Moral-Emotional Content and Patterns of Violent Expression and Hate Speech in Online User Comments

Jeffrey Javed and Blake Miller

Working Paper

Abstract:

What kinds of content are most associated with violent expression and hate speech? We hypothesize that articles rich in moral-emotional content are more likely to provoke outrage against targeted individuals and outgroups, which in turn increases individuals' propensity to use or support violent and hateful expression against them. We use a text analysis of populations of articles and user comments scraped from 72 online media sources that use the Disqus platform to measure correlations between moral-emotional content and user comments that express a desire for violence or contain some kind of attack on an individual or group. Using a novel typology for classifying violent and hateful speech, we annotated a sample of over 12,000 user comments and used this human-annotated corpus to train classification models to classify over 300 million user comments. We find support for our hypothesis that content that contains both moral and emotional rhetoric is significantly more likely to be associated with higher proportions of violent expression. Within comments, we find that user comments containing calls for violence are significantly more likely to contain dehumanizing and demonizing language, target groups as opposed to individuals, and include references to an ingroup identity.

Introduction

A peculiar incident shook the small town of Twin Falls, Idaho, just months before the 2016 presidential election. At a city council meeting, several locals asked why the government was covering up the rape of a local child by Muslim refugees; the mayor and city councilors, knowing nothing of the incident in question, were dumbfounded. Their confusion and the reticence of the police department only intensified locals' outrage and suspicions. A media firestorm ensued, spreading from "alt-right" news outlets to more mainstream conservative media, detailing an increasingly sensationalized government cover-up of a horrific child gang rape in Twin Falls. The story attracted national attention when, less than a week after the city council meeting, The Drudge Report released an article with the headline: "REPORT: Syrian

‘Refugees’ Rape Little Girl at Knifepoint in Idaho.” As public outrage swelled, citizens flooded the mayor with e-mails accusing him and the city council of treason and political misconduct, along with death threats against him, his wife, and even local journalists accused of participating in the cover-up. Yet the alleged rape never occurred, nor did the actual incident that generated the story involve Syrian refugees, who have never been resettled in Twin Falls. The boys involved were not young men, but children ages eight and ten. Somehow, a case of sexual misconduct between three children warped into a sensationalized story about Syrian refugees gang raping a child, which enraged citizens and provoked some to send death threats and demand the deportation of Muslims and refugees (Dickerson 2017).

What occurred in Twin Falls represents one of many cases in which online content alleging extreme morally transgressive behavior has provoked hateful and violent online and offline behavior. In the infamous “PizzaGate” incident, a Facebook post originated an elaborate conspiracy theory alleging that Hillary Clinton and John Podesta were running a Satanic child sex trafficking ring out of Comet Ping Pong, a pizza restaurant in Washington, DC. When Alex Jones, the creator of the far-right news website InfoWars, picked up the story he asserted that “Hillary Clinton has personally murdered and chopped up and raped [children].” This growing litany of alleged evil deeds generated a series of online attacks and death threats against the restaurant owner, and prompted one man to drive from North Carolina to DC with his semi-automatic rifle and pistol to the restaurant to “investigate” (Robb 2017).

Twin Falls, PizzaGate, and other related incidents raise a significant theoretical and empirical question regarding the role of information in the social mobilization of violent behavior: under what conditions can information provoke ordinary citizens to advocate for or engage in violent behavior that they might not otherwise support? While it is convenient to dismiss such incidents as aberrant episodes of media-induced hysteria, it ignores that many of the residents of Twin Falls who denounced the government and called for the deportation of fictitious Syrian refugees from their town acted out of a genuine sense of outrage and concern for a problem afflicting their community. The PizzaGate gunman, too, claimed good intentions, apologizing to the court and saying that he only “came to D.C. with the intent of helping people” (Kennedy 2017). These apocryphal stories of extreme immoral behavior, whether or not they were intended to do so, mobilized citizens to act out of a sense of righteousness.

We seek to understand how information mobilizes popular support for violence by analyzing the connection between moral-emotional information framing, outrage, and violence in online American media content. We hypothesize that *online content that portrays sensationalized accounts of moral violations elicits outrage that in turn mobilizes support for violence against alleged violators*. Similar content that provides identical facts without a strong moral-emotional frame or content will be far less likely, we believe, to elicit outrage and mobilize support for violence.

We assess these hypotheses through a text analysis of populations of articles and comments scraped from major American media outlets to test for correlations between moral-emotional content and expressions of violence and hate speech in users comments.

The rest of this paper is structured as follows. We begin with a discussion of the links between information and violent mobilization and the significance of moral outrage as a mobilizing mechanism. We then lay out our theory, hypotheses, and methodology for our text analysis. The final two sections present our results and conclude.

