Examining the influence of Moral Foundations on Polarization in Social Media Discourse: A context of vigilantism

Shalini Kapali Kurumathur
University of Texas at San Antonio, shalinikapali.kurumathur@my.utsa.edu

Rohit Valecha
University of Texas at San Antonio

Govind Hariharan
Kennesaw State University

H Raghav Rao
University of Texas at San Antonio

Examining the influence of Moral Foundations on Polarization in Social Media Discourse: A context of vigilantism

Shalini Kapali Kurumathur[1]
Alvarez College of Business,
The University of Texas at San Antonio,
San Antonio, Texas, USA

Rohit Valecha
Alvarez College of Business,
The University of Texas at San Antonio,
San Antonio, Texas, USA

Govind Hariharan
Coles College of Business,
Kennesaw State University,
Kennesaw, Georgia, USA

H. Raghav Rao
Alvarez College of Business,
The University of Texas at San Antonio,
San Antonio, Texas, USA

ABSTRACT

In August 2022, Kyle Rittenhouse's vigilantism in Kenosha, Wisconsin, resulted in two fatalities and one injury, sparking heightened public security concerns and eliciting intense moral reactions. These moral impulses tend to contribute to extreme judgments of right or wrong, thereby fostering polarization on social media. The phenomenon of polarization, recognized as a component of social cybersecurity, has recently gained attention. This study explores the impact of five Moral Foundations on polarization following vigilantism, utilizing moral foundation theory and vector autoregression (VAR) in the analysis of social media discourse. Our findings reveal that these Moral Foundations significantly influence polarization dynamics. This insight holds implications for both research and the development of practical strategies for managing the societal consequences of polarization on social media.

Keywords: social cybersecurity, social media discourse, polarization, vigilantism, econometrics

INTRODUCTION

In August 2022, Kyle Rittenhouse, in an incident of vigilantism, fatally shot two men and wounded the third in Kenosha, Wisconsin. Such vigilantism incidents heighten security alerts in public and can spark a lot of debates on social media. Vigilantism, defined as a violation of

---

[1] Corresponding author. shalinikapali.kurumathur@my.utsa.edu +1 612 516 3004

established socio-political order, can provoke extreme judgments, that may also result in increasing polarization where attitudes of users and their discussions are increasingly concentrated at opposing extremes. Such extreme judgements or attitude of right or wrong challenge the social cohesion in modern civil societies (Rapp 2016).

Highly polarized citizens may dismiss flaws of their own views and merits of the opponents (Heltzel and Laurin 2020). They would articulate their views in social media discourse. The resulting discourse results in what social cybersecurity literature refers to as "echo chamber" environments, where existing attitudes are continuously reinforced and can lead to increased polarization (Carley 2020). Social cybersecurity is a field that studies how to preserve internet as 'a free and open space for exchange of information' (Carley et al. 2018). It is particularly interested in how social media messages impact polarization (Carley 2020). Such polarization can lead to a fragmented society (Adamic and Glance 2005; Taber and Lodge 2006), which may initiate violence (Pearlman and Cunningham 2012).

Following (Garrett and Bankert 2020), we suggest that Moral foundations influence polarization in the online context. Moral Foundations are the innate (but modifiable) psychological mechanisms built on virtues, vices and moral values (Haidt and Graham 2007). Moral values correlate with psychological arousal in an individual, that would induce polarizing judgements (DeScioli and Kurzban 2009; Hetherington and Weiler 2009). These moral values are central to people's sense of identity and their perception of justice (Clayton and Opotow 2003; Miller 2001). People affirm their sense of self by possessing moral values when they identify themselves as part of a collective. i.e., the overall sense of justice in that collective is more important than one's individual perception (Clayton and Opotow 2003).

