA
PRO	undocumented(.924)	countries (.883)	neighbors (.897)	trump (.856)
	rights (.908)	sisters (.873)	young (.891)	florida (.856)
	goal(.815)	ban (.864)	muslim (.836)	texas (.834)
	youth (.811)	refugee (.836)	bans (.812)	defend (.807)
	standing(.784)	bans (.820)	sisters (.794)	utah (.790)
	family (.770)	travel (.818)	value (.780)	ending (.785)
	community(.766)	blocks (.803)	lowest (.762)	muslim (.772)
	lowest (.756)	trumps (.802)	trumps (.749)	program (.743)
	resettlement (.746)	donald (.799)	unamerican (.748)	ban (.740)
	families (.730)	stay (.780)	remain (.734)	state (.724)
ANTI	policy (.909)	laws (.817)	legal (.835)	now (.919)
	aliens (.864)	end (.774)	policy (.797)	sure (.910)
	immigration (.810)	ban (.763)	illegals (.772)	problem (.741)
	immigrants (.796)	act (.735)	alien (.748)	immigrants (.717)
	countries (.783)	countries (.692)	stand (.551)	change (.661)
	alien (.778)	ask (.663)	immigrant (.746)	left (.657)
	criminal (.775)	everything (.633)	stop (.743)	travel (.649)
	refugee (.746)	across (.632)	illegal (.743)	thank (.636)
	cut (.729)	control (.629)	means (.733)	damn (.622)
	means (.703)	terrorists (.628)	long (.690)	supports (.615)

word-proportion metric for each week, and then apply Welch's independent sample t-tests to compare the LIWC scores for all weeks across the anti-immigration and pro-immigration samples. Welch's t-test adjusts the degrees of freedom to account for the unequal variances (Delacre et al. 2017). This procedure gives a comparison of sustained difference across time, independent of word count for that week. Figure 2 shows the distribution of word count per week in each sample over the post-election period for each sample.

In addition to performing comparative statistics over the twelve-month post-election period, we also perform analyses on the Muslim Ban and DACA repeal sub-periods as well. As the original Muslim ban (Executive order 13769) was in effect for a 7-week period from January 27, 2017 until it was superseded by Executive order 13780 on March 16, 2017 (Vergani 2018), we chose to create a separate data set aggregating tweet data by week for the Muslim ban period for 8-weeks starting from three days before the announcement of the executive order (January 24, 2017), until four days after it was superseded (March 20, 2017). We did this to capture additional small windows of tweets leading up to and preceding the event. To be consistent across groupings, we also created another separate data set aggregating the DACA repeal data by week, starting from three days before the announcement of the DACA repeal (September 1) (Venkataramani and Tsai 2017) until 8 weeks later on November 4, 2018.

5.2.1. *Care/Harm and Fairness/Cheating Foundations in pro-Immigration Tweets*

Our first hypothesis addressed whether pro-immigration tweets used more language associated with the individualizing foundations (Care/Harm (CH) and Fairness/Cheating (FC)) than anti-immigration tweets. We first examine these tweets over the period of one year to see if there is a change over time between the pro and anti-immigration samples. Figure 3 presents graphs of the LIWC scores for the CH and FC foundations by week over the year-long period. Notice that we see an initial spike in the CH foundation in the pro-immigration tweets in the weeks following the 2016 presidential election. An F-test on a model to account for difference in best-fit slopes between pro-immigration and anti-immigration samples showed no significant increases in explained variance for either CH ($F(2,101) = 0.050$, $p = .82$), or FC ($F(2,101) = 0.15$, $p = .70$) over a null model with parallel slopes. This means that between pro-immigration and anti-immigration samples, there were no significant differences in the slope of the best fit trend line through the data.

We next test our hypothesis. Table 6 show the results of the Welch's t-tests applied to the data[2]. Looking at the data for the entire post-election period, we observe that the proportion of Care/Harm (CH) foundation words is consistently higher in the pro-

[2] Degrees of Freedom are rounded to the nearest whole number. Degrees of Freedom are adjusted in the Welch's t-test differently than standard paired t-tests.

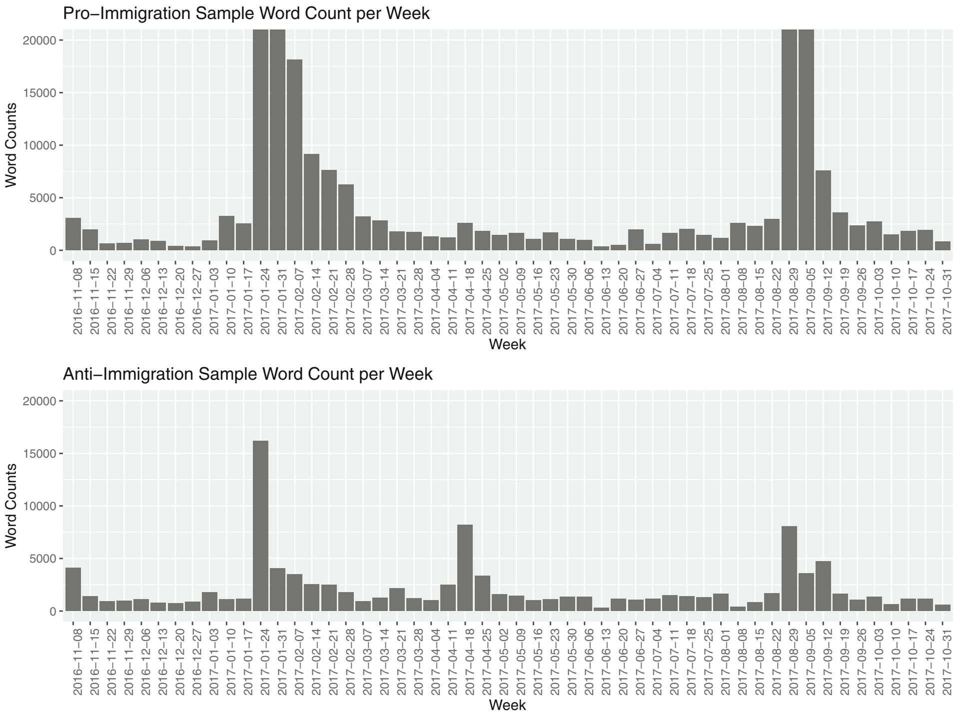

Figure 2. Distribution of word counts per week for Pro-Immigration and Anti-Immigration samples. On the Pro-Immigration graph the y-axis scale is cut at 20000 to maintain symmetry between the graphs. For the weeks of 2017-01-24, 2017-01-31, 2017-08-29, and 2017-09-05 the word counts are 207,538, 46,283, 69,986, and 28,098 respectively.

immigration sample ($p < .001$). We also observe during the entire post-election period, that the proportion of Fairness/Cheating (FC) foundation words is consistently higher in the pro-immigration sample ($p < .001$). Across both the pro-immigration and anti-immigration samples, the average proportion of CH words was higher than the average proportion of FC words (mean = .96, sd = .59 vs. mean = .22, sd = .18 for pro-immigration; mean = .58, sd = .25 vs. mean − .07, sd = .06 for anti-immigration).

Differences in expression of these foundations within the Muslim Ban period and DACA sub-periods were present. Compared to anti-immigration tweets, pro-immigration tweets used more FC words than CH words during the Muslim Ban sub-period ($t = 3.896$ vs. $t = 2.531$), while using more CH words than FC words during the DACA Repeal sub-period ($t = 4.343$ vs. $t = 1.628$). Taken together, these results support our hypothesis of pro-immigration tweets containing more CH and FC language than anti-immigration tweets.

5.2.2. *Loyalty/Betrayal, Authority/Subversion, Sanctity/Degradation Foundations in Anti-Immigration Tweets*

Hypothesis 1b addressed whether anti-immigration tweets used more language associated with the binding foundations (Loyalty/Betrayal (**LB**), Authority/Subversion (**AS**), and Sanctity/Degradation (**SD**)) than pro-immigration tweets. We first look at whether this language use changed over time between the samples. Figure 4 presents graphs of the LIWC scores for the LB, AS, and FC foundations by week, over the year-long period. An F-test on a model accounting for difference in regression slopes between pro-immigration and anti-immigration samples showed a weak trend of more explained variance in LB ($F(2,101) = 3.46$, $p = .07$), and SD ($F(2,101) = 3.36$, $p = .07$) over a null model with parallel slopes.

