Incorporating the Measurement of Moral Foundations Theory into Analyzing Stances on Controversial Topics

Rezvaneh Rezapour
rezapou2@illinois.edu
University of Illinois at Urbana-Champaign
Champaign, USA

Ly Dinh
dinh4@illinois.edu
University of Illinois at Urbana-Champaign
Champaign, USA

Jana Diesner
jdiesner@illinois.edu
University of Illinois at Urbana-Champaign
Champaign, USA

ABSTRACT

This paper investigates the correlation between moral foundations and the expression of opinions in the form of stance on different issues of public interest. This work is based on the assumption that the formation of values (personal and societal) and language are interrelated, and that we can observe differences in points of view in user-generated text data. We leverage the Moral Foundations Theory to expand the scope of stance analysis by examining the narratives in favor or against several topics. Applying an expanded version of the Moral Foundations Dictionary to a benchmark dataset for stance analysis, we capture and analyze the relationships between moral values and polarized online discussions. Using this enhanced methodology, we find that each social issue has different "moral and lexical profiles." While some social issues project more authority related words (Donald Trump), others consists of words related to care and purity (abortion and feminism). Our correlation analysis of stance and morality revealed notable associations between stances on social issues and various types of morality, such as care, fairness, and loyalty, hence demonstrating that there are certain morality types that are more attributed to stance classification than others. Overall, our analysis highlights the usefulness of considering morality when studying stance. The differences observed in various viewpoints and stances highlights linguistic variation in discourse, which may assist in analyzing cultural values and biases in society.

CCS CONCEPTS

• **Computing methodologies** → **Natural language processing**; • **Information systems** → *Information extraction*; • **Human-centered computing** → *Collaborative and social computing*.

KEYWORDS

Moral Foundations Theory, stance analysis, social media, controversial topics, text analysis

HT '21, August 30-September 2, 2021, Virtual Event, Ireland

ACM ISBN 978-1-4503-8551-0/21/08…$15.00
https://doi.org/10.1145/3465336.3475112

ACM Reference Format:
Rezvaneh Rezapour, Ly Dinh, and Jana Diesner. 2021. Incorporating the Measurement of Moral Foundations Theory into Analyzing Stances on Controversial Topics. In *Proceedings of the 32nd ACM Conference on Hypertext and Social Media (HT '21), August 30-September 2, 2021, Virtual Event, Ireland.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3465336.3475112

1 INTRODUCTION

People use social media platforms such as Twitter to share information, viewpoints and ideas, read the news, be entertained, and connect with others, among other purposes [16]. Online conversations can include discussions of social issues [34, 44], which may feature the expression of opinions, emotions, and polarizing discussions among community members. Social issues have some of their roots in subjective perceptions of cultural, societal, and political characteristics and differences. These personal or group-based differences in perceptions and beliefs can lead to debates and conflicts, which can be deeply divisive and contain biased justifications [41]. To solve social issues, legislative changes or gradual reforms at the individual, organizational, and societal level might be needed. Online conversations can serve as one among several sources for recognizing opposing points of view, also referred to as stances, which is a necessary ingredient for bridging gaps between groups, facilitating constructive conversations, and mitigating biases.

To extract stances, researchers from various fields have leveraged large amounts of online, user-generated texts and real-time conversations on social media platforms. One direction of scholarly work on this issue, also known as stance analysis, aims to detect people's position on a topic (in favor or against). Computational solutions for this purpose typically use a combination of linguistic features and machine learning algorithms to build binary classifiers [3, 37, 38]. While such approaches are helpful for categorizing people and terms along stances, in this paper, we study online debates from a different perspective, namely by considering morality as an additional independent variable for stance analysis. This work is based on our assumptions that (a) online controversies can have their roots in individual beliefs related to features such as gender and race, political orientation, and other cultural characteristics [1, 46], and (b) people's everyday language, written or verbal, represents some of these beliefs [52]. Moreover, Moral Foundations Theory (MFT) assumes that emotional and cognitive intuitions, also known as foundations, influence personal judgements and (moral) decision-making "through a psychological preparedness to attend, approve or disapprove particular aspects of situations or issues prior to any conscious reasoning process" [20, 21, 26]. MFT

Dimensions		Explanation
Virtue	**Vice**	
Care (CareVirtue)	Harm (CareVice)	Protecting versus hurting others
Fairness (FairnessVirtue)	Cheating (FairnessVice)	Cooperation/ trust/ just versus cheating in interaction with objects and people
Loyalty (IngroupVirtue)	Betrayal (IngroupVice)	Ingroup commitment (to coalitions, teams, brands) versus leaving group
Authority (AuthorityVirtue)	Subversion (AuthorityVice)	Playing within the rules of hierarchy versus challenging hierarchies
Purity (PurityVirtue)	Degradation (PurityVice)	Behavioral immune system versus spontaneous reaction

Table 1: Principles of Moral Foundations Theory (MFT)

captures people's moral reactions and categorizes human behavior into five basic principles that are characterized by opposing values (virtues and vices); fairness/cheating, care/harm, authority/subversion, loyalty/betrayal, and purity/degradation (as shown in Table 1).

In this study, we analyze text data to identify the type and magnitude of moral values that people refer to when expressing their opinion (stance) about a social issue. In brief, MFT, which is discussed in more detail in §2, postulates that several innate and universal psychological features are the foundation of "intuitive ethics." Each culture then constructs narratives, norms, and institutions that are influenced by these foundations, thereby creating the unique value systems we see around the world and that may sometimes trigger conflicts among groups [18].

To further investigate our hypothesis, we use a standard, publicly available dataset for stance classification [36], which contains tweets labeled as being in favor or against six social issues (abortion, climate change, feminism, Donald Trump, Hilary Clinton, and atheism). The Moral Foundation Dictionary (MFD) [18] operationalizes the MFT and can help in capturing morality from texts. In this paper, we leverage an expanded version of the MFD [43, 45] that expands the original lexicon in size and scope. We applied the expanded dictionary to the tweet corpus to measure and understand basic differences between the opposing sides, regardless of political orientation, which is often the focus of stance detection [14]. After categorizing each social issue with respect to morality types, we extract aspects (the most salient terms) per morality type from both sides of the discussion (in favor and against). Figure 1 summarizes the workflow of this study.

Overall, we aim to address the following research questions in this paper:

- **RQ1:** What basic morality types are contained in tweets about social issues?
- **RQ2:** What are the characteristics associated with each morality dimension, given there are opposing sides (known as stance) when discussing a social issue?
- **RQ3:** What are the correlations between each morality dimension and stance?