Information, Moral Outrage, and the Mobilization of Violence

What role does information play in the mobilization of individual support or participation in violence against targeted individuals and outgroups? The scholarship on collective violence, genocide, and war has long emphasized the role of information—i.e. propaganda, usually—in creating or inflaming hatred towards outgroups to generate popular support for violence (Charny 1982; Cohn 1967; Dower 1986; Fein 1979; Herf 2006; Hill 1995; Goldhagen 1996, 2009; Tsesis 2002). Few scholars, however, have directly tested the causal effects of information on violent behavior, and the results of these studies have been mixed. Looking at political attitudes, Kalmoe (2014) finds that violent rhetoric interacts with trait aggression to increase support for political violence. Some scholars argue that online hate speech (Müller and Schwarz 2017) and news incriminating an outgroup (King and Sutton 2013) may translate into on-the-ground violence. In the context of genocidal violence, Hagan and Rymond-Richmond (2008) find that the use of dehumanizing racial epithets during attacks on black African populations in Darfur correlated with more extreme violence. Other scholars have challenged the causal link between hateful content and on-the-ground violence. Looking at the Rwandan genocide, Straus (2007) argues that the spatial coverage of Hutu-controlled "hate radio" cannot explain the geographic variation in the onset of violence. Also looking at Rwanda, Fujii (2004) asserts that it was the combination of ubiquitous anti-Tutsi propaganda with staged attacks aimed to create an atmosphere of immediate threat that "normalized" genocidal behavior.

Moreover, it is unclear what kinds of information frames promote violence. Though scholars have emphasized the primacy of dehumanization frames in genocidal violence (Charny 1982; Fein 1979), there is little empirical evidence for their efficacy. A major underexplored alternative is moral frameworks that justify the "righteousness" of violence (Viterna 2014). Social psychologists have long emphasized the significance of morality—conceptions of right and wrong traits and behaviors—in understanding participation in violence, private and political (Bandura et al. 1975; Baumeister 1999; Beck 1999; Fiske and Rai 2015). Because using violence requires first overcoming formidable moral-emotional reservations, moral narratives that condone and justify violence can erode these otherwise powerful restraints on violent behavior (Baumeister 1999; Beck 1999). Intriguingly, these moral transgression frames may increase individuals' willingness to punish transgressors, in part, by dehumanizing them (Fincher and Tetlock 2016). Moralistic frames may also mobilize participation in violent causes by drawing upon individuals' latent moral convictions. Wood (2003) emphasizes the "pleasure of agency" that participants derive from acting on their moral convictions in a movement, which Viterna (2014, 191) extends to the willingness of citizens to accept violence in the name of "righteous" causes, movements where "interested publics believe that the enactors of political violence are defending society's most vulnerable and protecting a morally legitimate social order." Indeed, Jennet Kirkpatrick (2008) documents a long history of extrajudicial and vigilante violence in America that invokes revolutionary values of freedom, justice, and democracy to frame itself as righteous.

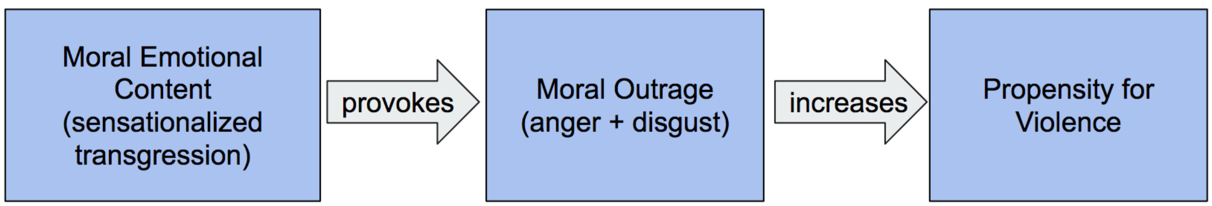
Understanding the effects of highly moralized information frames on violent behavior requires attention to the emotional responses these frames provoke. Despite an abundant literature on campaign messaging and voter behavior (Ansolabehere and Iyengar 1995; Freedman and Goldstein 1999; Mendelberg 2001; Valentino et al. 2002) and emotional response to campaign messaging (Brader 2005, 2006; Marcus et al. 2000), the emotional microfoundations of violent mobilization remain underexplored (Viterna 2013). Anger, or outrage, appears to be the most relevant mobilizing emotion in the context of violence. Unlike fear, which tends to demobilize, social psychologists have found that anger and outrage have mobilizing effects (Banks 2014; Jasper 1997, 2014; Lerner et al. 2001; Valentino et al. 2011). Significantly, content that conveys or sensationalizes moral violations may lower an individual's threshold for using violence through provoking moral outrage against violators, which "motivates people to shame and punish wrongdoers" (Crockett 2017). The presence of a moral framework appears critical to the strength and duration of outrage's mobilizing effect. Goldberg et al.'s (1999) experimental work on outrage demonstrates the importance of morality in provoking and sustaining a desire for retributive reprisal. They find that priming outrage through the revelation of unpunished "normative violations" triggers an "intuitive prosecutor" mindset, whereby affected individuals will more readily accept and propose harsh punishment, not just of the original transgressors but future, unrelated transgressors as well. Significantly, they do not find this effect when they prime anger without embedding it in a moral frame.