For LB, pro-immigration tweets showed a steeper negative slope over time compared to anti-immigration tweets ($\beta = -.012$ vs. $\beta = -.003$). For SD, pro-immigration tweets showed a slightly steeper positive slope over time compared to anti-immigration tweets ($\beta = .006$ vs. $\beta = .003$). However, a similar F-test procedure on AS showed no increase in explained variance for a model with differing slopes ($F(2,101) = 1.94$, $p = .166$). This suggests that for LB, language in the pro-immigration sample was decreasing over time, while SD language in the pro-immigration sample was increasing, although the fit is weak. For AS, there were no significant differences in the slope of the best fitting trend line through the data between pro-immigration and anti-immigration samples.

We next test our hypothesis. Table 8 presents the results of the Welch's t-tests applied to the data. Looking at the data for the entire post-election period, unexpectedly we observed that the proportion of LB foundation words was consistently higher in the pro-immigration sample ($p < .001$). We observe similar significant differences within the Muslim Ban ($p < .01$), and DACA repeal ($p < .001$), sub-periods as well.

We chose to investigate these surprising results further, and looked at the presence of LB words in the pro-immigration sample by calculating the frequency of each word in the LB MFD category in the sample and found that the top 5 most frequent LB words in the sample were *community, solidarity, family*, *united, and families*. In contrast, the most frequent LB words in the anti-immigration sample were *nation, terrorist, family, united, group, foreign*, and *families*. To better understand the results of the hypothesis test, we wanted to understand how the LB foundation was being expressed in the pro-immigration tweets. We randomly sub-sampled 300 tweets from the sample that contained these 5 most frequent terms and had one coder qualitatively code the chosen tweets using an in vivo coding technique (Saldaña 2015), looking at phrases that occurred with these top 5 most frequent terms in the tweets. We found strong recurrent themes of expressing one's support and allegiance to communities committed to fighting and protesting perceived injustices (for either the Muslim Ban or DACA repeal). Tables 7 and 8 presents some chosen example tweets (with URLs and @ mentions removed) that best portray these themes. The results from our word co-occurrence analysis for pro-immigration tweets also supports these conclusions,

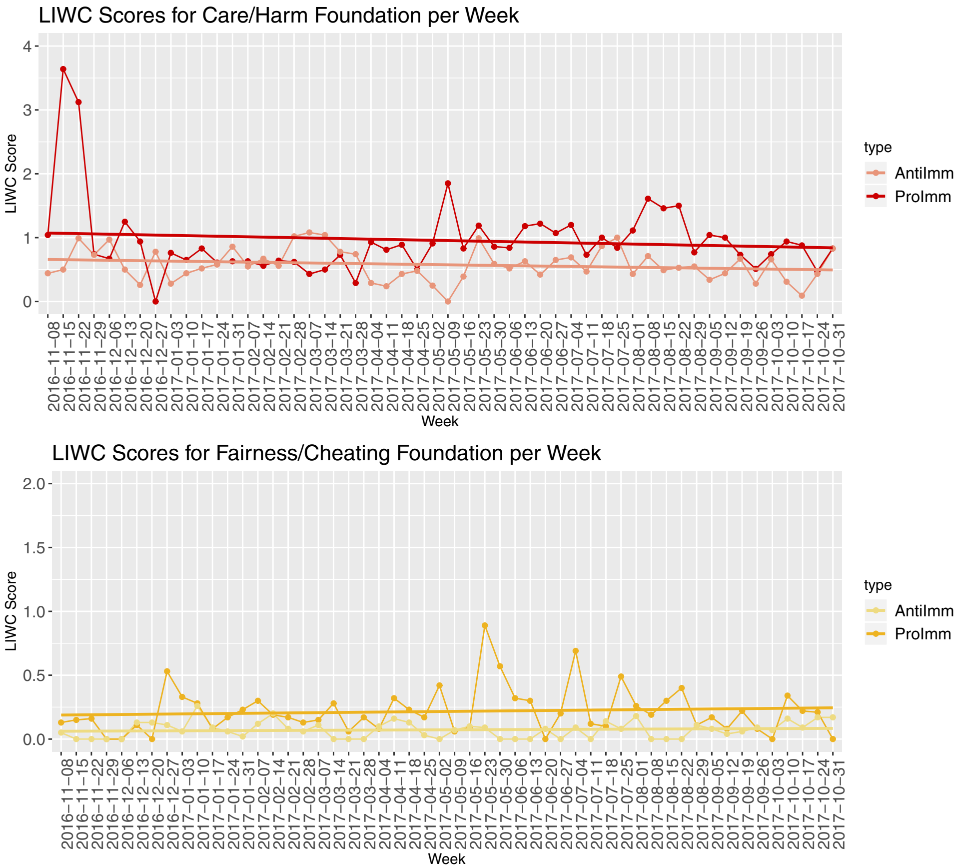

Figure 3. LIWC scores for Care/Harm and Fairness/Cheating Foundations per week. ('AntiImm' are anti-immigration tweets; 'ProImm' are pro-immigration tweets).

as many of these binding LB words (e.g. solidarity, community) were found to co-occur with the terms *immigrant, muslim,* and *refugee.* Overall, our results suggest that LB words were primarily used in the pro-immigration sample to convey solidarity with immigrants.

Over the entire election period, we also observed a higher proportion of AS words in the anti-immigration sample ($p < .001$). We observe similar significant differences for AS within the Muslim Ban ($p < .001$), and DACA repeal ($p < .001$), sub-periods as well.

The results for the SD foundation are inconclusive. There were no statistically significant differences in the proportion of SD words between the tweet samples over the post-election period. However, there was a weak trend of pro-immigration tweets having more SD words during the DACA repeal sub-period ($p = .07$).

Taken together, the results partially support our conclusion. As expected, anti-immigration tweets used more AS words. However, our finding that pro-immigration

Table 6. Independent sample t-test results on weekly data for CH and FC.

Foundation	Pro Prop. Mean	Anti Prop. Mean	t	p	df
Care/Harm – Muslim Ban	.80	.58	2.53	.03*	8.50
Care/Harm – DACA Repeal	.83	.42	4.34	<.001***	13.71
Care/Harm – Post Election	.96	.58	4.26	<.001***	69.10
Fairness/Cheating – Muslim Ban	.20	.08	3.90	.02*	13.99
Fairness/Cheating – DACA Repeal	.15	.09	1.63	.14	8.58
Fairness/Cheating – Post Election	.22	.07	5.45	<.01**	64.01

Pro Prop. Mean is the sample proportion mean for pro-Immigration tweets, *Anti Prop. Mean* is the sample proportion mean for anti-immigration tweets. Values are percentages of all words falling in the category (e.g. .600 is 0.60% of all words).
* - $p < .05$, ** - $p < .01$, *** - $p < .001$.

tweets used more LB words, contradicts our hypothesis, while the result for the presence of SD words remains inconclusive.

5.2.3. *Third person plural in anti-Immigration tweets*
Next, we focus on our hypothesis that anti-immigration tweets should use more third-person plural words than pro-immigration tweets. To address this hypothesis, we apply the LIWC '3rd Person Plural' lexical categories to the text (Pennebaker et al. 2015). We first look whether this language use changed over time between the samples. Figure 5 presents graphs of the LIWC scores for the 3rd person plural category by week over the year-long period. An F-test on a model comparing difference in regression slopes between pro-immigration and anti-immigration samples showed no significant increases in explained variance for 3rd person plural words ($F(2,101) = .10$, $p = .75$) over a null model with parallel slopes. This means that between pro-immigration and anti-immigration samples, there were no significant differences in the slope of the best fit line through the data for the 3rd person plural category.

We next test our hypothesis, i.e. for differences. Table 9 presents the results of the Welch's t-tests applied to the data. Considering the presence of 3rd person plural over the entire post-election period, we observe that the proportion of certainty words is consistently higher in the anti-immigration sample ($p < .001$). We observe similarly significant, but less strong differences within the Muslim Ban period ($p < .01$), but within the DACA Repeal sub-period there is no significant difference between anti-immigration tweets and pro-immigration tweets ($p = .32$). These results generally support our hypothesis, as we observe more third person plural words within anti-immigration tweets over the entire post-election period.