There are several research gaps in the extant literature on polarization. Much of the polarization literature has not considered morality as an issue. We argue that understanding of polarization is incomplete without accounting for morality because moral foundations can predict attitudes on social issues (Koleva et al. 2012; Stolerman and Lagnado 2020). Incidents of vigilantism provoke discourses related to moral values and may induce polarizing judgements on social media (Haas 2010). Examining the research on influence of moral foundations can help explain people's attitude toward vigilantism. Typically, vigilantism incidents result in the violation of social orders. This represents an insult to the integrity of various communities and can provoke the urge to punish the vigilante (Miller 2001). Harming a community member (victim) can arouse strong moralistic and punitive impulses (Miller 2001). Moreover, the literature on polarization which has dealt with morality has considered it to be a single construct (Garrett and Bankert 2020). We suggest that different Moral Foundations need to be treated separately. The five moral foundations can affect attitudes differently, thereby influencing what users write on social media (Day et al. 2014). Additionally, social media is updated at a high velocity, therefore an analysis of finer granularity would be critical. Moral foundations entail automatic gut-reactions of like and dislike when certain patterns are perceived in the social world, which in turn guide judgments expressed on social media (Koleva et al. 2012). The social world with news reports, trials proceedings on vigilantism has been changing rapidly, thereby influencing the Moral Foundation expressions on social media. Hence, it is important to understand how the relationship varies over time to reveal the temporal effect of the Moral Foundations on polarization.

This paper specifically poses the following research questions. 1. How do Moral Foundations influence polarization? 2. Do the different Moral Foundations influence polarization in the same

way? 3. Does the relationship between Moral Foundations and polarization differ over time? To examine these research questions, we propose to use Moral Foundations (MF)Theory from the social psychological literature (Graham et al. 2013, 2018; Haidt et al. 2009). We focus on a vigilantism incident regarding 'Rittenhouse'. Data was collected from the X platform (formerly known as Twitter) streaming API with developer accounts. The data was collected from the day the trials started through three days after the jury announced the verdict. We conducted human coding on a sample of random tweets to examine if the tweets expressed in favor or against vigilantism. We then used machine learning to predict the values of the remaining tweets. We identified MFs for the tweets using Linguistic Inquiry and Word Count (LIWC) and the polarization score was computed across all the tweets over time. We used time series analysis i.e., vector autoregression (VAR), which will help us to analyze the relationship of Moral Foundations and polarization at hourly frequency. This will provide an understanding of short-term and long-term impact of Moral Foundations on polarization.

This paper makes several important contributions. First, this study extends the social cyber-security literature by focusing on how and when (short-term and long-term impact) each of the Moral Foundations influence polarization after a vigilantism incident (Carley 2020). Second, our study helps to advance the polarization literature and provides implications to policy makers who wish to tackle the important issue of security in the digital age to those dealing with threat posed by polarization in the context of vigilantism.

RESEARCH BACKGROUND

Social cybersecurity

Social cybersecurity is defined as a 'multidisciplinary and multimethodological field that studies how to preserve internet as a free and open space for exchange of information' (Carley et

al. 2018; Uyheng et al. 2020). While cybersecurity is focused on machines, social cybersecurity is focused on humans. Social cybersecurity research involves to identify, counter, and measure the impact of communications (Carley 2020), specifically, focusing on understanding social communications. Social cybersecurity is focused on humans situated in society and how digital environment alters the community. For instance, some social media posts may manipulate group dynamics. i.e., they can isolate two groups because their beliefs clash with each other, creating polarization. Various incidents such as vigilantism can provoke social media expressions of moral foundations, which can influence expressions of extreme right or wrong judgements, thereby influencing polarization. This literature has largely neglected the context of vigilantism.

Vigilantism

Social media users have expressed their opinions in the context of extreme events such as Egyptian revolution (Venkatesan et al. 2021), Gaza conflict (Kwon et al. 2012), mass shooting (Demszky et al. 2019) and vigilantism (Kurumathur et al. 2022). Vigilantism is defined as "acts or threats of coercion in violation of the formal boundaries of an established sociopolitical order which, however, are intended by the violators to defend that order from some form of subversion" (Rosenbaum and Sederberg 1974). The vigilante may be a private agent who is not affiliated with law enforcement/government, or someone affiliated to law enforcement/ government. For any act to be one of 'vigilantism', there needs to be some planning or pre-meditation by those engaging in it (Johnston 1996).

The vigilantism incidents may lead concerned social media users to react to the changed environment because it is an insult to the integrity of the entire community (Miller 2001). When a community member (victim) is harmed, it can arouse strong moralistic and punitive impulses

in a social media user (Miller 2001). Thus, supporting the victim becomes a defense of honor to the public (Miller 2001). Such moralistic impulses can lead to polarization.

