THE UNIVERSITY OF CHICAGO

The Popularity of extremism:

A computational exploration of extremist political speech on Twitter

By

Shuyi Yang 2023

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Jon Clindaniel

The Popularity of extremism

Abstract

We define political extremism as a duality of internal ideology and external performance and explore whether its performative component on digital platforms is associated with greater popularity. We select three measures of extremism performance supported by prior literature and examine their ability to predict human-labeled extremism in crowd-sourced data. Then we observe how the social media posts' popularity is dependent on these measures. We find that, out of the three measures, sentiment has a statistically significant association with popularity, but could not yield evidence to confirm it to be a good predictor of extremism.

Introduction

Extremism is an anti-establishment political ideology emphasizing power struggle activities that transgresses normal peaceful competitions (Böttischer, 2017). Treating politics as a zero-sum game, extreme ideologies, and their adherents regardless of leaning to the left or the right, compared to their more neutral counterparts, pursue temporarily appealing yet irrational agendas, adopt more intolerant speech and hateful attitudes towards any dissent opinions, and attack political opponents with more relentless and militant means. For instance, left-wing extremists demand more rapid societal and institutional changes than moderate social democrats while condoning the use of violence and disorder; right-wing extremists safeguard rigid values and dream to rescind previous social progress utilizing even authoritarian vessels, while established conservatives tend to reserve flexibilities of liberal reforms. The key feature that distinguishes extremists, whether left or right, from more general progressivists or radicals is

their authoritarian nature, ultimately seeking to coerce all fellow citizens into submission through the absolute grip of power (Koselleck, 2010).

In an optimistic spirit, a set of universal values such as international collaboration and amalgamation, removal of trade and travel barriers, and the progress of feminism and minority rights had been the mainstream rhetoric for decades post-Cold War. But unresolved tensions due to immigration policies, income inequalities, and other debates exacerbated a social schism and gave rise to more extreme political tendencies, exemplified by populist demagogues like Donald Trump in the US, Boris Johnson in the UK, and Australia's Scott Morrison, with extreme political parties winning more parliamentary seats in many other countries. Their left-wing counterparts often do not achieve the same levels of success, except for Italy's *Movimento 5 Stelle* (Five Star Movement), with the more dominant position already occupied by the rights. Despite the defeat of Trump in 2020, the polarizing tendency persisted in the US, with far-right or far-left groups gaining more grounds in many countries (Doxsee et al, 2022); though Macron has secured his reelection, the victory was a precarious one with significantly fewer supporters compared to 2017 and his central-liberal party is likely to lose majority in parliament. In Germany, where a coalition of relatively moderate parties held power, there is a modest yet visible growth of a left-populist party, The Left (*Die Linke*), as well as the right-wing populist AfD (*Alternative für Deutschland*, Alternative for Germany), in the federal Parliament, the Bundestag (Klikauer, 2018). Extremism is growing ever stronger and its consequences should not be neglected: first, having observed a few administrations of such politicians we can see that their strength lies in boasting slogans and rousing sentiments rather than governing states and simply being unable to redeem promises made in elections. More importantly, the rise of extremist politicians amidst a period of stagnation left previously successful states such as the

US deeply divided, undermining citizens' enthusiasm to discuss and participate politics (Dimock and Gramlich, 2021). This can be seen as a blow to the democratic institution and gives people incentives to covet "strong leaders" reminiscent of Mussolini and Hitler, which could lead to even more catastrophic events. Therefore, we must be vigilant of this tendency and monitor it with discretion. Social scientists are well-positioned to reflect on the past, study the present, and provide the public with relevant information that can help shape the development of civic society.