5.2.4. *Third person plural in anti-immigration tweets*
Next, we focus on our hypothesis that anti-immigration tweets should use more language associated with cognitive rigidity than pro-immigration tweets. To address

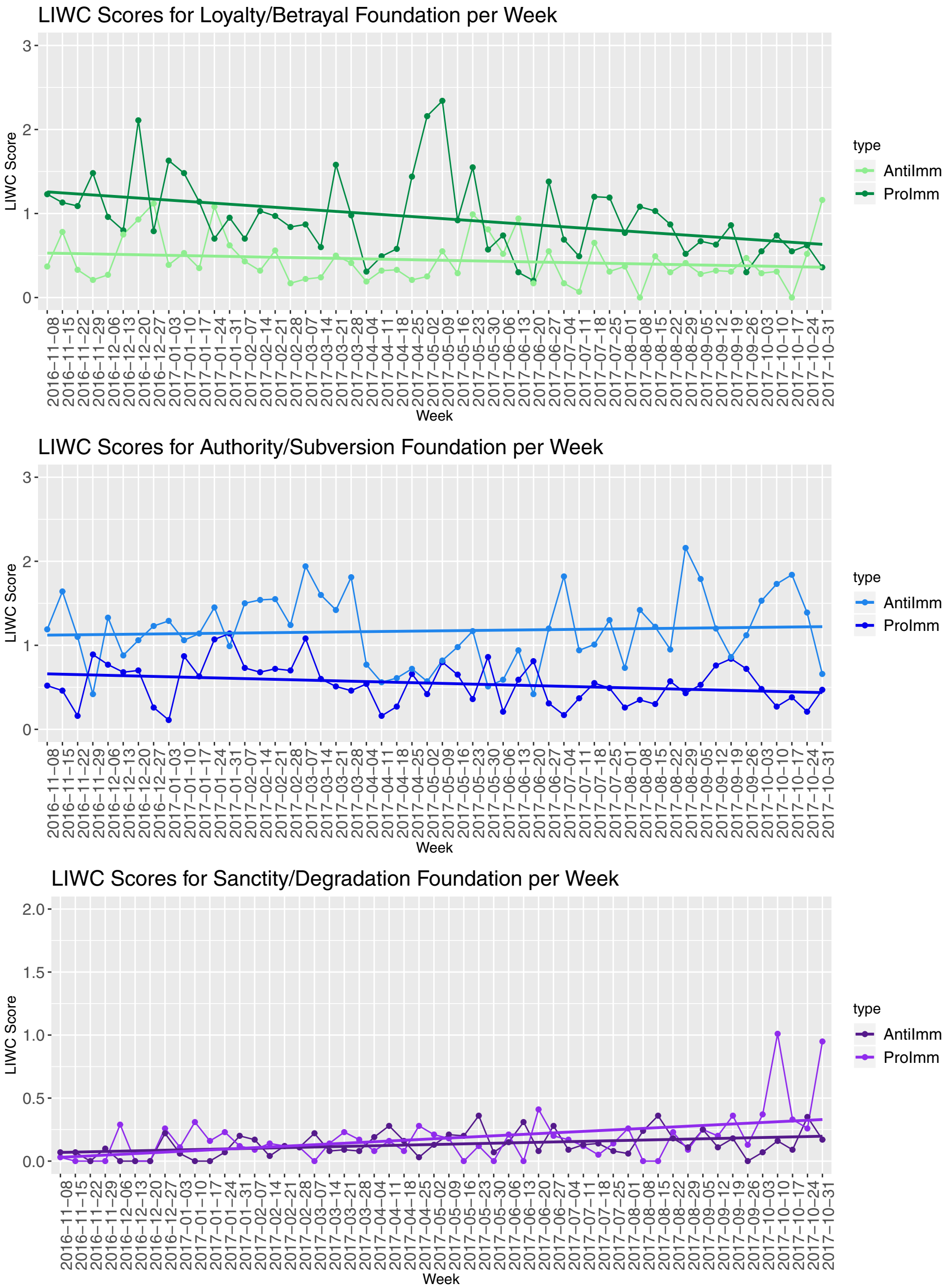

Figure 4. LIWC scores for Loyalty/Betrayal and Authority/Subversion Foundations per week. ('AntiImm' are anti-immigration tweets; 'ProImm' are pro-immigration tweets).

Table 7. Welch's t-test results for tweet data aggregated by week for LB, AS, and SD.

Foundation	Pro Prop. Mean	Anti Prop. Mean	t	p	df
Loyalty/Betrayal – Muslim Ban	.83	.46	3.17	<.01**	10.38
Loyalty/Betrayal – DACA Repeal	.60	.30	3.98	<.001**	13.51
Loyalty/Betrayal – Post Election	.95	.46	6.57	<.001***	82.57
Authority/Subversion – Muslim Ban	.84	1.48	−5.13	<.001**	13.26
Authority/Subversion – DACA Repeal	.55	1.53	−5.67	<.001**	9.89
Authority/Subversion – Post Election	.55	1.17	−9.11	<.001**	83.14
Sanctity/Degradation – Muslim Ban	.12	.13	.234	.82	14.00
Sanctity/Degradation – DACA Repeal	.34	.12	2.10	.07	7.96
Sanctity/Degradation – Post Election	.18	.12	1.54	.13	74.760

Pro Prop. Mean is the sample proportion mean for pro-Immigration tweets, *Anti Prop. Mean* is the sample proportion mean for anti-immigration tweets. Values are percentages of all words falling in the category (e.g. .600 is 0.60% of all words).
* - $p < .05$, ** - $p < .01$, *** - $p < .001$.

this hypothesis, we apply the LIWC 'Certainty' category to the text (Pennebaker et al. 2015). In LIWC, the 'Certainty' category is a sub-category under the 'Cognitive Processes' super-category (Pennebaker et al. 2015). Figure 6 presents graphs of the LIWC scores for the Certainty category by week over the year-long period. An F-test on a model comparing difference in regression slopes between pro-immigration and anti-immigration samples showed no significant increases in explained variance for the Certainty category ($F(2,101) = .48$, $p = .49$) over a null model with parallel

Table 8. Sample pro-immigration tweets containing loyalty/betrayal language.

- 'I stand in **solidarity** vs. @realDonaldTrump's agenda of fascism, sexism, racism, and xenophobia #heretostay #stoptrump'
- 'I stand in **solidarity** with the Yemeni business owners striking today. They've always stood with me #nobannowall'
- **'#Solidarity** #ResistTrump #NoBanNoWall let's protect our undocumented **family** from state violence!'
- 'So beautiful. **Unity** is something incredible. We need to remain **united** during these times #NoBanNoWall'
- 'My **family**, my **community** seek dignity. When our **community** is under attack we rise up and fight back bc we are #heretostay #DACA'

slopes. This means that between pro-immigration and anti-immigration samples, there were no significant differences in the slope of the best fit line through the data for the Certainty category.

We next test our hypothesis, i.e. for differences. Table 10 presents the results of the Welch's t-tests applied to the data. Looking at the data for the entire post-election period, we observe that the proportion of certainty words is consistently higher in the anti-immigration sample ($p < .001$). We observe similarly significant, but less strong differences within the DACA repeal sub-period ($p = .03$), but only a weak trend of difference during the Muslim Ban sub-period ($p = .07$). These results generally support our hypothesis, as we observe more certainty words within anti-immigration tweets over the entire post-election period.

5.2.5. *Negative emotion in pro-immigration tweets*

Our last hypothesis asserted that negative emotion should be higher in pro-immigration tweets than anti-immigration tweets. To do so we apply the LIWC 'Negative Emotion' word super-category, which is composed of separate sub-categories for 'Anger', 'Anxiety', and 'Sadness' (Pennebaker et al. 2015) to both tweet sample aggregated by week, and built linear models to analyze linear trends in the expression of the categories. Figures 7 and 8 present graphs of the LIWC scores for the Negative Emotion category and its sub-categories by week over the year-long period for the pro and anti-immigration tweets. An F-test on a model comparing difference in regression

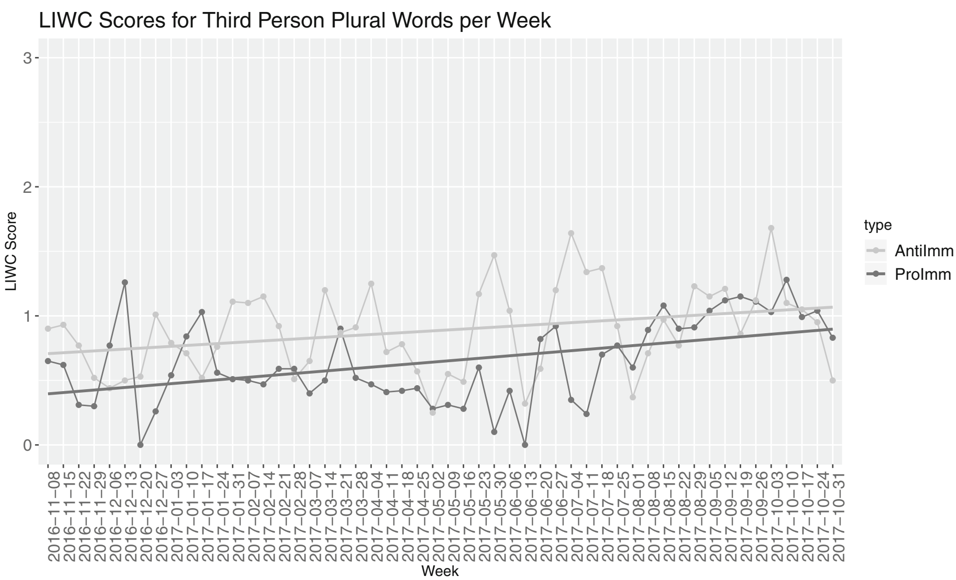

Figure 5. LIWC scores for 3rd person plural words per week. ('AntiImm' = anti-immigration tweets; 'ProImm' = pro-immigration tweets).