Under some circumstances, however, uncontextualized, or generalized, anger may be sufficient to mobilize. Antoine Banks (2014) argues that non-group specific appeals that stimulate generalized anger are sufficient to mobilize voters because they trigger latent racist feelings that, ultimately, are rooted in experiences of anger. These studies, however, tend to provoke outrage through means unrelated to information content: they use videos or ask respondents to recall emotional events to provoke emotional states, which are then used to test the link between emotional response and political behavior, often voting.

Theory and Hypotheses

We hypothesize that individuals' consumption of content that contains sensationalized or extreme moral transgressions will provoke outrage that increases their propensity for violence—namely, heightened acceptance of and support for severe punishment, including physical violence or the confiscation or destruction of assets. The causal pathway between this moral-emotional content and propensity for violence is outrage, a category of "feelings that stem from violating evaluative cultural codes, that is, codes that indicate what is good or bad or right or wrong in a society" (Stets 2012, 330; Turner and Stets 2006). Figure 1 illustrates this argument.

Figure 1. Schematic for the Central Argument



In the context of online content and expression, we conceptualize users' expression of their increased propensity for violence as *calls for violence*. Our principal hypothesis (H1) argues that exposure to online content containing sensationalized morally transgressive behavior—i.e. behavior that breaches norms regarding good and appropriate behavior—will increase calls for violence.

> *H1*: Online content that contains sensationalized moral transgressions is more likely to mobilize calls for violence against perceived transgressive individuals and/or groups.

In addition to calls for violence, we suspect that moral-emotional content will increase user attacks—expressions of hatred or disdain against a person or group. These attacks include both personal attacks and "hate speech"—i.e. attacks on entire categories of persons.

> H2: Moral-emotional content will increase users' propensity to make attacks on individuals or groups.

Since group validation or invalidation of individual emotions may increase or negate their effects, we consider the possibility that calls for violence are partly a function of ingroup identity. While the existence of an ingroup does not necessitate hostility toward an outgroup

(Brewer 1999), expressions of ingroup identity in user content may indicate strong ethnocentrism, which is associated with increased hostility towards outgroups (Kinder and Kam 2010). Thus, we expect that users who invoke ingroup identities to be more likely to make calls for violence.

> H3: Online users who invoke ingroup identities will be more likely to make calls for violence.

Methodology and Variable Operationalization

We conduct a series of observational tests to investigate the link between moral-emotional content and violent expression online. We scrape and analyze populations of articles and comments from 72 media outlets that use the Disqus platform, a popular commenting platform used by many media outlets. These sources are selected by identifying popular far-right-leaning, right-leaning, moderate, left-leaning, and far-left-leaning "soft news" publications such as Occidental Dissent, Breitbart, The Hill, Mother Jones, and Forward respectively. Using the Amazon Alexa platform, we used audience overlap tools to discover sites with similar audiences. For each site, we build a scraper to capture the population of articles on the site and the entire comment thread from each article. Because these comment sections are lightly moderated, we run scrapers persistently, fetching new comments every five minutes, before moderators can remove them. For the population of historical comments that were not gathered through persistent scraping, we know which comments have been deleted from the API, but the text content of these comments has been redacted.

Using human-coded labels of user comments, we train several binary classification models to classify unlabeled, out-of-sample comments, since labeling all 300 million comments would be intractable. Our research team hand coded a random subset of 9,587 of these scraped comments for the following variables: 1) calls for violence; 2) attacks; 3) group references; 4) ingroup references; and 5) person references. The first three variables are parent categories with several subcategories. We conceptualize and operationalize each of these categories as follows:

1. *Calls for violence* include user-generated text that mentions or implies credible support or desire for the punishment of or infliction of harm against persons, groups, and/or their assets. This category contains four subcategories: lethal violence, property-based violence, judicial violence, and forcible displacement or restriction of movement. *Lethal violence* refers to calls for violence involving violence that will or most likely will result in the death of the targeted party. *Property-based violence* refers to calls for violence that mention taking by force, destroying, or vandalizing property or assets—e.g. confiscating property, burning down homes, bombing community centers, etc. *Judicial violence* includes punishments that follows established judicial procedures—e.g. putting someone on trial, recommending a harsher sentence for a convicted criminal, etc. *Forcible displacement or restriction of movement* are calls for violence that involve deporting, detaining, or imprisoning persons and/or groups.
2. *Attacks* are expressions of hatred, irritation, disdain, or disgust towards persons and/or groups. The parent category of attack has three subcategories: dehumanizing attacks, demonizing attacks, and attacks containing foul language. *Dehumanizing attacks* state or imply that targeted persons or groups are subhuman or non-human or compare them with non-human animals or inanimate objects—e.g. beasts, dirt, etc. *Demonizing attacks* state or imply the perceived violation of moral norms by persons or groups, including but not limited to claims

of their criminality, depravity, or evil nature, etc. Finally, *foul language attacks* contain expletives, including those that are self-censored or written as acronyms.