Today, given the widespread use of decentralized social media platforms, opinions travel at unprecedented speed. Compared to Napoleon's printed press and Goebbels' radio, marked as the seventh and eighth great power in Goebbels' speech "The Radio as the Eighth Great Power"[1], social media allows extremist speech to ever more efficiently attract an audience, rally advocates, and merge them into a potent political force. The former two media innovations revolutionized political campaigns and sparked peaks of nationalism; internet social media is the more formidable successor employed by even more minuscule yet polarized parties. Until the late crackdown on extreme speech by major sites such as Twitter and Facebook after the Jan 2021's US Capitol riots, those on the far right, including Donald Trump himself, have actively engaged in conversations on these platforms and furthered their courses. Active participation in online discourses and thus closer ties with constituents were major advantages that contributed to Trump's victory in 2016 and the success of other populist politicians alike. Innovative at the time, this strategy was subsequently adopted by most candidates across the political spectrums.

* The first six being the six Great Powers of the period, England, France, Germany, Italy, Austria and Russia; the English translation of the speech is based on Goebbels, Joseph, "Der Rundfunk als achte Großmacht," *Signale der neuen Zeit. 25 ausgewählte Reden von Dr. Joseph Goebbels* (Munich: Zentralverlag der NSDAP., 1938), pp. 197-207.

Even after this seemingly major drawback in 2021, with many Trump supporters and more excessive alt-right groups' forced exodus to their smaller specialized counterparts, such as Gab or Truth Social,[2] their phantoms are still visible on mainstream social media platforms: a post with few MAGA slogans most ironically represented by "Trump won 2020" theories and Christian religious references could quickly become the forum of a right-wing cohort; and equally alarming are the active responses from the far-left like Antifa, who apply same levels of aggressive language and are threatening action, usually attacking the police and judiciary forces. Having two distinctive factions grimly barricading themselves in these camps of discourse, spaces are shrinking for the candid, rational voices, and traditional values of modern democracy.

That these kinds of speech carry so much power is concerning. In times of paper press and radio, extreme language was provoking and incendiary; it is even more so in the age of the internet. Certainly, the convenience and expediency of the technology linked all the once isolated extremists, and the common internet anonymity allowed them to fully express daring ideologies with impunity, but the present extreme communities are way more populous than these core members with many loose followers assembled under their banners. The question is: Do extreme speeches attract audiences and followers more than neutral calm expressions?

In this paper, I will answer these questions through a large-scale quantitative study of Twitter data. My definition of extremism, as described at the beginning, is comprised of its internal anti-establishment ideology and its external performance expressed by agitating actions. The internal ideology, on one hand, is ambivalent and difficult for digital tools to quantify; the performance, on the other hand, is straightforwardly visible and easier to be transformed into

[2] See the online news reference "Key facts about Gab" and "Truth Social's Influence grows" about the right-wing social networks.

interpretable statistics. Moreover, “digital cultures are performative cultures” (Leeker, Schipper, and Beyes, 2017). Under the social media environment, the external performance largely maps to the agent’s intentions and therefore the working definition of extremism in this paper shall be primarily based on the performative social media activities, leaving their underlying reasonings for a future study. This study explicates the performance of extremism in two stages: first, it develops a criterion based on natural language processing for identifying and categorizing social media posts of extremist political nature; then it provides statistical evidence to support the proposition that more radical speech wins more viewers/upvotes/comments on the platform, and thus the biased are more outspoken and visible. The relationship between the extremism level and the attention statistics of posts, measured by the number of upvotes and reposts, is determined via regression analysis. The current results will support some predictions on the future developments of social media politics, as well as suggest a few measures be taken that curb the issue – such as encouraging platforms to cultivate an environment that venerates neutrality and moderate political identities over extreme values.