slopes between pro-immigration and anti-immigration samples showed no significant increases in explained variance for either Negative Emotion ($F(2,101) = .21$, $P = .65$), Anger ($F(2,101) = .10$, $p = .75$), Sadness ($F(2,101) = 2.76$, $p = .10$), or Anxiety ($F(2,101) = .56$, $p = .46$) over a null model with parallel slopes. This means that between pro-immigration and anti-immigration samples, there were no significant differences in the increase or decrease of negative emotion over time.

We next test our hypothesis. Table 11 presents Welch's t-tests applied to the samples aggregated by week over the yearlong period. In the post-election period, we see that anti-immigration tweets have a higher proportion of negative emotion words ($p < .0001$) and anger words ($p < .0001$). There were no statistically significant differences in anxiety words ($p = .09$), or sadness words ($p = .12$).

Pro Prop. Mean is the sample proportion mean for pro-Immigration tweets, *Anti Prop. Mean* is the sample proportion mean for anti-immigration tweets. Values are percentages of all words falling in the category (e.g. .600 is 0.60% of all words). * - $p < .05$, ** - $p < .01$, *** - $p < .001$.

5.2.6. *Summary of results*

Table 12 presents a summary of our results. To summarize, we found direct support for hypotheses H1, H2b, and H2c, as we observed pro-immigration tweets contained more language associated with CH and FC, and anti-immigration tweets contained more 3rd person pronouns and language associated with cognitive rigidity. We only found partial support for H2a as anti-immigration tweets contained only more language associated with AS. Unexpectedly, we found pro-immigration tweets to contain more language associated with LB, and we found no significant differences in the presence of language associated with SD between the two tweet samples. Last, we observed more language associated with negative emotion in anti-immigration tweets rather than pro-immigration tweets, which contradicted our hypothesis in H3.

Table 9. Welch's t-test results for tweet data aggregated by week for 3rd person plural words.

Category	Pro Prop. Mean	Anti Prop. Mean	t	p	df
3rd Person Plural – Muslim Ban	.52	.93	−4.36	<.01**	7.87
3rd Person Plural – DACA Repeal	1.08	1.18	−1.05	.32	10.07
3rd Person Plural – Post Election	.65	.89	−3.71	<.001***	101.93

Pro Prop. Mean is the sample proportion mean for pro-Immigration tweets, *Anti Prop. Mean* is the sample proportion mean for anti-immigration tweets. Values are percentages of all words falling in the category (e.g. .600 is 0.60% of all words).
* - $p < .05$, ** - p < .01, *** - $p < .001$.

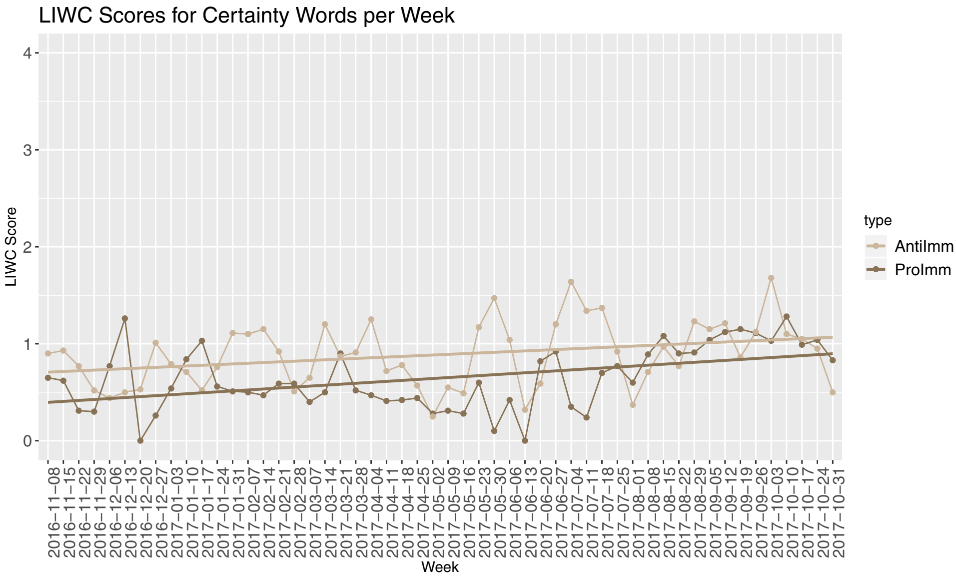

Figure 6. LIWC scores for Certainty words per week. ('AntiImm' = anti-immigration tweets; 'ProImm' = pro-immigration tweets).

6. Discussion

In our work, we aimed to provide a thorough understanding of the underlying moral, cognitive, and affective differences in linguistic style for both sides of the ongoing debate over American immigration policies since the election of U.S. president Donald Trump. To the best of our knowledge, our work is the first to examine the language of U.S. debate over immigration using these methods. As many countries become more polarized politically (Boxell et al. 2017), we contend that it is becoming increasingly necessary to investigate how this polarization is being expressed through the rhetoric of both opposing sides. Twitter has become a popular

Table 10. Welch's t-test results for tweet data aggregated by week for certainty words.

Category	Pro Prop. Mean	Anti Prop. Mean	t	p	df
Certainty – Muslim Ban	1.40	1.69	−1.98	.07	8.50
Certainty – DACA Repeal	1.51	1.83	−2.37	.03*	12.98
Certainty – Post Election	1.40	1.96	−6.23	<.001**	99.99

Pro Prop. Mean is the sample proportion mean for pro-Immigration tweets, *Anti Prop. Mean* is the sample proportion mean for anti-immigration tweets. Values are percentages of all words falling in the category (e.g. .600 is 0.60% of all words).
* - $p < .05$, ** - $p < .01$, *** - $p < .001$.

medium for political dialogue for both politicians and the public, and we are only starting to understand the implications of its ability to enable the creation and propagation of a diverse set of viewpoints immediately and broadly in a public sphere. Twitter, and social media in general, has the power to act as valuable tool to analyze trends, differences, and insights into ongoing rhetoric on contentious issues.

Our work first makes a conceptual contribution to the study of moral rhetoric in the context of the ongoing immigration debate within the United States. We found that pro-immigration tweets gathered over this year-long period consistently used more language associated with the individualizing foundations (Care/Harm (CH) and Fairness/Cheating (FC)), which aligns both with our hypothesis and past literature that shows these foundations to be the most heavily weighted by liberals (Graham et al. 2009; Graham et al. 2012; Haidt 2012), who generally prefer more open-borders and looser restrictions on immigration. We also observed that during the Muslim Ban, pro-immigration tweets had more of an emphasis on the FC foundation, while during the DACA repeal pro-immigration tweets had a stronger emphasis on the CH foundation. One straightforward explanation for this difference could be the differences in the perceptions of the groups affected by these policies and legislation. As the DACA repeal has direct implications just for young children under the age of 14 (Venkataramani and Tsai 2017), it makes sense that the CH foundation could be expressed more as young children are almost always regarded as a vulnerable