Polarization

Polarization is "a process whereby the normal multiplicity of differences in a society increasingly align along a single dimension, cross-cutting differences become instead reinforcing, and people increasingly perceive and describe politics and society in terms of "Us" versus "Them" (McCoy et al. 2018). Polarization has been studied in both political and non-political issues. It is found that there is greater polarization on political issues than non-political issues (Barberá et al. 2015). Researchers also suggest that the polarization on social media is due to echo chambers, where a person only encounters information or opinions that reflect and reinforce their own (Del Vicario et al. 2016).

Moral Foundations theory

In this paper, we argue that the polarization literature needs to give attention to the role of Moral Foundations after vigilantism incidents. Moral Foundations Theory is one of the popular theories in providing common language in the moral domain and has been used in different contexts such as same-sex marriage (Ball 1996), climate change (Landrum et al. 2016), U.S immigration policy debate (Grover et al. 2019). Moral Foundations is closely related to human rights principles, especially social security and equality (Stolerman and Lagnado 2020). According to moral foundations theory, all cultures build mainly on five universally available moral foundations (Haidt and Graham 2007). Each of the five moral foundations is theorized in terms of: Care, Fairness, Loyalty, Authority, and Sanctity. The theory claims that these moral foundations drive users instinctively and immediately react to any stimuli.

Care Moral Foundation

The Care Foundation evolved from the adaptive challenges faced due to the need to care for vulnerable, weak offspring for relatively long periods in relationships such as the mother-child relationship. (Graham et al. 2013). From such relationships, people eventually learned to care for others. The feeling of care for defenseless beings can transform into feelings of compassion for victims and anger at preparators of vigilantism. Similarly, care can be expressed towards the vigilante and anger towards the situation that made the vigilante to fight the victim.

Social media discourses with high Care Moral Foundation express dislike after observing victims harmed by vigilantes. The social media content may convey anger towards an aggressor or vigilante, leading to an extreme attitude. Due to the tendency to feel for vulnerable victims, we argue that social media discourse expressing Care Moral Foundation will exhibit polarization either containing attitudes of extreme anger or extreme sympathy.

There is also a possibility that social media discourse with high Care Moral Foundation can express dislike after seeing vigilantes being a victim of different circumstances. The social media content may convey anger towards different situations and care towards vigilantes (who are victims of circumstances), leading to an extreme attitude. Such extreme attitudes lead to polarization in social media discourse.

Fairness Moral Foundation

In a discourse, fairness, justice, and trustworthiness are the virtue words associated with this moral foundation (Graham et al. 2013). This Foundation is concerned with treating individuals fairly. For example, users may think that it is fair to preserve the rights of victim/aggressor after the recent vigilantism.

A social media discourse with high Fairness Moral Foundation will likely evaluate vigilantism incidents based on facts. Hence, the textual content will be neutral and unbiased in attitude. Thus, the social media discourses expressing Fairness Moral Foundation will not contain content expressions of extreme attitude. Thus, content with high Fairness Moral Foundation will not be polarized.

Loyalty Moral Foundation

Discourses prioritizing this foundation show the appreciation of the virtues of loyalty to the group an individual belongs to (Graham et al. 2013), for example, belonging to a particular race, religious group, political party. That is, loyalty to one's group is more important than one's individual concerns in a discourse. For instance, if a victim who is African American is injured, expressions of social media users of African American origin may convey texts of loyalty to the group. The group loyalty in the discourses would lead the expressions of social media user to have extreme attitudes. In other words, a social media user will convey either loyalty (towards one's group) or disloyalty discourses (towards other groups) regarding the Victim/aggressor depending on whether he/she is from the same group. Thus, expressions of Loyalty Moral Foundation lead to polarization in social media discourse.

Authority Moral Foundation

This foundation evolved in response to concerns related to social order and the obligations of hierarchical relationships, and relates to acts of obedience/ disobedience, respect/disrespect, and submission/rebellion, concerning legitimate authority (Graham et al. 2013). It makes individuals sensitive to symbols of authority and also makes them aware of whether others are following the social order according to law and order (Graham et al. 2013). This foundation is about commanding authority to fulfil its role of leading to maintain social

order as well as commending others' behavior of obeying the authority accordingly (Koleva et al. 2012).