Literature review

I. Background: Contemporary Extremism and the Internet

As early as 2006, scholars were aware of the emergence of right-wing extremism in the digital world, but its first mainstream interference was not seen until the 2008 US presidential election. Blee and Creasap (2008) outlined the initial motivations and basic appearances of such

movements and pointed out three directions of future research: the right-wing activists' surrounding environments and subcultures, cross-country connections and collaborations, and how they interacted with the established institutions in the US. Inadequate attention was given to the potential of combining the right-wing movements with the internet as a propaganda tool. The report considered the internet's major advantage for the rights to be its virtualness, on one side, shielding the participants from risking reputations, jobs, or family relations due to such involvements. On the other side, the internet also enabled the creation of virtual communities that offer a sense of belonging, companionship, and mutual support. Indeed, the protection of the PC or cellphone screen has encouraged people to be more outspoken about their opinions, but the feat benefits moderates and extremists alike and is far from being the main factor of the rise of internet extremism. As a major innovation of the last three decades, the internet has spawned novel functionalities at an incredible pace, and it is reasonable that studies at its dawn could not foresee the later rapid developments. It was the alarming triumph of Trump in 2016 that sparked more academic interest in the impact of social media on right-wing movements with improved understanding. The digital new media gave those "counterpublics" means to circumvent traditional journalistic gatekeepers and create direct linkages to the common people (Schroeder, 2018). It liberated them from the physical hideouts or clandestine gatherings and offered a straightforward, conspicuous platform for public demonstration with which they can draw folk of little political training or ideological inclination using enticing conditions and cheering slogans. Gounari (2018) also noticed the paradox that social media, though initially hailed as an emancipatory tool that promoted ordinary people's participation in democracy, was also the ideal medium of the right-wing populists. Gounari attributed the phenomenon to social media's lack of contextualization, a culture of amusement, and overall fragmented structure, thus escaping the

traditional frameworks and confinements and allowing the right to proliferate on the fertile soil. Social media packaged intricate political concepts into everyday merchandise of emotional arousal, excitement, and promises and made it accessible to all ordinary people, many of whom were limited by barriers of literacy, education, or social status, and were loaded with grievances seeking a channel to speak for themselves.

Matching the prevalent growth of the far-right movement, the far-left movement also became more visible in public forums. Long have social scientists compared left and right-wing radicals and concluded a series of resemblances. In fact, in the contemporary atmosphere where political assemblies are increasingly decentralized, their differences are rapidly disappearing. McClosky and Chong (1985) have already pointed out that, despite drastically different ideologies, ultra-lefts and ultra-rights bear similar organization methods, psychological inflexibility, political zeal, and most importantly, estrangement from established institutions. They have correctly predicted that in European countries, due to more radical traditions, left and right extremism would render greater impacts and reveal more similarities, as we observe in the last French election with the rise of both the ultra-right Le Pen and the aggressive left Mélenchon; according to Drake (2018), Italy's left-wing populist movement, *Movimento 5 Stelle,* while promoting environmental preservation, criticizing bankers and capitalists, as well as endorsing a citizen stipend in standard left-wing, socialist fashion, attacked the liberal globalization and immigrants for endangering local jobs in striking resemblances to ultra-right movements of the US and other European countries, even attracting former conservative voters. But McClosky and Chong mistakenly believed that the US would see primarily a proliferation of right-wing movements with their left counterparts historically repressed since the Cold War. In traditional politics McClosky and Chong's standpoint was until recently very much correct, with

the socialism sympathizing factions and leaders constantly restrained even within the Democratic Party; and though Bernie Sanders had some appealing performances early in the 2020 Democratic Primaries, he could not be chosen as the Democrats' common factor representing the party against Trump. However, on American social media lefts and far-lefts enjoyed a great deal of space as well. Starting in 2016 as responses to intensified right activities, far-left movements exemplified by the Antifa, bound by hatred for conservatives and loosely left-oriented ideals of wealth redistribution, social equality, and stringent feminism, adopted equally violent activities in real-life clashes and extreme language expressions in online spaces. Some of Sanders' aggressive reformation proposals also won numerous supporters from the younger generations who are the social media's chief players. Moreover, defensive media campaigns to ward off right extremists often quickly turn into 'witch hunts' against any who do not share the left's values, attacking neutral people with libels of alt-rights, racial supremacists, or anti-feminists and disrupting the peaceful causes of online discussion.