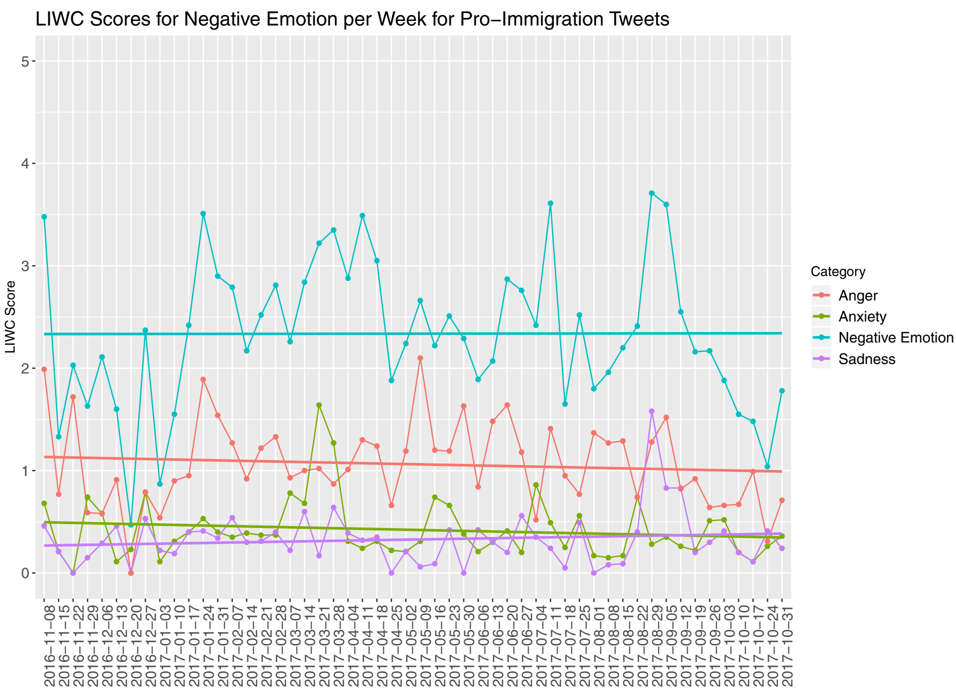

Figure 7. LIWC scores for Negative Emotion words for Pro-Immigration Tweets.

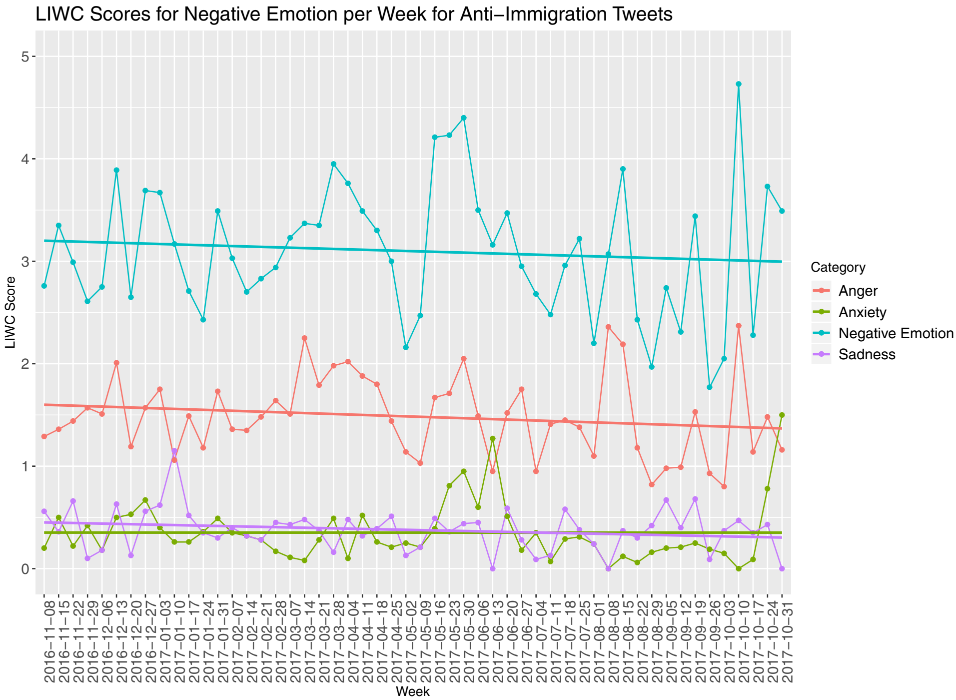

Figure 8. LIWC scores for Negative Emotion words for Anti-Immigration Tweets.

demographic. For the Muslim ban sub-period, people of all ages were potentially affected, so it is likely that the legislation was perceived as a general injustice for all, which might mitigate the expression of the CH foundation compared to the FC foundation.

We only found partial evidence for our hypothesis that anti-immigration tweets would use more language associated with the binding foundations (Loyalty/Betrayal (LB), Authority/Subversion (AS), and Sanctity/Degradation (SD)). The only foundation that we observed being expressed more in the language of the anti-immigration tweets was AS, which suggests that supporting the President as an authority figure and the legislation he sets forward was framed as a primary moral concern within our sample. There were no significant differences between anti-

Table 11. Welch's t-test results for tweet data aggregated by week for Negative Emotion words.

Category	Pro Prop. Mean	Anti Prop. Mean	t	p	df
Negative Emotion	2.34	3.10	−7.98	<.01**	100.85
Anger	1.06	1.48	−7.47	<.01**	101.99
Anxiety	.42	.35	1.70	.09	101.56
Sadness	.33	.38	−1.55	.12	96.89

Table 12. Summary of results.

Hypothesis	Result
H1 – More CH and FC language in Pro-Immigration Tweets	Supported
H2a – More LB, AS, and SD language in Anti-Immigration Tweets	Partially Supported
H2b – More 3rd person pronouns in Anti-Immigration Tweets	Supported
H2c – More cognitively rigid language in Anti-Immigration Tweets	Supported
H3 – More negative emotion language in Pro-Immigration Tweets	Rejected

immigration tweets and pro-immigration tweets with respect to the SD foundation, and the overall rate of SD language in both tweet samples was quite low, which may mean that in the context of immigration, the SD foundation may not be very relevant within public discourse. The results of the LIWC word co-occurrence analysis (Table 3) suggest that rhetoric associated with opposition to immigration and incoming immigrants is more likely to use language showing concern for national security and legal issues than language showing concern for pollution or tainting of American culture.

Surprisingly, we found that pro-immigration tweets used more language associated with LB, which partially contradicted our hypothesis on the presence of all binding foundations being higher only in anti-immigration tweets. Research from Clifford (2017) suggests that the individual differences in the LB foundation is a robust predictor of an individual's political partisan strength on both sides of the political spectrum. Based on our results and Clifford's findings, we theorize that those on the liberal pro-immigration side may express and gravitate towards language associated with LB on social media in response to perceived threats (e.g. Muslim Ban, DACA repeal) to perhaps show allegiance to and help strengthen their own political "team". Future research should investigate if large-scale expression of LB on twitter and social media is present on the more typically liberal sides of other contentious issues (e.g. gun control, police brutality etc.), and if so, if there has there been an overall increase in the expression of LB from liberal perspectives over spanning over the last few years as socio-political polarization has increased, in the U.S (Boxell et al. 2017) but also in Europe as well.

In contrast to the pro-immigration sample which showed themes of inclusion and fostering group-identity, we saw that the anti-immigration sample showed a stronger preoccupation with a perceived outgroup through their more frequent use of third person plural words compared to the pro-immigration sample (Chung and Pennebaker 2008). In addition, we observed that anti-immigration tweets contained more language associated with cognitive rigidity. Both observations may align with past research showing lowered tolerance for ambiguity and uncertainty as well as higher threat sensitivity as predictors of political conservatism (Jost et al. 2003). Identifying consistent markers of conservative (or liberal) thought on social media could help policy makers and social scientists track how public opinion on important issues shifts or evolves over time. Overall, we aim to grow an understanding of how these predictors relate to political discussion in the context of immigration.

Last, we observed that anti-immigration tweets had consistently higher rates of overall negative emotion and anger words than pro-immigration tweets, which contradicted our hypotheses. We find this result somewhat surprising, as we expected those who are on the side fighting against government legislation, and who value moral foundations of Care/Harm and Fairness/Cheating, to produce tweets with stronger negative undertones on social media. However, recent research suggests that there exists a link between negative sentiment in the social media posts of populist political leaders (Donald Trump, Geert Wilders, Nigel Farage, Narendra Modi), and their ability to gain rapid popularity (Pal et al. 2018). Pal et al. (2018) found that the tweets of these populist leaders frequently contained negative and antagonistic messaging, and the more antagonistic the tweets were (e.g. personal insults and sarcasm), the more widely the tweets were retweeted. Our results compliment these findings, and suggest that similar rhetorical styles seem to characterize the tweets of the supporters of populist leaders, at least in the context of immigration. More negative sentiment in anti-immigration tweets may be an effective rhetorical strategy in propagating information, as research suggests that political information expressed in negative language persists in individual memory longer compared to neutral expressions (Utych 2018). Future work should investigate if supporters of populist leaders have similar rhetorical patterns in debating different social policies beyond just immigration.