A social media discourse with Authority Moral Foundation would display respect to the State. In the context of vigilantism, social media discourses may convey that the vigilante needs to follow the State and expect vigilantes to follow the established social order to maintain the same. A social media expression with high Authority Moral Foundation may convey extreme anger towards the vigilante for not following social order.

Sanctity Moral Foundation

This foundation largely results in feelings of disgust. For example, repulsion against homosexuals (Haidt et al. 2009) can be based on the sanctity moral foundation. This foundation is shaped by the psychology of disgust and contamination.

The sanctity foundation makes social media discourses avoid anything that is not traditional in nature and concerns order. For example, in the context of vigilantism, social media discourses have often suggested that life is sacred and should not be sacrificed by policies such as gun control laws. Because vigilantism destabilizes the sanctity of the social structure, social media discourses with high sanctity Moral Foundation will display expressions of extreme attitude on vigilantism incidents. Thus, expressions of Sanctity Moral Foundation will lead to polarization in social media discourse.

DATA COLLECTION AND ANALYSIS

We analyzed tweets collected from the X platform (formerly known as Twitter) streaming API with developer accounts. The data was collected using the Twitter streaming API with the help of Tweepy python library using keywords about Kyle Rittenhouse and hashtags associated with Kyle Rittenhouse trial such as #kylerittenhouse, #RittenhouseTrial,

#RittenhouseVerdict, #Kenosha. The tweets were gathered from the trial commencement on 11/1/2021 to three days post the verdict announcement (11/22/2021), capturing a period of heightened public attention. This timeframe minimizes memory bias, ensuring more accurate responses immediately following the Rittenhouse incident. After deleting the non-English tweets, we got a total of 132,313 tweets. We aggregated these tweets by hour level to have 528 observations in the time series model.

Measures

Polarization

We examine polarization based on the prior research of polarization of public opinion on social media (Samantray and Pin 2019), using ideological divergence at the tweet level for each hour, known as ideological divergence of emotion-adjusted measure of belief (EAB).The EAB combines two aspects: expressed opinion in the message, and emotional content in a tweet.

A tweet carries an opinion, denoted by *op* and a sentiment denoted by *s*. Opinion (op) signifies whether a tweet is in favor or against vigilantism. We encode op as 1 if the message in the tweet is against the vigilantism statement and as −1 if the message is in favor of vigilantism. For this, we followed content coding. The content analysis was carried out in two stages. The first stage involved pilot coding with a sample of 100 tweets of three rounds. The second stage consisted of coding of 1500 tweets. For the remaining tweets, we used the Linear Support Vector Machine (SVM) classifier, which is the most dominant classification algorithms used in supervised learning (Vapnik 1991; Vapnik and Vapnik 1998).

Irrespective of the opinion, the sentiment (s) can be positive or negative depending on the way the message is communicated. We extract the sentiment (s) using VADER (Hutto and

Gilbert 2014) which is designed to conduct sentiment classification of short texts like tweets. For each tweet, a score is obtained on the scale −1 (most negative) to +1 (most positive).

As discussed earlier, an emotion-adjusted measure of belief called EAB combines these two aspects: expressed opinion in the message, and emotional content in the message. For a given tweet with attributes op and s, EAB is defined as the product of opinion and absolute value of sentiment, i.e., op·|s|, (Samantray and Pin 2019). To calculate the polarization, we use the measure of ideological divergence, which involves characterizing polarization based on kurtosis and skewness across all the tweets in a specific time period, in our case an hour. The polarization of EAB is defined as:

$$\frac{s^2 + 1}{k + 3 \frac{(n-1)^2}{(n-2)(n-3)}}$$

Where s is the skewness, k is the kurtosis of EAB. The variable, n is the sample size of the tweets that displayed the particular hour of the day. The value of 1 suggests that EAB is perfectly bimodal and value 0 signifies that the EAB is perfectly unimodal. Any value greater than 0.56, suggests that there is a tendency towards polarization (Lelkes 2016).