While scholarship on political extremism focused on the role of specific authoritarian leaders (i.e. Adorno et al. 1950), more recent work, especially work focusing on contemporary political developments, is mainly preoccupied with the role of the public. Studies from the previous century are becoming insufficient to analyze present social movements likely due to the development of "New" and even "More" media around the start of this millennium, where user-generated content, rather than centralized releases, has come to dominate the forums (Manovich, 2020). When the internet removed the barricades that had prevented most people from discussing politics, the performance of extremism evolved from planned actions of inner circles to relatively sporadic individual performativity. Consequently, modern examinations of political and cultural trends must focus on decentralized behaviors and publishing with the assistance of computational

tools. Conway, Scrivens, and Macnair (2019) also argued for giving attention to obscure common people than more organized packs of fanatics. They observed that initially the right-wing extremists, as Blee and Creascap described, established websites and relied on forums for community and asylum, but riding the tide of the social media expansion, their influence gradually permeated every internet channel from major social media to messaging apps. The crackdown after the Capitol riots may curb their appearance on Twitter and Facebook for a time but they could easily switch battlegrounds with greater resurgence. To explain this trend, computational tools are needed to dissect and analyze the fundamental causes of extremism's attractiveness, of both lefts and rights combined. Directions such as Spruyt, Keppens, and Droogenbroeck (2016) continued the study of the digital sector but focused on the movements' ideological shifts while neglecting their transmitting techniques. In comparison, we will pay attention to not only what ideas are being communicated but also how they're being communicated.

II. Computational studies of online political extremism

According to reviews of existing literature on automated extremism detection (Aldera et al. 2021, Torregrosa et al. 2021), common accurate methods to identify extreme behaviors use profile-related and time-related features, labeling users and their posts based on known extreme movement accounts they followed and conversations participated. This is indeed a commonly accepted criterion of effectiveness. A string of prior computational efforts paved the way for the solution. For example, Hughes, McCabe et. al (2021) sought to create a robust approach of sampling Twitter users and Tweets to characterize how the Twitter crackdown reduced right-wing influences on the platforms. The study utilized methods first introduced in Barbera et al. (2015), which defined extremism as people's polarized ideological position that can be

determined by their network positions. Utilizing an impressive sample of 150 million tweets, they were able to create a model predicting quantified extremism based on users' profiles of followed political elites. Their sampling strategies as well as its prior study identifying user ideologies provide a good addition to this project's primary focus of performance.

Other than identifying with users' profile data, Torregrosa et al. (2021) stated that extremist narratives bear specific linguistic styles, such as the higher use of first- and third-person plural pronouns, a more negative tone and more words related to negative topics, and cite specific biased sources, making them identifiable with machine learning techniques. Sentiment detection NLP is sufficiently apt in distinguishing the extreme posts as they are proven to be more temperamental. Notably, KhudaBukhsh et. al (2020), using machine learning to describe political polarization, found that the lefts and rights are almost discursive mirrors when debating with each other: substituting the attacked target is sufficient to turn a left slogan into a right one. Using substantial data comprised of 86.6 million comments from 200 thousand YouTtube videos, this study compared languages used by viewers of pro-left and pro-right news networks and found that their choices of adjectives are highly similar when describing each other: the similarity between CNN and Fox News users, though lower than that among networks of the same political camp, reached 89%. In other words, the far-left and far-right movements are simply using different nouns when marking their opponents, but the rest of their lexicons are highly alike in expressing belligerent, exclusionary sentiments. This conclusion provides the option of treating both political leanings' extreme speeches as equivalents when using sentiment-related classification tools. I have defined extremism as a combination of ideology and performance and argued for using the performance to represent the whole concept; the practice was prevalent in all prior research. Despite their varied approaches, these studies had

unanimously chosen to focus on the activities, including gatherings and rallies in the real world and virtual expressions in the digital world.