3. *Group references* include all references to broad, identity-based categories of persons. This category contains eight subcategories for groups whose boundaries are defined by political affiliation or views; race or ethnicity; religious affiliation; gender; sexuality; nationality or citizenship status; corporate affiliation or occupation; or affiliation with journalism or mass media.
4. *Ingroup references* invokes a group or groups to which the user states or implies they belong—e.g. "our country," "we as taxpayers," "as an American citizen," etc.
5. *Person references* include all references to a specific person or a discrete, countable number of identifiable persons. For example, "President Obama's aides" would count as a reference to a person, but "Trump's base" would not.

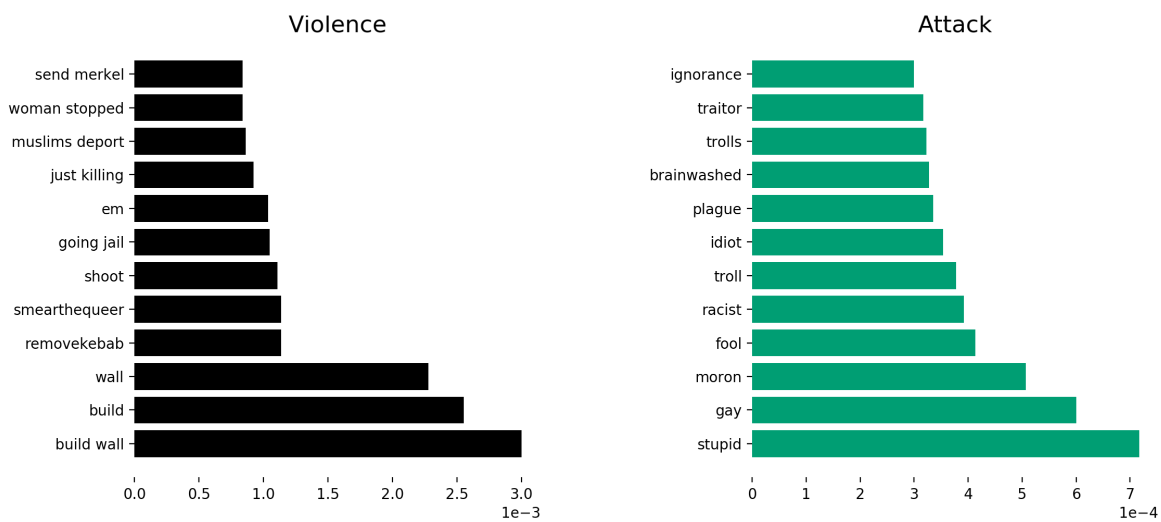
For each variable discussed above, we train a binary L2-normalized linear support vector machine. Instead of a simple bag of words model, we chose a continuous word embedding feature representation because of the availability of powerful pre-trained word embedding models, and because continuous representations of comments better capture covariances of the semantic features of the text.[1] For our purposes, we will make use of transfer learning and utilize pre-trained Glove word embeddings that have been trained on the Twitter firehose and contain 27 billion tokens. To use these embeddings to represent comments, we calculate a tfidf-weighted

[1] More specifically, we used tfidf (term frequency inverse document frequency) weighted mean average embeddings as features. Word embeddings are fixed dimensional vector representations of words, learned from co-occurrence in text corpora.

average of all words in the comment.[2] Tfidf weights are used in information retrieval to give a weight to the importance of a particular word. The top tfidf weights for the "attack" and "violence" variables is visualized in Figure 2.

Each model's hyperparameters are selected using a randomized hyperparameter search that maximized the F1 score on a held-out development set. Model performance measures can be found in the Table 1. With a total dataset of approximately 300 million comments, we use these models to automatically label comments that are outside of our training sample. Qualitative review of model predictions suggest that these models accurately detect the underlying concept.

Figure 2. Features with the Highest TFIDF Weight by Class



To test H1—that online content that contains sensationalized moral transgressions are more likely to mobilize calls for violence against perceived transgressive individuals and/or groups—we fit a negative binomial regression model with the count of violent comments in the article's comment section as the dependent variable, and the count of moral words, the count of

[2] For example, the word "the" and the word "plesiosaur" would have low weights in a corpus of political texts because "the" is very common and does not provide a lot of information, and because "plesiosaur" might occur only one time in a policy document about protection of archeological sites. A mean average embedding is the tfidf-weighted average of each fixed dimensional word vector. The advantage of using word embeddings is that "good," "gooooood," "great," and "nice" will all be projected into a similar vector space.

emotional words, and the interaction of moral and emotional words. We expect the interaction of moral and emotional words to be positively associated with the number of violent comments in each article's comment section. Since content that only contains emotional words or moral words will most likely not be content that contains information on morally transgressive behavior, we do not expect moral words or emotional words to have an independent positive association with calls for violence.