Moral Foundation

To code for moral foundations, we analyzed the tweets using Linguistic Inquiry and Word Count (LIWC). LIWC calculates the percentage of words in a sample of text belonging to several predefined categories (Boyd et al. 2022). LIWC has been used extensively across Information systems and social science research, including in research on Moral Foundations (Grover et al. 2019). We used the Moral Foundations Dictionary 2 (MFD2), which is an extension to the base LIWC dictionary that adds additional word categories for each of the above mentioned five moral foundations (Frimer 2020). The MFD dictionary categories are used as all

other built-in LIWC categories, where an overall percentage rate score is calculated from a sample of text. The updated Moral Foundations Dictionary 2 has enhanced psychometric properties compared to Moral Foundations Dictionary 1 (Frimer 2020). We got the score of each of the Moral Foundations of the tweets using these dictionaries. Further, the mean score of Moral Foundations was computed across tweets in an hour.

Control variables

Using tweets as the unit of analysis had its own limitations. The tweets can vary in length dramatically. Some social media users may express themselves a lot more compared to others on social media. Greater number of words in a message strongly relates to persuasion (Tan et al. 2016), which may influence polarization. Similarly, number of likes amplifies viewpoints of specific tweets, thereby leading to polarization (Masrani et al. 2023). Hence, word count and number of likes was considered as a control variable.

Missing values were imputed using structural model and Kalman Smoothing with the help of imputeTS package in R. Research suggests that Kalman smoothing provides best results for time series data (Moritz and Bartz-Beielstein 2017).

<div align="center">

RESULTS

</div>

We used time series technique called vector autoregression (VAR) to model the dynamic interaction between Moral Foundations and polarization (Lütkepohl 2005). Compared to other models, VAR has an advantage in addressing biases such as endogeneity and autocorrelations (Lütkepohl 2005).

The basic VAR model assumes a stationary time series process. Hence, we conducted Augmented Dickey Fuller test and Phillips Peron Unit Root test on all the variables to confirm that the time series data is stationary. Both the Augmented Dickey Fuller test and Phillips Peron

Unit Root test confirmed that the time series is stationary. Before performing VAR model, it is important to perform optimal lag selection test. The optimal lag for the model based on AIC was 10 hours. We found that Care Moral Foundation third lag, sixth lag and ninth lag are statistically significant but with different signs. The Care Moral Foundation third and sixth lag is positive whereas the ninth lag is negative. This also suggest an overreaction pattern. The Fairness Moral Foundation first and seventh lag are statistically significant and negative. The Loyalty Moral Foundation sixth lag is statistically significant and positive. The Authority Moral Foundation fourth and seventh lag is statistically significant and positive. The Sanctity Moral Foundation first lag is statistically significant and positive. The results are shown in Table1. The granger causality results (chi-square = 24.809, $p < 0.001$) also show that Moral Foundations have significant influence on Polarization at the hourly rate frequency.

Table 1. Coefficient Estimates: The Effect of Moral Foundations on Polarization

Impulse	Response: Polarization	Impulse	Response: Polarization
Care Moral Foundation		Authority Moral Foundation	
Lag1	0.0091	Lag1	-0.041
Lag2	0.0096	Lag2	0.0102
Lag3	0.0235*	Lag3	0.0091
Lag4	-0.0078	Lag4	0.0511*
Lag5	0.0127	Lag5	0.0201
Lag6	0.0283*	Lag6	0.0008
Lag7	-0.00561	Lag7	0.0421*
Lag8	0.0077	Lag8	0.0002
Lag9	-0.0295*	Lag9	-0.0193
Lag10	0.0131	Lag10	-0.0207
Fairness Moral Foundation		Sanctity Moral Foundation	
Lag1	-0.0443*	Lag1	0.0461**
Lag2	-0.0022	Lag2	-0.0068
Lag3	-0.0116	Lag3	-0.0036
Lag4	0.0205	Lag4	0.0139
Lag5	-0.0146	Lag5	-0.0264
Lag6	0.0197	Lag6	-0.0171
Lag7	-0.0434*	Lag7	-0.0161

Lag8	0.0162	Lag8	0.0021
Lag9	-0.0111	Lag9	-0.0234
Lag10	-0.0051	Lag10	0.0267
Loyalty Moral Foundation			
Lag1	0.0023		
Lag2	0.0099		
Lag3	0.0088		
Lag4	0.0106		
Lag5	0.0052		
Lag6	0.0620*		
Lag7	0.0019		
Lag8	0.0248		
Lag9	-0.0329		
Lag10	-0.0504		