Our work makes the following contributions: We enhance the status quo of knowledge about the application of the MFT to empirical data about controversial issues of general interest and independent of political orientation. We believe that the MFT is a suitable framework for examining morality and stance as it establishes the correlation between five fundamental moral foundations (care, fairness, ingroup, authority, purity) and moral behaviors (showing support/against a certain issue) [18]. Secondly, we advance psycholinguistic understanding of how individuals' personal beliefs and values such as morality and stance can be manifested through language and discourse [27]. Moreover, we expand the scope of stance analysis, which traditionally focuses on identifying binary polarization in discussions, by examining the narratives on either side of the topic in more depth; identifying patterns that describe moral foundations across topics. Finally, characterizing each moral foundation via aspects (key terms) from empirical data provides a window into individual values that are used when discussing controversial social issues.

2 RELATED WORK

Stance has been defined as the overall position of a person towards an idea, object, or proposition [6, 49]. Extant literature has studied the relationship between morality and stance through the lens of "moral politics" [30] and moral "rhetoric" [15, 48] surrounding political issues such as presidential debates [35]. Traditionally, moral politics and rhetoric were examined using discourse analytic methods such as critical narrative analysis [47, 50]. For instance, Rymes [47] used critical narrative analysis to examine how at-risk youths assert their moral stance and "moral agency" towards violence through narrative and grammatical techniques. Though discourse analyses give thorough attention to how language and discourse elements are used to convey an individual's or group's moral stance, they are often based on smaller quantities of text data collected in specific social contexts. With these limitations in mind, we aim to leverage a larger-scale corpora of text data that covers social topics to investigate the narrative and linguistic features grounded in language on individuals' moral values and stances towards contemporary social issues.

One of the first empirical studies related to stance classification dealt with perspective identification. Lin, Wilson, Wiebe, and Hauptman [32] leveraged Bitter-Lemons' articles on Palestine-Israel conflict to automatically detect people's perspective regarding that issue. Hoover and colleagues [23] used linear SVM (Support Vector

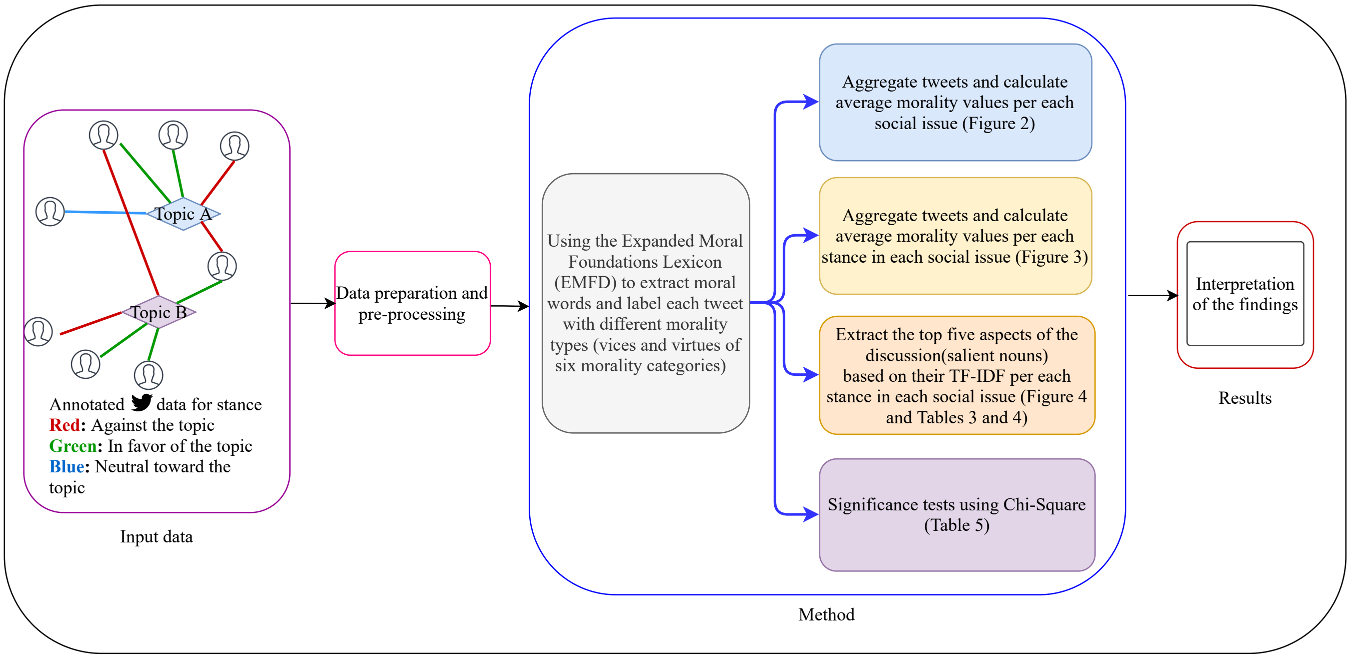

Figure 1: Project workflow and experimental design

Machines) to classify 35,108 tweets into "moral sentiments" (positive or negative) towards each of the five moral dimensions listed in the MFD. Similarly, Somasundaran and Wiebe [49] used a lexicon-based approach to identify arguments and sentiment in texts (on abortion, creationism, gun rights, and gay rights), and used these two features to classify stance. Both studies [23, 49] asserted that sentiment is a reliable indicator of an individual's perspective towards a social issue. Anand and colleagues [3] leveraged word-level features such as n-grams and syntactic dependencies to predict stance in debates. Li and Caragea [31] leveraged an existing sentiment lexicon [25] to predict stance by incorporating the lexicon words in the attention layer of a bidirectional LSTM.

Mohammad and colleagues [36] introduced the SemEval 2016 shared task for stance detection (they had an inter-annotator agreement of 73.1%). Using Twitter as a source, they released a baseline model and dataset for analyzing stance in user-generated texts. Elfardy and Diab [10] analyzed the SemEval dataset by leveraging perspective detection, where they used frame and semantic analysis as well as textual information such as sentiment and Linguistic Inquiry and Word Count (LIWC) as features to predict stance. Recent work by Zhang and colleagues [57] used the SemEval dataset to train a bidirectional LSTM that incorporates semantic and emotional valences as additional features to predict the stance of tweets. Aldayel and Magdy [2] leveraged social (media) network properties, such as a user interactions and connections, to study stance. Popat and colleagues [42] leveraged BERT (Bidirectional Encoder Representations from Transformers) to study stance, and showed the efficiency of their approach as it increased the state-of-the-art around 2-3%.