Table 1: Model Performance Measures

	F1[3]	**Precision**[4]	**Recall**[5]	**Training Set Proportion**[6]	**Out-of-sample Proportion**[7]
Violence	0.45	0.45	0.45	0.091	0.093
Attack	0.73	0.76	0.7	0.502	0.56
Ingroup Appeal	0.75	0.71	0.78	0.135	0.253
Group	0.84	0.82	0.87	0.626	0.706

To test H2—that moral-emotional content will correlate with a higher proportion of attacks on persons or groups—we will run an additional negative binomial regression model with the total count of "attacks" in the article's comment section as the dependent variable. The same count variables of moral and emotional words from the Brady et al. (2017) dictionary and their

[3] F1 is the harmonic mean of precision and recall. $$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

[4] Precision measures how accurate the guesses of the relevant class are. $$Precision = \frac{\texttrue positive}}{\text{true positive} + \text{false positive}}$$

[5] Recall measures how many of the relevant class were recalled. $$Recall = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}$$

[6] The proportion of the training set where this label is positive. Smaller proportions usually have lower baseline performance measures.

[7] The proportion of each class should be similar in the training set and the out-of-sample prediction.

interactions are the independent variables. We expect the interaction of moral and emotional words to be positively associated with the number of attacks in each article's comment section.

To test H3—that online users who invoke ingroup identities will be more likely to make calls for violence—we fit a logit regression model with presence of a call for violence in a given comment as the dependent variable, and the presence of ingroup references in that same comment as an independent variable. We expect the presence of an ingroup appeal in a comment to be positively associated with the presence of a call for violence.

Results

Both the model for calls for violence and for attacks bear out the predictions of H1 and H2, respectively. As Figure 3 illustrates, the combination of moral and emotional language in online articles predicts higher proportions of calls for violence and attacks in user comments. The coefficient on the interaction term for moral and emotional language is positive and significant at the 95 percent confidence level. Surprisingly, the individual coefficients for moral language and emotional language are negatively correlated with calls for violence and attacks. Taken together, these results suggest that moral-emotional content—compared to distinctly moral or emotional content—is more likely to provoke calls for violence and attacks by internet users.