*p < 0.05, ** p < 0.01, *** p < 0.0001

DISCUSSION

Through this work, we uncovered how social media messages of different Moral Foundations can impact polarization in social cybersecurity literature. We find that Loyalty Moral Foundation, Authority Moral Foundation and Sanctity Moral Foundation influences polarization positively. It is important because increased polarization on social media has translated to the ground level, manifested through attacks on minorities (Rafee 2020). We recommend that policy makers implement soft moderation on social media platforms to counter polarization (Singhal et al. 2023). For example, users can be informed about potential polarization in the content by adding a warning label (Zannettou 2021). Social media platforms such as X can have a strike system, where users can be discouraged to post information with high Loyalty, Authority and Sanctity Moral Foundations (Singhal et al. 2023). Fairness Moral Foundations negatively influences polarization because it is more based on facts and figures. Social media can modify their algorithm to prioritize content that foster Fairness Moral Foundation and reduce visibility of other MFT to create more balanced online environment (Cinus et al. 2023). The Care Moral Foundations suggest an overreaction pattern, where its short-

term influence is positive and long-term influence is negative on polarization. The initial positive influence of Care Moral Foundation on polarization may be driven by a sense of shared concern or collective empathy because of initial surge in sentiments related to caring for the well-being of individuals affected by the incident. However, differing perspectives, emerging details because of media coverage or societal responses may contribute to observed shift in the relationship with polarization.

CONCLUSION

In summary, the purpose of the study is to investigate what, how and when Moral Foundations messages influence polarization in the context of vigilantism. Social cybersecurity is particularly interested in how social media messages impact polarization (Carley 2020). We find that the Loyalty Moral Foundation, Authority Moral Foundation and Sanctity Moral Foundation influences polarization positively whereas Fairness Moral Foundation influences polarization negatively. Thus, Fairness Moral Foundation can douse polarization.

The study makes important contributions, yet it has some potential limitations. First, we have identified key aspects only using X platform. In future studies we plan to use data from multiple social media platforms. Second, we only chose one vigilantism incident in America and did not consider other incidents or those outside America. We plan to replicate the study in other countries as well.

REFERENCES

Adamic, L. A., and Glance, N. 2005. *The Political Blogosphere and the 2004 US Election: Divided They Blog*, presented at the Proceedings of the 3rd international workshop on Link discovery, pp. 36–43.

Ball, C. A. 1996. "Moral Foundations for a Discourse on Same-Sex Marriage: Looking beyond Political Liberalism," *Geo. LJ* (85), HeinOnline, p. 1871.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R. 2015. "Tweeting from Left to Right: Is Online Political Communication More than an Echo Chamber?," *Psychological Science* (26:10), Sage Publications Sage CA: Los Angeles, CA, pp. 1531–1542.

Baughan, A., Petelka, J., Yoo, C. J., Lo, J., Wang, S., Paramasivam, A., Zhou, A., and Hiniker, A. 2021. "Someone Is Wrong on the Internet: Having Hard Conversations in Online Spaces," *Proceedings of the ACM on Human-Computer Interaction* (5:CSCW1), ACM New York, NY, USA, pp. 1–22.

Boyd, R. L., Ashokkumar, A., Seraj, S., and Pennebaker, J. W. 2022. "The Development and Psychometric Properties of LIWC-22," *Austin, TX: University of Texas at Austin*, pp. 1–47.

Carley, K. M. 2020. "Social Cybersecurity: An Emerging Science," *Computational and Mathematical Organization Theory* (26:4), Springer, pp. 365–381.

Carley, K. M., Cervone, G., Agarwal, N., and Liu, H. 2018. *Social Cyber-Security*, presented at the International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation, Springer, pp. 389–394.

Cinus, F., Gionis, A., and Bonchi, F. 2023. *Rebalancing Social Feed to Minimize Polarization and Disagreement*, presented at the Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, pp. 369–378.

Clayton, S., and Opotow, S. 2003. *Introduction: Identity and the Natural Environment.*, MIT Press.

Day, M. V., Fiske, S. T., Downing, E. L., and Trail, T. E. 2014. "Shifting Liberal and Conservative Attitudes Using Moral Foundations Theory," *Personality & Social Psychology Bulletin* (40:12), pp. 1559–1573. (https://doi.org/10.1177/0146167214551152).