While prior research on stance detection has thoroughly investigated different linguistic and non-linguistic features to categorize stance, there is limited work on leveraging moral values as indicators of stance in text data. Baly and colleagues [5] used MFD as one of seven features to predict factuality of news articles (whether an article is unbiased or biased, fake or real). Johnson, Lee, and Goldwasser [28] conceptualized morality as a frame of reference that politicians take to express their stance towards six political issues on Twitter. Ferreira and Vlachos [13] used moral values from the MFD as one of several lexical features to train a multi-label stance classifier that predicts either a presence or absence of stance in a tweet. Their classifier with moral values incorporated yielded a 12% higher performance compared to the baseline model.

Prior research has also shown that moral values can be observed in language through the notion of stance [5, 27, 45]. Moreover, [45] studied the impact of using MFD as an additional (NLP) feature in predicting social effects such as stance, and showed that both classical feature-based (93%) and deep learning (85.7%) machine learning models benefited from moral words in the majority of test cases. However, in-depth knowledge about the relationship between morality and stance is still underexplored. Hence, in this study we further investigate the effect that moral values may have on opinion formation and expression, i.e., stance related to six social issues.

3 DATA

We use a previously annotated dataset for stance detection that was made available for SemEval 2016 Task 6 [36]. This dataset contains 4,870 tweets in total that were manually annotated for stance (in favor, against, or none (Table 2)). The SemEval 2016 dataset was used in a shared task to determine users' stance in tweets on six selected topics (abortion, atheism, climate change, feminism, Donald Trump, and Hillary Clinton). For the purpose of

	Abortion	Atheism	Climate Change	Feminism	Hillary Clinton	Donald Trump
Favor	167	124	335	268	163	148
Against	544	464	26	511	565	299
None	222	145	203	170	256	260

Table 2: Number of tweets in the SemEval-2016 Stance dataset labeled in favor or against of six social issues

this study, we only considered using tweets labeled as in favor or against, excluding the tweets labeled as none.

4 METHOD

4.1 Expanded Moral Foundations Dictionary

To analyze the relationship between stance and morality and extract the relevant words, we leveraged the "Expanded Morality Lexicon" (EMFD), an enhanced version of the MFD [43, 45]. The original MFD associates 324 unique words with the virtues and vices of five morality dimensions, as shown in Table 1. Moreover, the MFD contains a 6th category, called "General", that consists of words indicative of morality, which were not assigned to any specific morality dimension. While the MFD is a theoretically motivated starting point for studying morality, it contains only a few indicator terms, and lacks additional disambiguating information, such as parts of speech (POS) per token. For example, the term *"safe"*, which is listed in the MFD, is relevant when occurring as an adjective, but does not entail any morality weight and hence is not relevant when occurring as a noun.

To address these issues, we leveraged the EMFD developed by [43, 45]. This lexicon divides the "General" dimension into virtue and vice, the same as the other five categories, and leverages a combination of WordNet [11] and human-centered evaluation to expand each word in the original morality lexicon and evaluate all original and added entries.

As a result, the EMFD consists of six morality dimensions (care, fairness, ingroup, authority, purity, and general), 12 morality types (vice and virtues of each dimension; care (carevirtue), harm (carevice), fairness (fairnessvirtue), cheating (fairnessvice), authority (authoriyvirtue), subversion (authorityvice), loyalty (ingroupvirtue), betrayal (ingroupvice), purity (purityvirtue), degradation (purityvice), general-virtue, and general-vice), two polarities (virtue and vice), and 4,636 words with each word tagged with their part of speech and morality type.

4.2 Morality Across Social Issues

To extract morality from the tweet corpus, we first preprocessed the data by converting all words to lower case, removing usernames and URLs, symbols, numbers, punctuations, and additional whitespace, and truncated repetitions of the same letters to two consecutive occurrences). We then used *NLTK* [7] to tokenize the tweets and tag each token with its respective POS. We then searched the preprocessed texts for the terms listed in the EMFD. If a term in text and its POS coincided with a lexicon entry and its POS, we considered the term for our analysis and labeled the word with its respective morality type per tweet.

We then clustered the tweets based on social issues and analyzed the differences and similarities between these social issues with respect to their average moral values (Figure 2). We further grouped the tweets based on their labeled stances to further investigate the relationship between stance, social issues, and morality (Figures 3).

4.3 Extracting Aspects based on Morality

When discussing an issue or topic, people mostly refer to various aspects of that issue to better position their opinion with respect to their stance. For instance, to discuss the topic of "abortion", tweets expressing favor towards this topic may discuss *women's right* while tweets against it may talk about *intention to harm*. Based on this assertion, we investigated the potential relationship between aspects related to each social issue and morality. According to [19, 33, 55, 56], nouns are key factors for representing aspects or topics in texts. Following this, we extracted the top 50 nouns, including hashtags, from each social issue with respect to the *TF*IDF* score of the nouns.

To identify the aspects related to each morality type, we first clustered the tweets based on morality types (resulting in 12 clusters for each social issue) and averaged the *TF*IDF* score of each extracted aspect (noun) across these clusters. We then selected the top five aspects (if applicable), per morality type and social issue. Figure 4 visualizes the top aspects across morality types and social issues. In addition, Table 3 and Table 4 list the top aspects.

4.4 Significance Testing

Given that our variables are categorical in nature, chi-square (χ^2) tests of associations were performed on each of the six social issue. We examined a series of correlations between stance and (1) different morality dimensions (number of words per tweet that match the five morality dimensions plus general category, as shown in Table 1), (2) different morality types (number of words per tweet that match words in the vice or virtue category of each morality dimension), and (3) morality polarities (number of words per tweet that match any virtue or vice category regardless of the morality dimension or type). We set our confidence level for statistical significance for considering any pairs of correlations to 95% ($p = 0.05$).

5 RESULTS

5.1 Morality Across Social Issues

To understand how moral foundations are manifested in tweets, we identified the occurrence of 12 morality types in each of the six social issues by applying the expanded MFD to the tweets (§4.2). Figure 2 visualizes the average morality across six social issues. We observe that social issues entail a distinctive distribution of

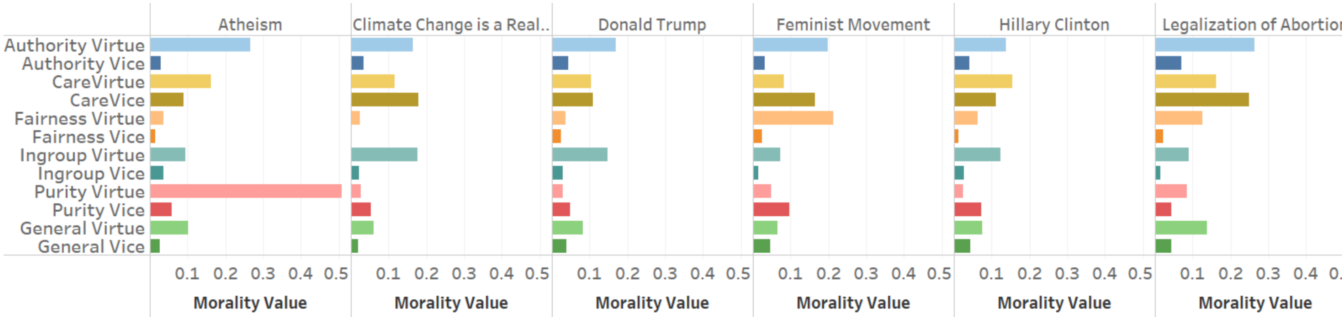

Figure 2: Average morality values across each social issue

morality types. While the highest value of fairness was found in the topic of feminism, atheism contains the highest morality value of purity. Moreover, harm and authority are more prevalent in the tweets related to the topic of abortion, and authority is among the top morality types in all social issues.