Figure 3. Coefficient Plot for Violence and Attack Models

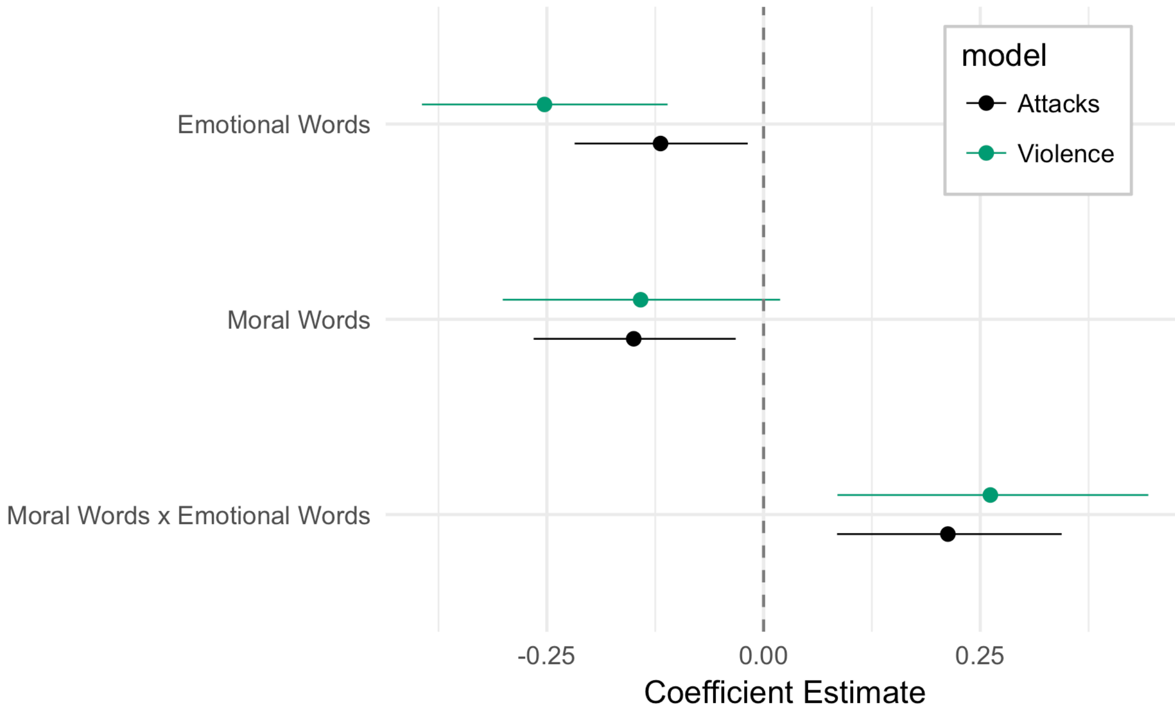
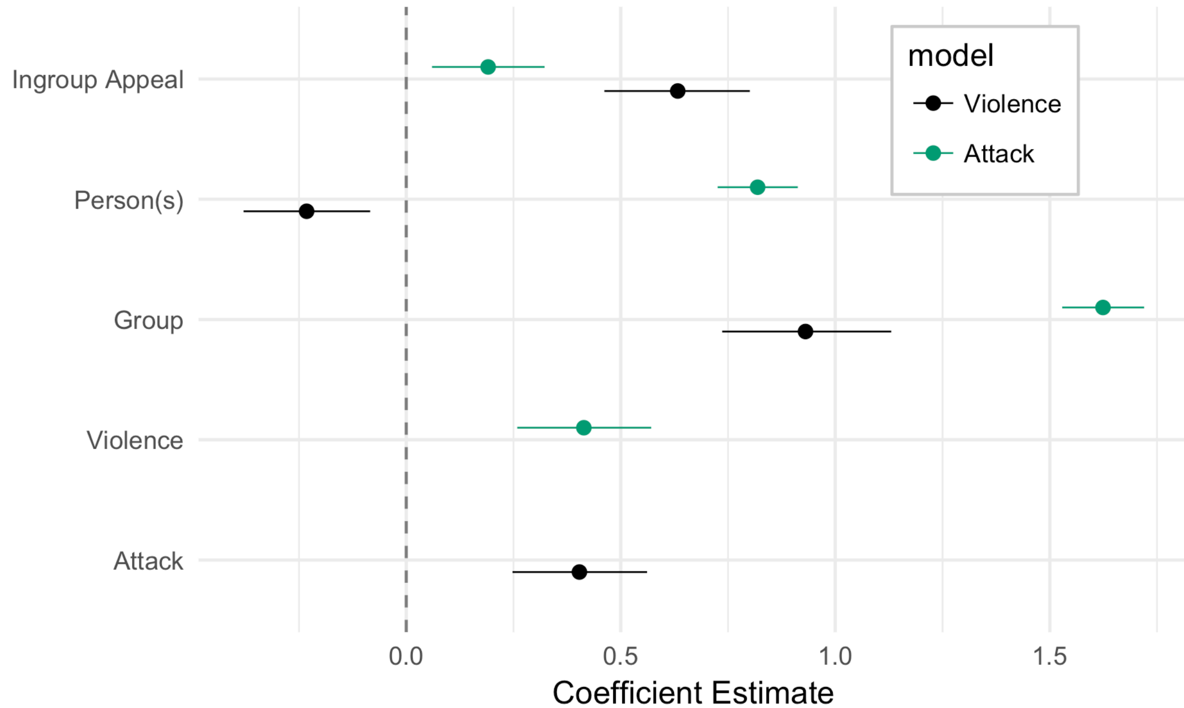


Figure 4 presents results for the logit regression model of the characteristics of calls for violence. We find that, in accordance with H3, ingroup references correlate positively and significantly with calls for violence. This means that invoking a “we” or an “us” in a comment is associated with a higher probability of making a call for violence in that same comment. Any reference to a group also correlates positively and significantly with the presence of a call for violence, though references to a person correlates negatively. Calls for violence are more likely to be directed against groups than persons. Along with the finding that ingroup references are positively associated with calls for violence, it appears that group identities—be they outgroups or ingroups—figure prominently in calls for violence.

Figure 3: Coefficient Plot for Top-Level Categories



Discussion and Conclusion

We find that moral-emotional content correlates strongly with calls for violence and attacks in user comments. Group references, particularly ingroup references, often appear alongside comments that contain calls for violence, which suggests that this content may activate latent ethnocentric dispositions or simply attract more ethnocentric users.

Understanding the importance of moral emotions and violence-promoting content in political life has perhaps never been more important than in the present. Since misleading content by its definition warps or fabricates content, it is critical to understand the conditions under which such content can provoke violent reactions online. Moreover, a great deal of this content is moral emotional in nature, as it tends to sensationalize alleged misdeeds; it contains stories of corruption, sexual misconduct, criminal behavior, etc. by high-profile politicians and celebrities or by ordinary people who are associated with a targeted outgroup—e.g. a Muslim, an undocumented immigrant, a transgender person, etc. This content may intend to provoke on-the-ground violence. For example, a leaked "style manual" from the neo-Nazi website *The Daily Stormer* instructs its writers that: "It's illegal to promote violence on the Internet. At the same time, it's totally important to normalize the acceptance of violence as an eventuality/inevitability" (Feinberg 2017). Lastly, an analysis of the possible effects of online content on individual propensity for violence will provide empirical evidence that can contribute to the current debate on the regulation of "hate speech" that has been dominated mainly by legal and philosophical concerns about freedom of speech (Gates et al. 1994).