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., and Quattrociocchi, W. 2016. "Echo Chambers: Emotional Contagion and Group Polarization on Facebook," *Scientific Reports* (6:1), Nature Publishing Group, pp. 1–12.

Demszky, D., Garg, N., Voigt, R., Zou, J., Gentzkow, M., Shapiro, J., and Jurafsky, D. 2019. "Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings," *arXiv Preprint arXiv:1904.01596*.

DeScioli, P., and Kurzban, R. 2009. "Mysteries of Morality," *Cognition* (112:2), Elsevier, pp. 281–299.

Frimer, J. A. 2020. "Do Liberals and Conservatives Use Different Moral Languages? Two Replications and Six Extensions of Graham, Haidt, and Nosek's (2009) Moral Text Analysis," *Journal of Research in Personality* (84), Elsevier, p. 103906.

Garrett, K. N., and Bankert, A. 2020. "The Moral Roots of Partisan Division: How Moral Conviction Heightens Affective Polarization," *British Journal of Political Science* (50:2), Cambridge University Press, pp. 621–640.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., and Ditto, P. H. 2013. "Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism," in *Advances in Experimental Social Psychology* (Vol. 47), Elsevier, pp. 55–130.

Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwitch, C., and Mooijman, M. 2018. "Moral Foundations Theory," *Atlas of Moral Psychology* (211), Guilford Publications.

Grover, T., Bayraktaroglu, E., Mark, G., and Rho, E. H. R. 2019. "Moral and Affective Differences in U.S. Immigration Policy Debate on Twitter," *Computer Supported*

Cooperative Work (CSCW) (28:3), pp. 317–355. (https://doi.org/10.1007/s10606-019-09357-w).

Haas, N. E. 2010. *Public Support for Vigilantism*, Leiden University.

Haidt, J., and Graham, J. 2007. "When Morality Opposes Justice: Conservatives Have Moral Intuitions That Liberals May Not Recognize," *Social Justice Research* (20:1), Springer, pp. 98–116.

Haidt, J., Graham, J., and Joseph, C. 2009. "Above and below Left–Right: Ideological Narratives and Moral Foundations," *Psychological Inquiry* (20:2–3), Taylor & Francis, pp. 110–119.

Heltzel, G., and Laurin, K. 2020. "Polarization in America: Two Possible Futures," *Current Opinion in Behavioral Sciences* (34), Political Ideologies, pp. 179–184. (https://doi.org/10.1016/j.cobeha.2020.03.008).

Hetherington, M. J., and Weiler, J. D. 2009. *Authoritarianism and Polarization in American Politics*, Cambridge University Press.

Hutto, C., and Gilbert, E. 2014. *Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text*, in (Vol. 8), presented at the Proceedings of the international AAAI conference on web and social media, pp. 216–225.

Johnston, L. 1996. "WHAT IS VIGILANTISM?," *The British Journal of Criminology* (36:2), pp. 220–236. (https://doi.org/10.1093/oxfordjournals.bjc.a014083).

Koleva, S. P., Graham, J., Iyer, R., Ditto, P. H., and Haidt, J. 2012. "Tracing the Threads: How Five Moral Concerns (Especially Purity) Help Explain Culture War Attitudes," *Journal of Research in Personality* (46:2), Elsevier, pp. 184–194.

Kurumathur, S., Bhatt, P., Hariharan, G., Valecha, R., and Rao, H. R. 2022. "AIS Electronic Library (AISeL) - ICIS 2022 Proceedings: Examining the Public Response to Vigilantism: A Multi-Dimensional Model of Social Media Discourse." (https://aisel.aisnet.org/icis2022/social/social/19/, accessed May 3, 2023).

Kwon, K. H., Oh, O., Agrawal, M., and Rao, H. R. 2012. "Audience Gatekeeping in the Twitter Service: An Investigation of Tweets about the 2009 Gaza Conflict," *AIS Transactions on Human-Computer Interaction* (4:4), pp. 212–229.

Landrum, A., Lull, R., Akin, H., and Jamieson, K. H. 2016. "Making It about Morals: Pope Francis Shifts the Climate Change Debate," *Available at SSRN 2997490*.

Lelkes, Y. 2016. "Mass Polarization: Manifestations and Measurements," *Public Opinion Quarterly* (80:S1), Oxford University Press US, pp. 392–410.