Frimer et al. (2018) and Alizadeh et al. (2019) are works that treated both left-wing and right-wing extremism and displayed promising directions to work upon yet need to be referred to with caution: the formerly used sentiment-detecting methods but did not see significant variations between extreme and other groups; the latter used crowd-sourced classification data of Twitter users and concluded that extreme speeches are emotionally more negative and angry than those of moderates, effectively distinguishing themselves from ordinary social media users. Alizadeh et al. (2019) also suggested that this result is consistent with the reasonings of the Moral Foundation Theory. The theory was first introduced in Graham et al. (2009) to evaluate political culture using six foundational morals, each having a "virtue" and a "vice"[3]. Alizadeh et al. claimed to provide statistical evidence to Haidt and Graham's argument that the individuals' ideological differences are associated with an emotional resonance which originates from the different moral foundations.

Another supporting study is Wei, Singh, and Martin (2016), which compared the account/profile-based methods to the sentiment analysis method in detecting extremist Twitter posts. Using both English and Arabic tweets with tags related to extremist behavior, the study examined approximately 2 million accounts and built a custom sentiment classifier with unigrams, emoticons, and negation expressions. It found that sentiment analysis tools are precise at 85% accuracy when Naive Bayes and Logistic Regression classifiers are used. Support vector machine is also a reliable classifier option at a rate of 70% while other methods did not suit the

[3] The six main foundations are care/harm, fairness/cheating, loyalty, betrayal, authority/subversion, sanctity/degradation and liberty/oppression.

application very well. It also showed that sentiment analysis is an accurate method whether used on its own or combined with other methods; conversely, the following or mentioned accounts method has hardly over 50% accuracy when used individually, and therefore is unreliable. Still, when used as an addition to other models it does slightly boost accuracy so it is still viable to be included. Chronologically this paper was done before Trump's triumph and subsequent academic attention given to the ultra-right movements; therefore one must note that the authors defined extremism completely differently from this study. Wei, Singh, and Martin were focused on the possible presence of terrorism in the likes of Al-Qaeda and the use of social media to solicit religious extremist attacks and had not the opportunity to apply the methods on the extreme right and left political wings of American politics described here. We have defined the extremism ideology as anti-establishment politics which differs from the religious ideologies in this study. Still, since we have proposed to focus on performance and the religious extremists' digital performance of verbal attacks and seditions are not different from that of political extremists, such results are generalizable references that support my decision to use the combination of account information and post sentiments but give more leverage and focus on the sentiment analysis.

Another frequent approach to analyzing social media broadcasts is modeling based on the notion of contagion. It has been repeatedly argued in Guilbeault et. al (2018) and Goldberg et. Stein (2018), as well as tested using empirical evidence in Sprague et. House (2017), that social media transmissions are usually complex contagions, especially when political topics are involved. Goldberg et Stein (2018) further coined the term "associative diffusion" to describe the process where the recipients consciously interpret and evaluate the message before deciding on the transmission. A practical problem that could occur in the ensuing analysis is that social media

users may have drastically different definitions of political concepts or have insufficient knowledge to tell the subtle differences. This is another reason to support the decision to keep the classification broad, identifying the extreme languages without distinguishing precise political standings. This is to avoid potential confusion and shall be reflected in the statistical treatment. So far, a process of studying extremism performance has been assembled based on successful experience; a relevant aspect should be included to represent the definition's ideology side.

Based on these preconditions, this project shall measure extremism based on these available options. It should first prepare a training set based on clearly defined features and realize a model detecting and classifying extreme speech to treat the raw sampled data, based on which whether extremist performance relates to social media sharing behavior.

Data and Methods

The project will employ two datasets: a main tweets collection randomly scraped from the social media site and an artificial training set generated by Amazon Mechanical Turk surveys.