The proposed theory here is not restricted to the contemporary "fake news" phenomenon or online content: the use of morally charged content to mobilize violence has a long genealogy. The circulation of Cotton Mather's book *Memorable Providences* detailing "real" accounts of bewitchings of innocent children fueled the persecution of "witches" in the Massachusetts Bay Colony (Hill 1995). Decades before the rise of Nazism, the *Protocols of the Elders of Zion*, which described a plan for Jewish world domination, circulated throughout Europe and primed the German public for Nazi mobilization (Cohn 1967; Herf 2006). To bolster the morale and fighting spirit of American soldiers in the Asia-Pacific in World War II, American anti-Japanese propaganda used stories of Japanese atrocities against American POWs, along with dehumanizing tropes, to stoke soldiers' outrage against the Japanese foe (Dower 1986). Why political actors and social movement leaders choose these strategies for mobilizing moral outrage, when they are effective, and the degree of individual responsiveness to them are questions belonging to a larger research agenda on morality, emotions, and political mobilization to which this study hopes to contribute.

References

Ansolabehere, Stephen, and Shanto Iyengar. *Going Negative: How Political Advertisements Shrink and Polarize the Electorate*. New York: The Free Press, 1995.

Bandura, Albert, Bill Underwood, and Michael E. Fromson, "Disinhibition of Aggression through Diffusion of Responsibility and Dehumanization of Victims," Journal of Research in Personality 9.4 (1975): 253-69.

Banks, Antoine J. *Anger and Racial Politics: The Emotional Foundation of Racial Attitudes in America*. New York: Cambridge University Press, 2014.

Baumeister, Roy F. *Evil: Inside Human Violence and Cruelty*. New York: Henry Holt, 1999.

Beck, Aaron T. *Prisoners of Hate: The Cognitive Basis of Anger, Hostility, and Violence*. New York: Perennial, 1999.

Brader, Ted. *Campaigning for Hearts and Minds: How Emotional Appeals in Political Ads Work*. Chicago: University of Chicago Press, 2006.

---. "Striking a Responsive Chord: How Political Ads Motivate and Persuade Voters by Appealing to Emotions." *American Journal of Political Science* 49, no. 2 (2005): 388-405.

Brady, William J., Julian A. Wills, John T. Jost, Joshua A. Tucker, and Jay J. Van Bavel. "Emotion shapes the diffusion of moralized content in social networks." *Proceedings of the National Academy of Sciences* 114, no. 28 (2017): 7313-7318.

Brewer, Marilynn B. "The Psychology of Prejudice: Ingroup Love and Outgroup Hate?" *Journal of Social Issues* 55, no. 3 (1999): 429-444.

Charny, Israel W., and Chanan Rapaport. *How Can We Commit the Unthinkable? Genocide, the Human Cancer*. Boulder, CO: Westview Press, 1982.

Cohn, Norman. *Warrant for Genocide: The Myth of the Jewish World-conspiracy and the Protocols of the Elders of Zion*. New York: Harper & Row, 1967.

Crockett, M.J. "Moral outrage in the digital age," *Nature Human Behaviour* 1, no. 11 (2017): 769-771.

Dickerson, Caitlin. "How Fake News Turned a Small Town Upside Down." *The New York Times Magazine*. September 26, 2017. https://www.nytimes.com/2017/09/26/magazine/how-fake-news-turned-a-small-town-upside-down.html

Dower, John W. *War Without Mercy: Race And Power In the Pacific War*. New York: Pantheon Books, 1986.

Fein, Helen. *Accounting for Genocide: National Responses and Jewish Victimization during the Holocaust*. New York: Free Press, 1979.

Feinberg, Ashley. “This is The Daily Stormer’s Playbook.” *HuffPost*. December 13, 2017. https://www.huffingtonpost.com/entry/daily-stormer-nazi-style-guide_us_5a2ece19e4b0ce3b344492f2.

Fincher, Katrina M., and Philip E. Tetlock. “Perceptual Dehumanization of Faces is Activated by Norm Violations and Facilitates Norm Enforcement.” *Journal of Experimental Psychology* 145, no. 2 (2016): 131-146.

Fiske, Alan P. Fiske and Tage S. Rai. *Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships*. Cambridge: Cambridge University Press, 2015.

Freedman, Paul, and Ken Goldstein. “Measuring Media Exposure and the Effects of Negative Campaign Ads.” *American Journal of Political Science* 43, no. 4 (1999): 1189-1208.

Fujii, Lee Ann. “Transforming the Moral Landscape: The Diffusion of a Genocidal Norm in Rwanda.” *Journal of Genocide Research* 6, no. 1 (2004): 99-114.