Lütkepohl, H. 2005. *New Introduction to Multiple Time Series Analysis*, Springer Science & Business Media.

Masrani, T. W., Jamieson, J., Yamashita, N., and He, H. A. 2023. "Slowing It Down: Towards Facilitating Interpersonal Mindfulness in Online Polarizing Conversations Over Social Media," *Proceedings of the ACM on Human-Computer Interaction* (7:CSCW1), ACM New York, NY, USA, pp. 1–27.

McCoy, J., Rahman, T., and Somer, M. 2018. "Polarization and the Global Crisis of Democracy: Common Patterns, Dynamics, and Pernicious Consequences for Democratic Polities," *American Behavioral Scientist* (62:1), Sage Publications Sage CA: Los Angeles, CA, pp. 16–42.

Miller, D. T. 2001. "Disrespect and the Experience of Injustice," *Annual Review of Psychology* (52:1), Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, pp. 527–553.

Moritz, S., and Bartz-Beielstein, T. 2017. "imputeTS: Time Series Missing Value Imputation in R.," *R J.* (9:1), p. 207.

Pearlman, W., and Cunningham, K. G. 2012. "Nonstate Actors, Fragmentation, and Conflict Processes," *Journal of Conflict Resolution* (56:1), Sage Publications Sage CA: Los Angeles, CA, pp. 3–15.

Rafee, A. A. 2020. "Polarization on Social Media Platforms Consequences for Politics and Security," *The Digital Age, Cyber Space, and Social Media The Challenges of Security & Radicalization*, p. 173.

Rapp, C. 2016. "Moral Opinion Polarization and the Erosion of Trust," *Social Science Research* (58), pp. 34–45. (https://doi.org/10.1016/j.ssresearch.2016.02.008).

Rosenbaum, H. J., and Sederberg, P. C. 1974. "Vigilantism: An Analysis of Establishment Violence," *Comparative Politics* (6:4), Comparative Politics, Ph.D. Programs in Political Science, City University of New York, pp. 541–570. (https://doi.org/10.2307/421337).

Samantray, A., and Pin, P. 2019. "Credibility of Climate Change Denial in Social Media," *Palgrave Communications* (5:1), Springer Science and Business Media LLC.

Singhal, M., Ling, C., Paudel, P., Thota, P., Kumarswamy, N., Stringhini, G., and Nilizadeh, S. 2023. *SoK: Content Moderation in Social Media, from Guidelines to Enforcement, and Research to Practice*, presented at the 2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P), IEEE, pp. 868–895.

Stolerman, D., and Lagnado, D. 2020. "The Moral Foundations of Human Rights Attitudes," *Political Psychology* (41:3), Wiley Online Library, pp. 439–459.

Taber, C. S., and Lodge, M. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs," *American Journal of Political Science* (50:3), Wiley Online Library, pp. 755–769.

Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., and Lee, L. 2016. *Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-Faith Online Discussions*, presented at the Proceedings of the 25th international conference on world wide web, pp. 613–624.

Uyheng, J., Magelinski, T., Villa-Cox, R., Sowa, C., and Carley, K. M. 2020. "Interoperable Pipelines for Social Cyber-Security: Assessing Twitter Information Operations during NATO Trident Juncture 2018," *Computational and Mathematical Organization Theory* (26), Springer, pp. 465–483.

Vapnik, V. 1991. "Principles of Risk Minimization for Learning Theory," *Advances in Neural Information Processing Systems* (4).

Vapnik, Vladimir, and Vapnik, Vlamimir. 1998. "Statistical Learning Theory Wiley," *New York* (1:624), p. 2.

Venkatesan, S., Valecha, R., Yaraghi, N., Oh, O., and Rao, H. R. 2021. "Influence in Social Media: An Investigation of Tweets Spanning the 2011 Egyptian Revolution," *MIS Quarterly* (45:4), MIS Quarterly, pp. 1679–1714. (https://doi.org/10.25300/MISQ/2021/15297).

Zannettou, S. 2021. *" I Won the Election!": An Empirical Analysis of Soft Moderation Interventions on Twitter*, in (Vol. 15), presented at the Proceedings of the International AAAI Conference on Web and Social Media, pp. 865–876.