Furthermore, tweets mentioning Donald Trump have higher value of loyalty and authority compared to tweets mentioning Hillary Clinton. On the other hand, tweets mentioning Hillary Clinton have higher morality values of care compared to the ones discussing Trump. In addition, the topic of climate change features higher value of loyalty and is the only social issue in our dataset that has no morality value of cheating. This may have happened as a result of sparsity in our dataset, and asks for a more detailed investigation in the future using additional datasets to confirm or reject this finding.

As shown in Figure 3, we also investigated similarities and differences between morality types with respect to stance and social issues. Similar to the previous analysis, the results show different profiles of morality types when considering stance. For instance, discussions against the topic of climate change do not indicate care while it has a high value of harm. Similarly, discussions against the topic of abortion and feminism demonstrate higher values of harm compared to discussions in favor of this topic. Tweets in favor of feminism also contain higher value of fairness compared to those against it. Moreover, discussions against the topic of Hillary Clinton project higher value of harm and those written in favor of her consist of higher value of care and fairness. In contrast, tweets written in favor of the topic of Donald Trump consists of higher values of harm, care, as well as authority and loyalty.

Our analysis highlights the importance of considering morality when studying stance. The differences that we observed in various viewpoints (stances) highlight linguistic differences in discourse and can assist in analyzing cultural values and biases in society.

5.2 Extracting Aspects based on Morality

To further explore the connections between morality and aspects, we extracted aspects of the discussions for each morality type and social issue. Figure 4 visualizes the top five topics and their connected moral dimensions using a word network graph. In this network, nodes represent (12) morality types as well as top extracted aspects (listed in Table 3 and Table 4). The connection between a term and a morality type is represented by the edges in our network. The weight of the edge represents the average *TF*IDF* value of the aspects, while the colors of the edges represent social issues and stances.

Our results show that tweets written against the topic of atheism reference quotes and verses from the Bible and other holy books. For instance, for the word "lamb" in care (Tables 3 and 4), people bring quotes such as *I am washed and cleansed by the blood of the Lamb -Rev. 1:5; 7:14*[1] and for "acts" tweets include verses such as *Jesus commands you to follow Acts 2:38-39 to be saved.* On the other hand, tweets in favor of atheism discuss "country" and its rules and "establishment clauses"; i.e., *The establishment clause sets our country apart and prevents the radical religious zealots from taking charge.* Moreover, our findings show that the topic of atheism is associated with words related to purity and sanctity. Based on the MFT, this dimension is inspired by the notion of living "in an elevated, less carnal, more noble way" [18].

Furthermore, context analysis of the tweets against the idea of climate change shows that they include aspects such as *tooth* "fairy", "fraud", and *flawed* "computer" *models* showing that climate change may have been perceived as a *hoax*. Those concerned about climate change include aspects such as "disaster", "government", "extinction", and "wildfire" showing concerns about *human and species extinction*. In addition, we found that people in this group discuss events such as "Paris climate change" and "climate summit of the Americas (csota)" and show their support and interest regarding the event and topics discussed. Moreover, discussions of climate change feature both high loyalty and authority. Based on the MFT, loyalty is active and high "anytime people feel that it's one for all, and all for one" [18]. Also, aspects such as "countries", "homes", "communities" are frequent in tweets related to loyalty. Authority, on the other hand, refers to "virtues of leadership and followership, including deference to legitimate authority and respect for traditions" [18]. The usage of aspects such as "government", "authority", "wakeupamerica" may have resulted in the higher ratio of authority in this topic.

Additionally, tweets labeled as against the topic of Hillary Clinton discuss aspects such as "scandal", her daughter "Chelsea", her role in Clinton's "foundation", as well as some African-American ("black") voters who were *not supporting her in the election.* Tweets written in favor of her contain aspects such as "justice", "pride", "faith". Also, in terms of stance, tweets labeled as in favor of the

[1] *Italicized texts* in this section represent parts of the contexts.

Figure 3: Average morality values of each social issue with respect to stance

topic of Hillary Clinton used less negative and vulgar words related to harm and more positive words related to authority and care. On the other hand, tweets in favor of Donald Trump include more aspects related to loyalty, highlighting the likelihood of commitment to this figure. Moreover, the results show that aspects such as "racist", "China", and "border" are more prevalent in discussions against him while those in favor include aspects such as "Trump brigade", "RNC", and "honesty".

For the topics of feminism and abortion, we observe a high ratio of fairness. Also, the results show that tweets labeled as against feminism and abortion are concerned with lack of fairness compared to those in favor of them. Furthermore, tweets showing support for feminism discuss *breaking the* "glass" *ceiling* and the fact that they *refuse to accept that there is an unbreakable glass ceiling*. For the topic of abortion, we observe more aspects related to both care and harm. For instance, tweets against the topic of abortion talk about "pain", "love", and "pain" while those in favor of it use aspects such as "care", "babies", and "women's right".

5.3 Significance Tests

Chi-square procedures yield significant associations between stance and morality dimensions, types, and polarities in three social issues; abortion (n = 711), atheism (n = 588), and Hillary Clinton (n = 728). As a result of this observation, we excluded the other three social issues (climate change, Donald Trump, and feminism) from the results presented in Table 5. Our analysis shows that most number of associations are for the topic of abortion (n = 711), with significant relationships found between stance and morality types of harm, subversion, purity, and general-virtue, as well as stance and morality dimensions of purity, and general (as shown in Table 5). A similar number of associations were also found on the topic of Hillary Clinton, with stance having significant relationships with morality types of harm, fairness, and betrayal, as well as morality dimension of fairness, and both morality polarities (all virtues and vices). Stances on the topic of atheism have significant relationships with morality types of purity and general-vice. To further test for the strengths of association between these pairs, *Lambda* test for association between nominal variables were conducted. We did not find any significance for the *Lambda* tests. The results in mind, we believe that morality types and dimensions can be considered as one of the features that contribute to predicting stance, but they may not be the only variables with full explanatory power.