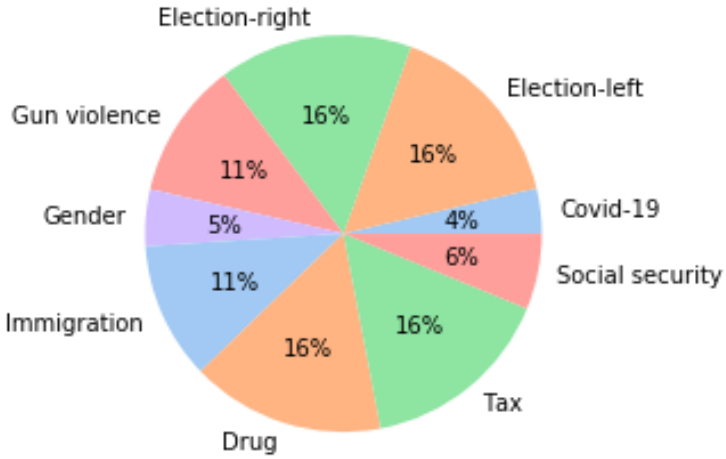
(I). The main dataset uses Tweepy and Twitter API to acquire randomly sampled tweets on specifically selected trending topics that typically reflect political identities. The initial filters are inspired by the 2022 version of World's Smartest Political Quiz. The quiz asked questions mainly on personal rights and economic policies to determine the political identification of those surveyed and we paraphrased these questions into search query inputs, including gun violence, gender, immigration, drugs, tax and social security issues, while supplementing the list with more updated discourses on the Covid-19 and the 2020 Presidential Election. The scraper

searches tweets matching the search query since Jan 1, 2019, until all available tweets are recorded, or a 100-thousand cap is reached for each category. The tweets' texts are scraped along with their Likes and Retweet counts. The set is imbalanced between different topics: more popular topics like covid or election conspiracy theories occupy the overwhelming majority of the set while gun control or drug concerns do not raise enough awareness. To complement the Quiz's topics and get more results, posts associated with more recent 2022 midterm was added to the election category. Overall, the dataset contains several thousand tweets for each category, varied due to disparities in these topics' discussion popularity, and builds a selected dataset consisting of around 600 thousands tweets . This dataset is gathered and treated to produce regression variables. The main dependent variables are the tweets' likes, retweets, and comment count, which reflect their influence levels.

Table 1. Composition of the scraped tweets

Topic	Count
Covid-19	23209
2020 US Presidential Election	200000
Gun violence	70394
Gender/Feminism	29162
Immigration	70717
Drugs	100000
Taxes	100000
Social security	39075
Total	**632557**

Figure 1. Distribution of each topic for Twitter scraping



(II). The purpose of the Mturk-generated dataset is to provide a validation set to examine the possible extremism identifiers. The dataset is produced by an Amazon Mturk survey which asks workers to determine sample tweets' extremism level based on a given instruction[4]. Under my working definition's emphasis on performance over ideology in measuring extremism, the instruction requires workers to only judge tweets based on how their expressions resemble each extremism level's activity without having to contemplate their positions in the political spectrum or their implicit messages. The process is a simplified reproduction of actual digital encounters on the Twitter platform and should provide results mimicking Twitter users' natural reactions. Each tweet will be labeled by multiple workers and an average of their scorings will be recorded. We assume that this crowd-sourced data represents a relatively accurate measure of extremism performance, which can be used to test the following methods.

(I). The first method is sentiment analysis. While Wei, Singh, and Martin (2016) argued for its accuracy in detecting extremism, Frimer et al. (2018) suggested that sentiments only have

[4] See Appendix 1 for the Mturk survey details, including the operating interface on Amazon and instructions for workers.

a loose correlation with extremism and can also be confusing when the moral words method is applied. Suspicions from Frimer et al. (2018) and others is why we implemented the Mturk validation experiment to first examine their relationships: we have to first establish that these machine-generated features predict human-labeled tweet extremism before using them to predict tweet popularity in the main analysis. The basic sentiment measurer is the SentimentDetector classifier in PySpark NLP by John Snow Labs, which returns texts' numerical sentiment values to represent a range from very negative to positive. This is a robust and readily available model trained on a reliable sample and should always work in case an improvised sentiment model using methods from Pennebaker, Booth, Boyd, & Francis (2015) and Frimer et al. (2018) does not work. This model is no different from Spark NLP's SentimentDectector in design: both count words in texts' segments and then use emotion dictionaries to make evaluations. The former offers simple directions on positive or negative, and increase or decrease, while the latter incorporates dictionaries for more subtle emotions of anxiety, anger, and sadness. We use SentimentDetector's negative scoring for the main regression.