6.1. Twitter as a public sphere for analyzing political debate

The format of Twitter as a social media platform seems to make it an ideal communication medium for studying moral foundations and related topics through language use *in the wild* as it allows and encourages people to provide their feedback, in a variety of different ways (e.g. sharing content, retweeting, liking), on all sorts of topics from the most endearing to the most enraging, in a restricted character format. Twitter especially affords users the opportunity to share their “snap judgements” (i.e. potentially unstudied thought), on contentious issues, where they can be easily spread, propagated, and popularized within online communities. As a result, Twitter may be an ideal candidate to study the social intuitions underlying “snap judgments” and how they relate to moral psychology on a massive scale (Haidt 2001). Future research should investigate Twitter in comparison to other social media platforms (e.g. Facebook, Reddit), which offer different affordances, for studying moral intuitions on a large scale.

In our work, we found clear differences in the expression of moral themes and attitudes expressed through Twitter between the two samples on opposing sides of the ongoing American immigration debate. Twitter allows for finer grained “snapshots” into the characteristics of rhetoric of important socio-political issues than traditional survey measures can realistically gather, and therefore our work has implications for social scientists and policy makers who aim to obtain real-time insight into the underlying moral themes and emotional patterns inherent to opposing sides of ongoing debates, beyond just surface level differences in topics discussed. For example, obtaining ongoing insight into conflicting moral frameworks in relation to public issues through social media can help inform swift changes to policies and political communication to dynamically adapt to current moral consensus of a target group (e.g. Zhang and Counts 2016). Furthermore, dynamically detecting emotional sentiment of rhetoric related to new government policy through social media can help policy makers better track, understand, and react to how their actions are perceived by the public at large (Zhang and Counts 2015, 2016). Overall, our work contributes to the growing body of literature in CSCW and HCI that situates Twitter as an increasingly important and useful research tool to understand societal behavior, opinions, and rhetorical strategies.

6.2. Incorporating MFT into future value sensitive design research

Although our work presents only a retrospective descriptive account of moral differences in rhetorical framing across both sides of a debate on a contentious public issue, we hope to start a broader conversation on how empirically grounded and established theories of morality and individual differences in human values can be integrated into Value Sensitive Design (VSD) and CSCW research more generally. As modern VSD approaches have stressed the importance of establishing a strong empirical grounding of values of system users (Le Le Dantec et al. 2009; Borning and Muller 2012), we see MFT as being one method for understanding the values of system users, especially when considering differences in values on an individual level, and the net effects that these individual differences may have when aggregated at a population level. Le Le Dantec et al. (2009) note that when a strong empirical investigation into the values of users is conducted, this can then help designers reflectively evaluate these findings onto conceptual values that are traditionally employed in VSD research (e.g. human welfare, trust, accountability, universal usability, ownership and property, freedom of expression etc. (Friedman et al. 2013)).

For instance, for conceptual values like universal usability, individual differences amongst system users in how much they value the fairness/cheating foundation may be an important factor in determining the degree to which concepts of equality and reciprocity are actively championed and promoted across users. Likewise, the degree to which ownership of intellectual property developed within a system is valued by users may have important ties to the degree in which its users generally value a clear hierarchy of governance (the authority/subversion foundation). The degree to which promoting human welfare within a system has clear ties to how much the care/harm foundation is valued by its users. Conflicts over freedom expression across system users may also arise from a conflict over the perceived need to protect fellow users from harm (the care/harm foundation), and the perceived rights of everyone to express themselves freely (the fairness/cheating foundation).

An important utility of MFT that can allow for empirical investigations of the prevalence of these values within a population of users are the methodological tools that accompany the theory. Past work exploring methods for studying value dimensions in VSD suggests that when trying to explore potential value conflicts between users and designers, survey or content analysis methods are ideal (Shilton et al. 2014). MFT provides both survey measures, in the form of the Moral Foundations Questionnaire (MFQ) designed to obtain the degree to which individuals' value each of the five foundations in the moral domain (Graham et al. 2011), and a textual content analysis tool, the Moral Foundations Dictionary (MFD), which we applied in our own work. Both tools provide well-validated and easily employable methods for obtaining information about the moral values of system users and designers, across a variety of design contexts that may differ in both size and scope.

Across many of these different and common conceptual values that have historically been a focus of VSD, we see clear connections to how MFT and theories from moral psychology more generally can bring additional insight towards understanding the values of system users, and using this insight to inform system design and governance. We hope that MFT can not only be employed as a theoretical construct in the retrospective analysis of the nature of conflict and cooperation within online spaces, but also proactively for grass-roots system design and maintenance. To the best of our knowledge, this is an unexplored area of future interdisciplinary research and investigation.

7. Limitations

Our study has several limitations. First, it should be restated that the distribution of tweets per week in our samples was imbalanced, with an inordinate number of tweets in some weeks (e.g. first week of the Muslim ban) compared to others. This was primarily due to two factors: (1) to the nature of the tweet hashtags we filtered for, as some had a shorter surge in popularity than others, and (2), that the Twitter API only allowed to access only up to 1% of all Twitter traffic. To try and counteract the imbalanced distribution of tweets, we chose to use independent t-tests on the raw LIWC scores on the tweets aggregated for each week in each sample, which let us shift the emphasis from word counts to differences across time (ie. analysis by week independent of the word count for that week). Still, for the weeks with less tweets, it is difficult to be sure that the tweets obtained are not an accidental biased sample that is not representative of the overall discourse for the week. However, we found that accounting for word count in our analysis had negligible effects. We performed independent t-tests weighted by the logarithm of the word count per week to see if this would change our results, and found that these results were consistent with our reported results; all significant results remained significant, and all non-significant results remained non-significant. As we only sampled a small number of specific hashtags related to immigration related issues, we cannot claim that our results can be generalized to the entire immigration debate within the US over this period. These hashtags were chosen primarily to give insight into key controversial events (Muslim Ban and DACA repeal) over our period of interest. Our results are better interpreted as being representative of a smaller more focused debate over specific U.S. immigration policy related issues.

Another limitation of our work is a result of the limitations of LIWC as a text analysis tool. As LIWC relies on just detecting the presence of predefined categories of words, it unfortunately has no ability to consider the context in which these words are used. By using LIWC categories to address our hypotheses, we take these categories to be a proxy for examining our phenomenon of interest, but they are not perfect indicators. One issue is that LIWC gives all words in a category an equal weighting, even though their direct relationship to a specific trait may vary. For instance, words like *family*, which is a word in the MFD for the Loyalty/Betrayal foundation, can be used in contexts that do not necessarily portray a moral sentiment.

Third, a potential is the role that bots may have played in our tweet samples. However, as stated in the data and methods section, we made a deliberate decision to include the effect of bots in our analysis. We argue that the presence of suspected bot accounts may not only influence the way the rest of the users in our data form and express their opinion, but also is an inevitable component of reality that shapes discourse in the online sphere. Future research should investigate how the linguistic style of suspected bot accounts compares to that of regular users across a variety of policy issues, to see if there are any significant differences in rhetorical strategies.

Our study also has a potential limitation in that there was no way to know the proportion of tweets in our sample that were produced by immigrants. It is possible that the moral concerns of immigrant populations might be different than those of non-immigrant populations. However, in a Twitter data study like ours, there is no way to identify demographic variables like this from the Tweets themselves.

We also examined the rhetoric of the immigration debate in a U.S. pro and anti-immigration sentiment, which is not only a U.S. phenomenon, but it is found worldwide, especially currently in Europe. We cannot say for certain the extent to which our results would generalize to a non-U.S. population. However, there is ample evidence for the same liberal and conservative divide in moral foundations across natural and cultural contexts (Graham et al. 2011).

Last, Twitter and survey methods each have their own benefits and drawbacks, and ideally both methods should supplement each other. We see our work as putting forward a simple and replicable framework for studying large-scale patterns of moral and emotional sentiment, that can be applied to a wide range of important topics of discussion, beyond just immigration.

8. Conclusion

In this paper, we present an empirical investigation into the moral, emotional, and cognitive differences in linguistic style between two large samples of tweets representing opposing sides of the political debate over immigration related issues in the United States since the 2016 election. We observe significant differences between the two sides in terms of the moral themes expressed, expression of negative emotion, and use of cognitively rigid language. Our work contributes to an understanding of modern moral rhetoric in social media in the context of immigration, and presents a repeatable framework for social scientists and policy makers to study political communication over social media in a wide variety of different contexts. We see unexplored opportunities to integrate similar empirical investigations of system user values into future research in CSCW, primarily within the Value Sensitive Design (VSD) paradigm.