6 DISCUSSION AND CONCLUSION

In this paper, we performed a theory-driven and vocabulary-controlled detection of moral foundations from text data to expand the knowledge about stance analysis. Moral foundations can capture the influence of personal values and cultural differences on polarized or controversial discussions. Using a standard stance dataset, we applied an expanded version of a moral foundation lexicon [45] to analyze people's discussions on six distinct social issues. The main objective of this task is to expand the scope of stance analysis by examining the narratives on either side of the topic in more depth.

Our first research question asks what basic morality types are contained in tweets about social issues? To answer this, we identified basic morality types contained in the sample of tweets we obtained on six different social issues, namely abortion, atheism, climate change, Hilary Clinton, Donald Trump, and feminism. Our

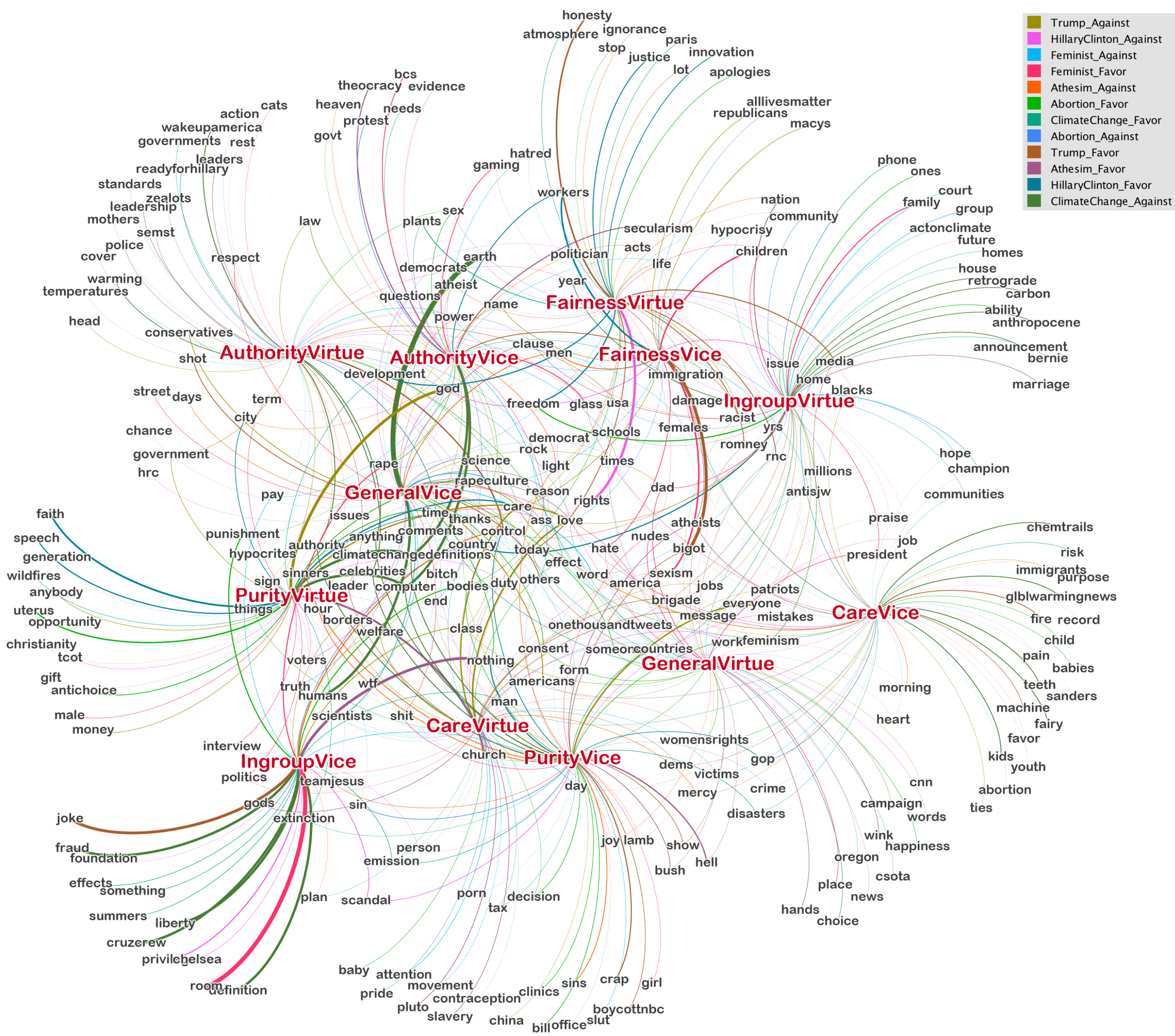

Figure 4: Network of top aspects and their connection to the 12 morality types (color=stance + aspect, thickness of the edge= weight of the word)

results (Figure 2 and Figure 3) show that various social issues have distinctive distribution of morality types. While some project more authority related words (Donald Trump), others consists of words related to care and purity (abortion and feminism). This finding suggests that social issues and individuals' stances on them are not morally equivalent. Additionally, our findings slightly confirm prior studies that found liberals' (e.g. Hillary Clinton) moral judgments more aligned with fairness and care [22, 51]. This finding is further exemplified in Stewart and Morris [51]'s study that found liberals' tended to exhibit "individualizing" moral foundations such as fairness and harm/care, while conservatives embraced group-based, "binding" foundations such as ingroup and authority. For the issue of climate change, some prominent morality types we observe such as authority, loyalty, harm are also found in other studies [9, 54] as important moral concerns regarding environmental conservation. However, our finding that climate change discussions score low on fairness and cheating is in contrast with the existing evidence that fairness is a salient indicator of attitude towards climate change [9]. One possible reason for this difference may be that prior studies often included political orientation as either a moderating [54] or mediating variable [8, 53] of moral types and stances found in the discussion of climate change. However, due to the sparsity in our dataset, more detailed investigation is needed to confirm or reject these findings.