(II). The second method is the Twitter user profile search derived from Hughes, McCabe et. al (2021) and Barbera et al. (2015). An issue raised by KhudaBukhsh et al. (2020) is that in fact, ultra-left and right wings are sometimes mirrors of each other, speaking extremist languages alike. The project needs to be equipped with trained models on the political spectrum reflecting languages to be able to tell the difference, with the assistance of profile data. To generate the score representing extremeness based on profiles, we shall rely on a published repository by Stefan McCabe, which has already identified extreme users based on their past profiles, activities, and following status. The transforming method is a Bayesian point estimation which first filters politically active users and takes in their lists of followers and then estimates their

political identification. From the repository, we collect model in Barbera et al. (2015) and generate outputs representing political inclinations from far-left to far-right on a scale from -1 to 1; since we have argued to only focus on the intensity rather than the direction of extremism performance, we will take the scores' absolute value. A potential flaw with the above measure is that it rates these users' tweets based only on their profile without at all considering the text, thus likely overstating the political inclinations of the text.

(III). A third addition for measuring extremism is based on the Moral Foundation method from Alizadeh et al. (2019). Still using the profile dataset for training, this classifier takes into consideration the word usage, phrase structure, and other language features that are commonly observed in more extreme speeches and utilizes NLP models (word-2-vec python package) to derive extremeness based on the posts' text. Having determined the sample's extremism levels using both pure sentiments and profiles, we examine how each moral foundation value using the lexicon in Graham et al. (2009) would make a difference. Similar to the treatment in sentiment analysis, we will use the scoring for negative morals. We calculate the term frequency of the words classified as "vices" in the Moral Foundations Dictionary by Jesse Graham and Jonathan Haidt.

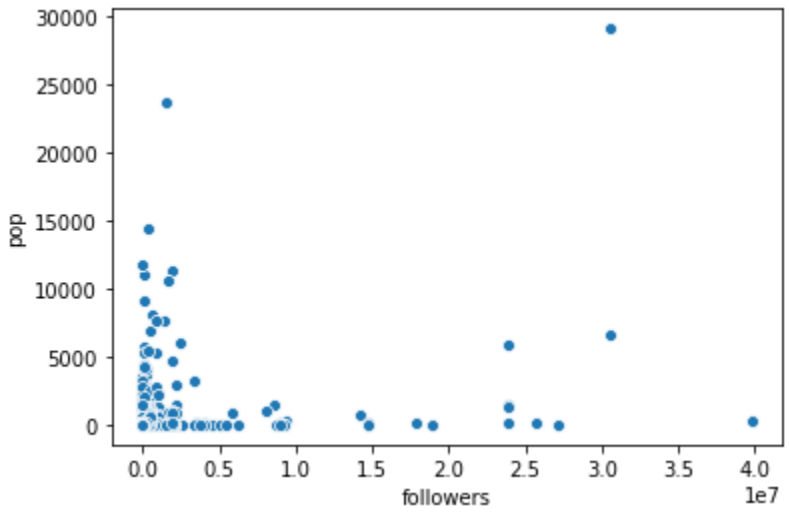
One should take note that the effects demonstrated by the project could be mechanically weaker than the actual effects given the time of data gathering since the turmoil and contention circumventing the 2020 Presidential Election and subsequent 2021 Capitol riots, social media, especially Twitter, has significantly tightened control over the spread of right-wing speeches, including banning Trump himself from using his once frequently used Twitter account. It is obvious that right-wing users and posts are somewhat suppressed and contained, but their influences are far from dwindling.