References

Abokhodair, Norah; and Sarah Vieweg (2016). Privacy & social media in the context of the Arab Gulf. *Proceedings of the 2016 ACM conference on designing interactive systems, Brisbane, Australia, 4 – 8 June 2016*. New York: ACM Press, pp. 672–683.

Abrajano, Marisa; and Zoltan L. Hajnal (2017). *White backlash: Immigration, race, and American politics*. Princeton: Princeton University Press.

Borning, Alan; and Michael Muller (2012). Next steps for value sensitive design. *CHI '12*: *Proceedings of the SIGCHI conference on human factors in computing systems. Austin, Texas, 5 - 10 May 2012*. New York: ACM Press, pp. 1125–1134.

Boulus-Rødje, Nina; Pernille Bjørn; and Ahmad Ghazawneh (2015). "It's about business not politics": Software development between Palestinians and Israelis. In *ECSCW 2015: Proceedings of the 14th European conference on computer supported cooperative work, Oslo, Norway, 19 – 23 September 2015*. Cham: Springer, pp. 43–61.

Boxell, Levi; Matthew Gentzkow; and Jesse M. Shapiro (2017). *Is the internet causing political polarization? Evidence from demographics*, No. w23258. National Bureau of Economic Research.

Boyd, Ryan L.; and James W. Pennebaker (2017). Language-based personality: A new approach to personality in a digital world. *Current opinion in behavioral sciences*, vol. 18, pp. 63–68.

Chung, Cindy K.; and James W. Pennebaker (2008). Computerized text analysis of Al-Qaeda transcripts. *The Content Analysis* reader. Thousand Oaks, California: Sage.

Chung, Wen-Ting; Kai Wei; Yu-Ru Lin; and Xidao Wen (2016). November. The dynamics of group risk perception in the US after Paris attacks. In *SocInfo '16*: *International Conference on Social Informatics, Bellevue, USA, 11 – 14 November, 2016*. Cham: Springer. pp. 168–184.

Clifford, Scott (2017). Individual differences in group loyalty predict partisan strength. *Political Behavior*, vol. 39, no. 3, pp. 531–552.

Clifford, Scott; and Jennifer Jerit (2013). How words do the work of politics: Moral foundations theory and the debate over stem cell research. *The Journal of Politics*, vol. 75, no. 3, pp. 659–671.

Cohen, Shuki J. (2012). Construction and preliminary validation of a dictionary for cognitive rigidity: Linguistic markers of overconfidence and overgeneralization and their concomitant psychological distress. *Journal of psycholinguistic research*, vol. 41, no. 5, pp. 347–370.

Conover, Michael; Jacob Ratkiewicz; Matthew R. Francisco; Bruno Gonçalves; Filippo Menczer; and Alessandro Flammini (2011). Political polarization on twitter. In *ICWSM'11*: *Proceedings of AAAI International Conference on Weblogs and Social Media, Barcelona, Spain, 17 – 21 July, 2011*. Menlo Park: AAAI. pp. 89–96.

Dawson, Sharon L; and Graham A. Tyson (2012). Will morality or political ideology determine attitudes to climate change. *Australian Community Psychologist*, vol. 24, no. 2, pp. 8–25.

Day, Martin V.; Susan T. Fiske; Emily L. Downing; and Thomas E. Trail (2014). Shifting liberal and conservative attitudes using moral foundations theory. *Personality and Social Psychology Bulletin*, vol. 40, no. 12, pp. 1559–1573.

De Choudhury, Munmun; Nicholas Diakopoulos; and Mor Naaman (2012). Unfolding the event landscape on twitter: Classification and exploration of user categories. In *CSCW'12*: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, Seattle, USA, 11 – 15 February 2012*. New York: ACM Press. pp. 241–244.

De Choudhury, Munmun; Scott Counts; and Eric Horvitz (2013). Major life changes and behavioral markers in social media: Case of childbirth. In *CSCW'13*: *Proceedings of the 2013 conference on Computer supported cooperative work, San Antonio, USA, 23 – 27 February 2013*. New York: ACM Press. pp. 1431–1442.

De Choudhury, Munmun; Andrés Monroy-Hernandez; and Gloria Mark (2014). Narco emotions: Affect and desensitization in social media during the Mexican drug war. In *CHI'14*: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems, Toronto, Canada, 26 April – 1 May 2014*. New York: ACM Press. pp. 3563–3572.

Dehghani, Morteza; Kenji Sagae; Sonya Sachdeva; and Jonathan Gratch (2014). Analyzing political rhetoric in conservative and liberal weblogs related to the construction of the "ground zero mosque". *Journal of Information Technology & Politics,* vol. 11, no. 1 pp. 1–14.

Delacre, Marie; Daniël Lakens; and Christophe Leys (2017). Why psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology,* vol. 30, no. 1.

Esses, Victoria M.; Stelian Medianu; and Andrea S. Lawson (2013). Uncertainty, threat, and the role of the media in promoting the dehumanization of immigrants and refugees. *Journal of Social Issues*, vol. 69 no. 3, pp. 518–536.

Ferrara, Emilio; Onur Varol; Clayton Davis; Filippo Menczer; and Alessandro Flammini (2016) The rise of social bots. *Communications of the ACM*, vol. 59, no. 7 pp. 96–104.

Friedman, Batya; Peter H. Kahn; Alan Borning; and Alina Huldtgren (2013). Value sensitive design and information systems. *Early engagement and new technologies: Opening up the laboratory.* Dordrecht: Springer. pp. 55–95.

Graham, Jesse; and Jonathan Haidt (2012). Sacred values and evil adversaries: A moral foundations approach. In Mario Mikulincer; and Phillip R. Shaver (eds): *The social psychology of morality: Exploring the causes of good and evil.* Washington, DC: American Psychological Association, pp. 11–31.

Graham, Jesse; Jonathan Haidt; and Brian A. Nosek (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, vol. 96, no. 5, pp. 1029–1046

Graham, Jesse; Brian A. Nosek; Jonathan Haidt; Ravi Iyer; Spassena Koleva; and Peter H. Ditto (2011). Mapping the moral domain. *Journal of personality and social psychology,* vol. 101, no. 2, pp. 366–385

Graham, Jesse; Brian A. Nosek; and Jonathan Haidt (2012). The moral stereotypes of liberals and conservatives: Exaggeration of differences across the political spectrum. *PloS one*, vol. 7, no. 12, article e50092.

Haidt, Jonathan, (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological review*, vol. 108, 4, pp. 814–834.

Haidt, Jonathan, (2012). *The righteous mind: Why good people are divided by politics and religion*. New York: Vintage.

Haidt, Jonathan; and Jesse Graham (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, vol. 20, no. 1, pp. 98–116.

Hainmueller, Jens; and Daniel J. Hopkins (2014). Public attitudes toward immigration. *Annual Review of Political Science*, vol. 17, pp. 225–249.

Hanna, Alexander; Chris Wells; Peter Maurer; Lew Friedland; Dhavan Shah; and Jörg Matthes (2013). Partisan alignments and political polarization online: A computational approach to understanding the French and US presidential elections. *Proceedings of the 2nd Workshop on Politics, Elections and Data*. New York: ACM Press. pp. 15–22.

Hemphill, Libby; and Andrew J. Roback (2014). Tweet acts: How constituents lobby congress via twitter. In *CSCW '14*: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, Baltimore, USA, 15 – 19 February 2014*. New York: ACM Press. pp. 1200–1210.

Hoover, Joe; Kate Johnson; Reihane Boghrati; Jesse Graham; and Morteza Dehghani (2018). Moral Framing and Charitable Donation: Integrating exploratory social media analyses and confirmatory experimentation. *Collabra: Psychology*, vol. 4, no. 1, pp. 9

Howard, Philip N.; and Bence Kollanyi. (2016). Bots, #StrongerIn, and #Brexit: Computational propaganda during the UK-EU referendum (20 June 2016). Rochester: SSRN. https://doi.org/10.2139/ssrn.2798311

Ji, Qihao, and Arthur A. Raney (2015). Morally judging entertainment: A case study of live tweeting during Downton Abbey. *Media Psychology,* vol. 18, no. 2, pp. 221–242.

Jost, John T.; Jack Glaser; Arie W. Kruglanski; and Frank J. Sulloway (2003). Political conservatism as motivated social cognition. *Psychological bulletin,* vol. 129, no. 3, pp. 339–375.