	Atheism	Climate Change	Hillary Clinton	Feminist	Donald Trump	Abortion
CareVirtue	mercy, lamb, joy, sign, person		duty, rights, voters, work, berniesanders	attention, movement, victims, porn, USA	love, welfare, care, borders, jobs	love, truth, humans, end, contraception
CareVice	morning, America, mercy, favor, purpose	fairy, teeth, kids, glblwarmingnew, chemtrails	duty, champion, sanders, record, issue	victims, hope, word, females, abortion	blacks, immigrants, ties, care, patriots	times, pain, crime,youth, heart
FairnessVirtue	acts, nation, ignorance, atheist, light		rights, media, blacks, questions, yrs	god, USA, hypocrisy, reason, hatred	care, life, immigration, lot, racist	rape, sex, community, control, democrats
FairnessVice	men, acts, nation, ignorance		issue, god	power, hypocrisy, rapeculture, antisjw, sexism	racist, macys, name, thanks, republicans	millions, alllivesmatter, love
IngroupVirtue	nation, men, secularism, acts, praise	announcement, carbon, retrograde, year, anthropocene	politician, work, yrs, job, mistakes	group, court, hope, hypocrisy, word	patriots, love, jobs, life, everyone	country, family, community, USA, control
IngroupVice	control, plan	liberty, Cruzcrew, fraud, definition	privilege, scandal, Chelsea, foundation, person	pay, day	class, borders, welfare	country
AuthorityVirtue	power, hour, sinners, control, sign	celebrities, authority, anything, wakeupamerica, warming	duty, politician, questions, democrat, rest	power, standards, cover, pay, hypocrites	law, conservatives, government, head, god	mothers, control, government, men, USA
AuthorityVice	light, acts, city, atheist, heaven	climatechangede-initions, computer	evidence, comments, issues, bitch, term	others, hatred, pay, females, life	law, class, borders, welfare, name	protest, times, end, rape, sex
PurityVirtue	sinners, hour, control, teamjesus, christianity		god, duty, voters, tcot	god, word, opportunity, today, wtf	god, love. government, anybody, money	sin, man, freedom, gift, generation
PurityVice	sinners, sins, hour, light, teamjesus	climatechangede-initions, computer, celebrities, authority, anything	scandal, mistakes, person, bitch, comments	slut, day, porn, word, today	patriots, everyone, China, thanks, end	sin, everyone, others, crime, control
GeneralVirtue	joy, lamb, science, morning, praise		democrat, job, news, words, yrs	antisjw, form, USA, work, rapeculture	class, racist, today, CNN, Americans	heart, everyone, millions, others, control
GeneralVice	men, control, days, atheist, power	earth, humans	voters, questions, anything, hrc, god	word, god, wtf, form, rapeculture	Americans, today, conservatives, dad, name	truth, times, end, democrats, chance

Table 3: Top five extracted aspects (if applicable) from tweets tagged as "against" each social issue with respect to their morality type

Our second research question focuses on examining different characteristics associated with each of the five morality dimensions plus general category, given there are two separate stances on each social issue. Referring to the Results section above (§5), we find that each social issue and stance has distinctive lexical profiles, with different aspects representing the prominent discussions surrounding each issue (Figure 3, and Tables 3 and 4). For instance, tweets that stood in favor of atheism contain more discussions on purity and authority, while tweets in support of Donald Trump are more related to loyalty and authority. It is interesting to note

	Atheism	Climate Change	Hillary Clinton	Feminist	Donald Trump	Abortion
CareVirtue	slavery, love, gods, church, county	tax, countries, plan, scientists, emission	care, pride, love, country, control	work, politics, womensrights, interview, sexism	brigade, onethousandtweets, message, today, dems	care, effect, baby, decision, things
CareVice	atheists, love, hate, nothing	risk, damage, disasters, man, communities	GOP, home	glass, hate, sexism, day, ass	fire, machine, media, dems, immigration	care, effect, babies, child, womensrights
FairnessVirtue	atheists	plants, Paris, development, atmosphere	rights, justice, development, freedom, innovation	children, nudes, earth, gaming, home	honesty, today, RNC, Romney, media	apologies, freedom, bodies, stop, issue
FairnessVice			workers	children, sexism, hate	bigot	
IngroupVirtue	marriage, clause, nation, country, schools	countries, homes, communities, house, actonclimate	home, country, phone, champion, Bernie	home, family, dad, children, president	brigade, onethousandtweets, message, control, future	freedom, ones, Issue, ability, feminism
IngroupVice	nothing	effects, emission, summers, man, something		room, rape	joke	punishment, bodies, control
AuthorityVirtue	leaders, country, clause, zealots, schools	governments, temperatures, science, plants, action	time, control, readyforhillary, rock, love	respect, glass, year, cats, anyone	leader, leadership, control, police, USA	control, effect, care, sex, things
AuthorityVice	secularism, theocracy	needs, plants, damage	workers	glass, bcs, gaming, ass	immigration	rapeculture, punishment, govt, rape, sex
PurityVirtue	nothing, gods, reason, church, hate	development, extinction, wildfires	faith, care, speech, city, thanks	politics, ass, term, male, interview		church, uterus, rapeculture, antichoice, time
PurityVice	hell, bush, everyone, atheists	emission, disasters, countries, extinction, tax	nothing, control, GOP	shit, nudes, girl, everyone, office	crap, bigot, show, leader, boycottnbc	bill, decision, feminism, bodies, clinics
GeneralVirtue	hands, wink, hell, Bush, schools	reason, csota, oregon, countries, science	choice, time, rock, love, work	nudes, dad, president, place, hate	show, today, campaign, RNC, Romney	consent, happiness, someone, womensrights, church
GeneralVice	love, clause, reason, science	scientists		shit, ass, respect, America, street	shot, others	nothing, rapeculture, consent, someone, care

Table 4: Top five extracted aspects (if applicable) from tweets tagged as "in favor" of each social issue with respect to their morality type

that while tweets in support of atheism, Donald Trump, and Hillary Clinton have different lexical profiles, they allude to the importance of group-based moral values (e.g. purity and authority) that are essential in the discussions of religion [17] and politics [51].

In general, we see more instances of negative polarities (i.e. vice) in most morality dimensions, namely care, authority, ingroup, which was also supported by chi-square analyses. We assume that the discussions surrounding controversial issues on social media, especially on Twitter, are often polarized to either positive or negative sentiments [12, 39]. Specifically, for several controversial issues in our analysis such as those related to politics (i.e. Hillary Clinton and Donald Trump) or those related to climate change, abortion, and atheism which all have longstanding debates, we expect to see

		Abortion (N=711)	**Atheism** (N=588)	**Clinton** (N=728)
Morality Types	CareVirtue (Care)	-	-	-
	CareVice (Harm)	X2(3) = 10.99, p = 0.012**	-	X2(2) = 6.16, p = 0.046*
	Authority Virtue (Authority)	-	-	-
	AuthorityVice (Subversion)	X2(2) = 11.04, p = 0.004***	-	
	FairnessVirtue (Fairness)	-	-	X2(4) = 13.45, p = 0.009**
	FairnessVice (Cheating)	-	-	-
	IngroupVirtue (Loyalty)	-	-	-
	IngroupVice (Betrayal)	-	-	X2(1) = 5.93, p = 0.015**
	PurityVirtue (Purity)	X2(2) = 10.57, p = 0.005***	X2(3) = 8.06, p = 0.045*	-
	PurityVice (Degradation)	-	-	
	GeneralVirtue	X2(3) = 15.11, p = 0.002***	-	-
	GeneralVice	-	X2(1) = 5.08, p = 0.024*	-
Morality Dimension	Care (Care/Harm)	-	-	-
	Authority (Authority/Subversion)	-	-	-
	Fairness (Fairness/Cheating)	X2(2) = 6.84, p = 0.033*	-	-
	Purity (Purity/Degradation)	-	-	X2(4) = 13.49, p = 0.009**
	Ingroup (Loyalty/Betrayal)	-	-	-
	General (GeneralVirtue/GeneralVice)	X2(3) = 0.04, p = 0.045*	-	-
Morality Polarity	Virtue (All virtues)	-	-	X2(5) = 12.93, p = 0.024*
	Vice (All Vices)	-	-	X2(4) = 22.52, p = 0.000****

Table 5: Result of significance tests using chi-square (χ^2) ($p = 0.05$)

distinctive characteristics in the discussions on opposing sides of each social issue.