Gates, Jr. Henry Louis, Anthony P. Griffin , Donald E. Lively, and Nadine Strossen. *Speaking of Race, Speaking of Sex: Hate Speech, Civil Rights, and Civil Liberties*. New York: New York University Press, 1994.

Goldberg, Julie H., Jennifer S. Lerner, and Phillip E. Tetlock. “Rage and Reason: The Psychology of the Intuitive Prosecutor.” *European Journal of Social Psychology* 29 (1999): 781-95.

Goldhagen, Daniel J. *Hitler's Willing Executioners: Ordinary Germans and the Holocaust*. New York: Knopf, 1996.

---. Worse than War: *Genocide, Eliminationism, and the Ongoing Assault on Humanity*. New York: Public Affairs, 2009.

Herf, Jeffrey. *The Jewish Enemy: Nazi Propaganda During World War II And the Holocaust*. Cambridge: Harvard University Press, 2006.

Hill, Frances. *A Delusion of Satan: The Full Story of the Salem Witch Trials*. New York: Doubleday, 1995.

Jasper, James M. *The Art of Moral Protest: Culture, Biography, and Creativity in Social Movements*. Chicago: University of Chicago Press, 1997.

---. “Constructing Indignation: Anger Dynamics in Protest movements.” *Emotion Review* 6, no. 3 (2014): 208-213.

Kalmoe, Nathan P. "Fueling the Fire: Violent Metaphors, Trait Aggression, and Support for Political Violence." *Political Communication* 31, no. 4 (2014): 545-563.

Kennedy, Merrit. "'Pizzagate' Gunman Sentenced to 4 Years in Prison." *NPR*. June 22, 2017. https://www.npr.org/sections/thetwo-way/2017/06/22/533941689/pizzagate-gunman-sentenced-to-4-years-in-prison.

Kinder, Donald R. and Cindy D. Kam. *Us Against Them: Ethnocentric Foundations of American Opinion.* Chicago: University of Chicago, 2010.

Kirkpatrick, Jennet. *Uncivil Disobedience: Studies in Violence and Democratic Politics*. Princeton: Princeton University Press, 2008.

Lerner, Jennifer S. and Dacher Keltner. "Fear, Anger, and Risk." *Journal of Personality and Social Psychology* 81, no. 1 (2001): 146-159.

Marcus, George E., W. Russell Neuman, and Michael MacKuen. *Affective Intelligence and Political Judgment*. Chicago: University of Chicago Press, 2000.
Mendelberg, Tali. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton: Princeton University Press, 2001.

Müller, Karsten and Carlo Schwarz. "Fanning the Flames of Hate: Social Media and Hate Crime" Working Paper (December 6, 2017). Available at SSRN: https://ssrn.com/abstract=3082972

Perry, Elizabeth J. "Moving the Masses: Emotion Work in the Chinese Revolution." *Mobilization: An International Journal* 7, no. 2 (2002): 111-128.

Robb, Amanda. "Anatomy of a Fake News Scandal." *Rolling Stone*. November 16, 2017. https://www.rollingstone.com/politics/news/pizzagate-anatomy-of-a-fake-news-scandal-w511904

Stets, Jan E. "Current emotion research in sociology: Advances in the discipline." *Emotion Review* 4, no. 3 (2012): 326-334.

Stenner, Karen. *The Authoritarian Dynamic*. Cambridge University Press, 2005.

Straus, Scott. "What is the Relationship between Hate Radio and Violence? Rethinking Rwanda's 'Radio Machete.'" *Politics & Society* 35, no. 4 (2007): 609-637.

Tsesis, Alexander. *Destructive Messages: How Hate Speech Paves the Way for Harmful Social Movements*. New York: NYU Press, 2002.

Turner, Jonathan H., and Jan E. Stets. "Sociological Theories of Human Emotions." *Annual Review of Sociology* 32 (2006): 25-52.

Valentino, Nicholas A., Ted Brader, Eric W. Groenendyk, Krysha Gregorowicz, and Vincent L. Hutchings. "Election Night's Alright for Fighting: The Role of Emotions in Political Participation," *The Journal of Politics* 73, no., 1 (2011): 156-70.

Valentino, Nicholas A., Vincent L. Hutchings, and Ismail K. White. "Cues That Matter: How Political Ads Prime Racial Attitudes during Campaigns." *American Political Science Review* 96, no. 1 (2002): 75-90.

Viterna, Jocelyn. "Radical or Righteous? Using Gender to Shape Public Perceptions of Political Violence." In *Dynamics of Political Violence: A Process-Oriented Perspective on Radicalization and the Escalation of Political Conflict*, edited by Lorenzo Bosi, Charles Demetriou, and Stefan Maltaner, 189-216. Surrey, UK: Ashgate Publishing, 2014.

---. *Women in War: The Micro-processes of Mobilization in El Salvador*. New York: Oxford University Press, 2013.

Wood, Elisabeth. *Insurgent Collective Action in El Salvador*. New York: Cambridge University Press, 2003.