Kaati, Lisa; Amendra Shrestha; Katie Cohen; and Sinna Lindquist (2016). Automatic detection of xenophobic narratives: A case study on Swedish alternative media. In *ISI'16*: *2016 IEEE Conference on Intelligence and Security Informatics, Tucson, USA*, 27 – 30 September, 2016. Piscataway, New Jersey: IEEE. pp. 121–126.

Kaur, Rishemjit; and Kazutoshi Sasahara (2016). Quantifying moral foundations from various topics on Twitter conversations. In *IEEE Big Data 16'*: *2016 IEEE International Conference on Big Data, Washington D.C., USA, 5 – 9 December, 2016*. Piscataway, New Jersey: IEEE. pp. 2505–2512.

Koleva, Spassena P.; Jesse Graham; Ravi Iyer; Peter H. Ditto; and Jonathan Haidt (2012). Tracing the threads: How five moral concerns (especially purity) help explain culture war attitudes. *Journal of Research in Personality*, vol. 46, no. 2, pp. 184–194.

Kollanyi, Bence; Philip N. Howard; and Samuel C. Woolley (2016). Bots and automation over twitter during the first US presidential debate. *Comprop data memo*, pp. 1–4.

Kou, Yubo; Yong Ming Kow; Xinning Gui; and Waikuen Cheng (2017). One social movement, two social media sites: a comparative study of public discourses. *Computer Supported Cooperative Work (CSCW)*, vol. 26, nos. 4–6, pp. 807–836.

Le, Huyen T.; G. R. Boynton; Yelena Mejova; Zubair Shafiq; and Padmini Srinivasan (2017). Revisiting the American voter on twitter. In *CHI'17*: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, *Denver, USA, 6 – 11 May, 2017*. New York: ACM Press. pp. 4507–4519.

Le Dantec, Christopher A; Erika Shehan Poole; and Susan P. Wyche (2009). Values as lived experience: Evolving value sensitive design in support of value discovery. In *CHI '09*: *Proceedings of the SIGCHI conference on human factors in computing systems, Boston, USA, 4 – 9 April, 2009*. New York: ACM Press. pp. 1141–1150.

Maddock, Jim; Kate Starbird; Haneen J. Al-Hassani; Daniel E. Sandoval; Mania Orand; and Robert M. Mason (2018). Characterizing online rumoring behavior using multi-dimensional signatures. In *CSCW'18*: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, New York City, USA, 3 – 7 November, 2018*. New York: ACM Press. pp. 228–241.

Mäkiniemi, Jaana-Piia; Anna-Maija Pirttilä-Backman; and Michelle Pieri (2013). The endorsement of the moral foundations in food-related moral thinking in three European countries. *Journal of agricultural and environmental ethics,* vol. 26, no. 4, pp. 771–786.

Manders-Huits, Noëmi (2011). What values in design? The challenge of incorporating moral values into design. *Science and engineering ethics*, vol. 17, no. 2, pp. 271–287.

McCright, Aaron M.; and Riley E. Dunlap (2011). The politicization of climate change and polarization in the American public's views of global warming, 2001–2010. *The Sociological Quarterly,* vol. 52, no. 2, pp. 155–194.

Mohammad, Saif M.; and Peter D. Turney (2010). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, Los Angeles, USA, 5 February, 2010*. Stroudsberg: Association for Computational Linguistics. pp. 26–34.

O'Rourke, Kevin H.; and Richard Sinnott (2006). The determinants of individual attitudes towards immigration. *European journal of political economy,* vol. 22, no. 4, pp. 838–861.

Pal, Joyojeet; Udit Thawani; Elmer van der Vlugt; Wim Out; and Priyank Chandra (2018). Speaking their mind: Populist style and antagonistic messaging in the tweets of Donald Trump, Narendra Modi, Nigel Farage, and Geert Wilders. *Computer Supported Cooperative Work (CSCW)*, vol. 27, nos. 3–6, pp. 1–34.

Park, Kunwoo; Ingmar Weber; Meeyoung Cha; and Chul Lee (2016). Persistent sharing of fitness app status on twitter. In *CSCW'16*: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, San Francisco, USA, 27 February – 3 March, 2016.* New York: ACM Press. pp. 184–194.

Pennebaker, James W.; Ryan L. Boyd; Kayla Jordan; and Kate Blackburn (2015). *The development and psychometric properties of LIWC2015.*

Pennington, J.; R. Socher.; and C. Manning (2014). Glove: Global vectors for word representation. In *EMLNP'14*: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25 – 29 October, 2018*. New York: ACM Press. pp. 1532–1543.

Pew Research Center: Internet, Science & Tech (2018). *Twitter bots: An analysis of the links automated accounts share*. Online. Available at: http://www.pewinternet.org/2018/04/09/bots-in-the-twittersphere/. Accessed 2 November 2018.

Quercia, Daniele; Jonathan Ellis; Licia Capra; and Jon Crowcroft (2012). Tracking gross community happiness from tweets. In CSCW'12: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, *Seattle, USA, 11 – 15 February 2012*. New York: ACM Press. pp. 965–968.

Rai, Tage Shakti; and Alan Page Fiske (2011). Moral psychology is relationship regulation: moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review,* vol. 118, no. 1, pp. 57-75.

Rho, Eugenia H.R.; Gloria Mark; and Melissa Mazmanian (2018). Fostering civil discourse online: Linguistic behavior in comments of #MeToo Articles across political perspectives. *Proceedings of the ACM on Human-Computer Interaction – CSCW*, vol. 2, article 147.

Sagi, Eyal; and Morteza Dehghani (2014). Measuring moral rhetoric in text. *Social Science Computer Review,* vol. 32, no. 2, pp. 132–144.

Saldaña, Johnny, (2015). *The coding manual for qualitative researchers*. Thousand Oaks, California: Sage.

Schein, Chelsea and Kurt Gray (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, vol. 22, no. 1, pp. 32–70.

Sharma, Eva, Koustuv Saha, Sindhu Kiranmai Ernala, Sucheta Ghoshal and Munmun De Choudhury (2017). Analyzing ideological discourse on social media: A case study of the abortion debate. In CSS'17: *Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas*, Santa Fe, USA, 19 October – 22 October, 2017. New York: ACM Press. pp. 3.

Shilton, Katie, Jes A. Koepfler, and Kenneth R. Fleischmann (2014). How to see values in social computing: methods for studying values dimensions. *Proceedings of the 17th ACM conference on*

Computer supported cooperative work & social computing, Baltimore, USA, 15 February – 19 February, 2014. New York: ACM Press. pp. 426–435.

Tausczik, Yla R. and James W. Pennebaker (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, vol. 29, no. 1, pp.se, USA, 7 – 12 May, 2016. New York: ACM Press. pp. 3378–3389.

Utych, Stephen M. (2018). Negative affective language in politics. *American Politics Research*, vol. 46, no. 1, pp. 77–102.

Vainio, Annukka; and Jaana-Piia Mäkiniemi (2016). How are moral foundations associated with climate-friendly consumption? *Journal of Agricultural and Environmental Ethics,* vol. 29, no. 2, pp. 265–283.

Van de Vyver; Julie Diane M. Houston; Dominic Abrams; and Milica Vasiljevic (2016). Boosting belligerence: How the July 7, 2005, London bombings affected liberals' moral foundations and prejudice. *Psychological science,* vol. 27, no. 2, pp. 169–177.

Venkataramani, Atheendar S.; and Alexander C. Tsai (2017). Dreams deferred—The public health consequences of rescinding DACA. *New England Journal of Medicine*, vol. 377, no. 18, pp. 1707–1709.

Vergani, Matteo (2018). *How is terrorism changing us?*. Basingstoke: Palgrave Macmillan.

Vieweg, Sarah; Amanda L. Hughes; Kate Starbird; and Leysia Palen (2010). Microblogging during two natural hazards events: What twitter may contribute to situational awareness. In *CHI'10*: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *Atlanta, USA, 10 – 15 April, 2010*. New York: ACM Press. pp. 1079–1088.

Zhang, Amy X.; and Scott Counts (2015). Modeling ideology and predicting policy change with social media: Case of same-sex marriage. In *CHI'15*: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, *Seoul, Korea, 18 – 23 April, 2015*. New York: ACM Press. pp. 2603–2612.

Zhang, Amy X.; and Scott Counts (2016). Gender and ideology in the spread of anti-abortion policy. In *CHI'16*: *Proceedings of 2016 CHI Conference on Human Factors in Computing Systems*, *San Jose, USA, 7 – 12 May, 2016*. New York: ACM Press. pp. 3378–3389.