Our third research question focuses on the correlations between morality and stances of each tweet according to the social issues. Furthermore, our chi-square tests of associations (Table 5) find variances in the number of statistically significant relationships on morality and stance across different social topics. For instance, we find six significant correlations on the topic of abortion, six correlations on Hillary Clinton, two correlations on atheism, and none for Donald Trump, climate change, and feminism. Among the significant associations we found, we observe stance to be most correlated with *vice* morality type on various dimensions such as harm, subversion, general-vice, and betrayal. A post-hoc analysis using *Lambda* test to find direction of associations was performed, however it does not yield significant relationships. We hypothesize that more words that falls under *vice* spectrum would correlate with the increase in against stance, however the results does not reflect this. There are several reasons to suspect that we should perform further analyses to explicate all correlation pairs. Firstly, each social topic has a different sample of tweets, as well as some topics may be more controversial comparing to others (i.e. the topic of abortion compared to Donald Trump). In the future, we hope to perform further aspect analysis to further examine cross-cutting communication as well as the conceptual complexities within each social issue.

Our work is limited in several aspects. First, our work follows the standard model of stances being a binary problem (in favor or against), while on certain issues, there might be more than these two points of views. Additionally, negation detection was not considered in this study, which might influence the amount of moral loading for contrasting polarities. We are also aware that there are state-of-the-art expansions of the MFD [4, 24], and we will compare the utility of our extended lexicon to other versions of the MFD in future work. Finally, we recognize the constraints and sparsity of our data sample, focusing only on US-related social issues and English-only language use. We recognize prior studies that have found that moral values and stances towards social issues significantly vary across cultures ([40]'s study with Swedish individuals found prominent concerns with fairness and harm; [29]'s study with Korean individuals found emphasis on purity values). In the future, we hope to expand the study of morality and stance towards more issues from a more diverse set of social contexts and cultures.

ACKNOWLEDGMENTS

We thank Ming Jiang and Aseel Addawood for their assistance in the first part of this study.

REFERENCES

[1] Aseel Addawood, Rezvaneh Rezapour, Omid Abdar, and Jana Diesner. 2017. Telling apart tweets associated with controversial versus non-controversial topics. In *Proceedings of the Second Workshop on NLP and Computational Social Science*. 32–41.

[2] Abeer Aldayel and Walid Magdy. 2019. Your stance is exposed! Analysing possible factors for stance detection on social media. *Proceedings of the 22nd ACM Conference on Computer Supported Cooperative Work & Social Computing* 3 (2019), 1–20.

[3] Pranav Anand, Marilyn Walker, Rob Abbott, Jean E Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. Association for Computational Linguistics, 1–9.

[4] Oscar Araque, Lorenzo Gatti, and Kyriaki Kalimeri. 2020. MoralStrength: Exploiting a moral lexicon and embedding similarity for moral foundations prediction. *Knowledge-based Systems* 191 (2020), 105184.

[5] Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2018. Predicting factuality of reporting and bias of news media sources. *arXiv preprint arXiv:1810.01765* (2018).

[6] Douglas Biber and Edward Finegan. 1988. Adverbial stance types in English. *Discourse Processes* 11, 1 (1988), 1–34.

[7] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.

[8] Sharon L Dawson and Graham A Tyson. 2012. Will morality or political ideology determine attitudes to climate change. *Australian Community Psychologist* 24, 2 (2012), 8–25.

[9] Janis L Dickinson, Poppy McLeod, Robert Bloomfield, and Shorna Allred. 2016. Which moral foundations predict willingness to make lifestyle changes to avert climate change in the USA? *PloS One* 11, 10 (2016), e0163852.

[10] Heba Elfardy and Mona Diab. 2016. CU-GWU Perspective at SemEval-2016 Task 6: Ideological Stance Detection in Informal Text. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, 434–439.

[11] Christiane Fellbaum. 2010. WordNet. In *Theory and Applications of Ontology: Computer Applications*. Springer, 231–243.

[12] Emilio Ferrara and Zeyao Yang. 2015. Measuring emotional contagion in social media. *PloS One* 10, 11 (2015), e0142390.

[13] William Ferreira and Andreas Vlachos. 2019. Incorporating label dependencies in multilabel stance detection. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 6350–6354.

[14] Dean Fulgoni, Jordan Carpenter, Lyle H Ungar, and Daniel Preotiuc-Pietro. 2016. An empirical exploration of Moral Foundations Theory in partisan news sources. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*.

[15] Justin Garten, Reihane Boghrati, Joe Hoover, Kate M Johnson, and Morteza Dehghani. 2016. Morality between the lines: Detecting moral sentiment in text. In *Proceedings of IJCAI 2016 Workshop on Computational Modeling of Attitudes*.

[16] Homero Gil de Zúñiga, Nakwon Jung, and Sebastián Valenzuela. 2012. Social media use for news and individuals' social capital, civic engagement and political participation. *Journal of Computer-Mediated Communication* 17, 3 (2012), 319–336.

[17] Jesse Graham and Jonathan Haidt. 2010. Beyond beliefs: Religions bind individuals into moral communities. *Personality and Social Psychology Review* 14, 1 (2010), 140–150.

[18] Jesse Graham, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P Wojcik, and Peter H Ditto. 2013. Moral Foundations Theory: The pragmatic validity of moral pluralism. In *Advances in Experimental Social Psychology*. Vol. 47. Elsevier, 55–130.

[19] Vishal Gupta and Gurpreet Singh Lehal. 2010. A survey of text summarization extractive techniques. *Journal of emerging technologies in web intelligence* 2, 3 (2010), 258–268.

[20] Jonathan Haidt. 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review* 108, 4 (2001), 814.

[21] Jonathan Haidt, Craig Joseph, et al. 2007. The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. *The Innate Mind* 3 (2007), 367–391.

[22] Wilhelm Hofmann, Daniel C Wisneski, Mark J Brandt, and Linda J Skitka. 2014. Morality in everyday life. *Science* 345, 6202 (2014), 1340–1343.

[23] Joe Hoover, Gwenyth Portillo-Wightman, Leigh Yeh, Shreya Havaldar, Aida Mostafazadeh Davani, Ying Lin, Brendan Kennedy, Mohammad Atari, Zahra Kamel, Madelyn Mendlen, et al. 2020. Moral Foundations Twitter Corpus: A collection of 35k tweets annotated for moral sentiment. *Social Psychological and Personality Science* 11, 8 (2020), 1057–1071.

[24] Frederic R Hopp, Jacob T Fisher, Devin Cornell, Richard Huskey, and René Weber. 2021. The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods* 53, 1 (2021), 232–246.

[25] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the 10th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. 168–177.

[26] Johan Jansson and Ellen Dorrepaal. 2015. Personal norms for dealing with climate change: results from a survey using Moral Foundations Theory. *Sustainable Development* 23, 6 (2015), 381–395.

[27] Katarzyna Jaszczolt. 1999. *Discourse, beliefs, and intentions: Semantic defaults and propositional attitude ascription*. Elsevier.

[28] Kristen Johnson, I-Ta Lee, and Dan Goldwasser. 2017. Ideological phrase indicators for classification of political discourse framing on Twitter. In *Proceedings of the 2nd Workshop on NLP and Computational Social Science*. 90–99.

[29] Kisok R Kim, Je-Sang Kang, and Seongyi Yun. 2012. Moral intuitions and political orientation: Similarities and differences between South Korea and the United States. *Psychological Reports* 111, 1 (2012), 173–185.

[30] George Lakoff. 1995. Metaphor, morality, and politics, or, why conservatives have left liberals in the dust. *Social Research* (1995), 177–213.

[31] Yingjie Li and Cornelia Caragea. 2019. Multi-Task Stance Detection with Sentiment and Stance Lexicons. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China.

[32] Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, and Alexander Hauptmann. 2006. Which side are you on? Identifying perspectives at the document and sentence levels. In *Proceedings of the 10th Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 109–116.

[33] Bing Liu et al. 2010. Sentiment analysis and subjectivity. *Handbook of natural language processing* 2, 2010 (2010), 627–666.

[34] Gilad Lotan, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce, et al. 2011. The Arab Spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International Journal of Communication* 5 (2011), 31.

[35] Morgan Marietta. 2009. The absolutist advantage: sacred rhetoric in contemporary presidential debate. *Political Communication* 26, 4 (2009), 388–411.

[36] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. 31–41.

[37] Saif Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)* 17, 3 (2017), 26.

[38] Akiko Murakami and Rudy Raymond. 2010. Support or oppose?: classifying positions in online debates from reply activities and opinion expressions. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, 869–875.

[39] Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, and Kristina Miler. 2015. Tea party in the house: A hierarchical ideal point topic model and its application to republican legislators in the 112th congress. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Vol. 1. 1438–1448.

[40] Artur Nilsson and Arvid Erlandsson. 2015. The Moral Foundations taxonomy: Structural validity and relation to political ideology in Sweden. *Personality and Individual Differences* 76 (2015), 28–32.

[41] Marco Pennacchiotti and Ana-Maria Popescu. 2010. Detecting controversies in Twitter: a first study. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*. Association for Computational Linguistics, 31–32.

[42] Kashyap Popat, Subhabrata Mukherjee, Andrew Yates, and Gerhard Weikum. 2019. STANCY: Stance classification based on consistency cues. *arXiv Preprint arXiv:1910.06048* (2019).

[43] Rezvaneh Rezapour and Jana Diesner. 2019. Expanded Morality Lexicon. https://doi.org/10.13012/B2IDB-3805242_V1.1

[44] Rezvaneh Rezapour, Priscilla Ferronato, and Jana Diesner. 2019. How do moral values differ in tweets on social movements?. In *Conference Companion Publication*

of the 2019 on Computer Supported Cooperative Work and Social Computing. 347–351.

[45] Rezvaneh Rezapour, Saumil H Shah, and Jana Diesner. 2019. Enhancing the measurement of social effects by capturing morality. In *Proceedings of the 10th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. 35–45.

[46] Sonia Roccas and Marilynn B Brewer. 2002. Social identity complexity. *Personality and social psychology review* 6, 2 (2002), 88–106.

[47] Betsy Rymes. 1995. The construction of moral agency in the narratives of high-school drop-outs. *Discourse & Society* 6, 4 (1995), 495–516.

[48] Eyal Sagi and Morteza Dehghani. 2014. Measuring moral rhetoric in text. *Social Science Computer Review* 32, 2 (2014), 132–144.

[49] Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing Stances in Online Debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. Association for Computational Linguistics, Suntec, Singapore, 226–234.

[50] Mariana Souto-Manning. 2014. Critical narrative analysis: The interplay of critical discourse and narrative analyses. *International Journal of Qualitative Studies in Education* 27, 2 (2014), 159–180.

[51] Brandon D Stewart and David SM Morris. 2021. Moving morality beyond the in-group: liberals and conservatives show differences on group-framed Moral Foundations and these differences mediate the relationships to perceived bias and threat. *Frontiers in Psychology* 12 (2021).

[52] Harry C Triandis. 1989. The self and social behavior in differing cultural contexts. *Psychological Review* 96, 3 (1989), 506.

[53] Annukka Vainio and Jaana-Piia Mäkiniemi. 2016. How are moral foundations associated with climate-friendly consumption? *Journal of Agricultural and Environmental Ethics* 29, 2 (2016), 265–283.

[54] Christopher Wolsko, Hector Ariceaga, and Jesse Seiden. 2016. Red, white, and blue enough to be green: Effects of moral framing on climate change attitudes and conservation behaviors. *Journal of Experimental Social Psychology* 65 (2016), 7–19.

[55] Samira Zad, Maryam Heidari, H James Jr, and Ozlem Uzuner. 2021. Emotion detection of textual data: An interdisciplinary survey. In *2021 IEEE World AI IoT Congress (AIIoT)*. IEEE, 0255–0261.

[56] Samira Zad, Maryam Heidari, James H Jones, and Ozlem Uzuner. 2021. A survey on concept-level sentiment analysis techniques of textual data. In *2021 IEEE World AI IoT Congress (AIIoT)*. IEEE, 0285–0291.

[57] Bowen Zhang, Min Yang, Xutao Li, Yunming Ye, Xiaofei Xu, and Kuai Dai. 2020. Enhancing cross-target stance detection with transferable semantic-emotion knowledge. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 3188–3197.