Summing up these derived features, the main regression will report how levels of extremism impact how well tweet posts are received based on the following structure:

Popularity measure ~ feature (sentiment/profile/moral words) + control

All feature scores have a range of [0, 1]. Each extremism identifier will be separately used to explain popularity as if used all together, they could dilute each other's effects; the popularity measure is the sum of the likes and retweets statistics. Meanwhile, the control will be the follower number of each tweet's author so that it does not confound the result: in the main dataset its correlation with popularity is 0.2673 (P=0.000). Finally, to compare the overall effect sizes of different generated criteria, sentiments, following accounts, and moral words, a general regression involving all extremism values is conducted.

Figure 2. Scatterplot representation showing popularity and follower number.



Results

The validation results of the main feature-generating models' ability to predict human-labeled tweet extremism are displayed as follows:

Table. 2 Each method's accuracy on Mturk dataset measured by Pearson's R with mturkers' score.

Detection method	SentimentDetector	User profile	Moral foundation
correlation	0.05024	0.10468	0.06461
P-value	0.547	0.209	0.438

We expected the sentiment analysis model to produce a similar level of accuracy to that in Wei, Singh, and Martin (2016), so that it confirms Wei, Singh, and Martin's conclusion that sentiment works as a robust detector of extreme performances and give us more confident to represent extremism with this feature in the subsequent analysis. However, the sentiment analysis method as well as the profile-based detection model sourced from Hughes, McCabe et. al (2021) and the moral words-based model based on Graham et al. (2009)'s moral foundation dictionary does not yield sound accuracy. The latter's relative ineffectiveness can be somewhat anticipated: Hughes, McCabe et. al (2021) completely disregards elements in the actual text, generates no result for users with insufficient profiles, and has been proven less accurate than sentiment by Wei, Singh, and Martin; the moral foundations method latter employs a potential dictionary and is relied on Graham et al.'s attaching different political identities to specific moral values in 2009, which could use some updates to better fit digital platforms today. But the equally lacking result for the sentiment analysis method could suggest issues with our validation

dataset itself. Considering the poor accuracy performances, we could not convincingly argue that these features can directly represent extremism in the main analysis.

Table. 3 Each method's accuracy on Mturk dataset measured by Pearson's R with mturkers' score.

	SentimentDetector & User profile	SentimentDetector & Moral foundation	User profile & Moral foundation
correlation	0.25294	-0.01849	-0.02942
P-value	0.017	0.863	0.784

The above summarization of correlations between the measures in the MSturk dataset backs up our decision to include both individual regressions and a regression with all three measures. We can at least observe that the sentiment score and the profile score are correlated at a significant level and could confound each other. While using all three predictors simultaneously are at risk of having diluted effect sizes, we also need to address the issue that the individual effect size gets overestimated, and thus we will compare results from both regression sets.

Table. 4 Results for predicting post popularity using Mturk data.

predictor variables	*Regression using Mturk validation set*
SentimentDetector	18.8103
User profile	19.3715*
Moral foundation	-295.3069
R^2	0.081
Adjusted R^2	0.049

*p<0.05, **p<0.01, ***p<0.001

We also conducted a preliminary regression using the Mturk dataset. Under this setup, the profile score displays a sound effect size and is useful in predicting popularity. Therefore, the profile-based model does help cement our primary definition of extremism as a duality of both ideology and performance and our operation to use one's performance to represent the underlying ideology. Despite its lower accuracy, the profile-based model nevertheless produces significant results showing that users' posts display political inclinations related to the accounts they regularly visit. This provides evidence that digital performances are steady reflections of the performers' inner ideology.

Table. 5 Results for predicting post popularity based on different setups.

predictor variables	*SentimentDetector only*	*User profile only*	*Moral foundation only*	*All included*
SentimentDetector only	21.7541*			1.459E-18***
User profile only		70.7928		5.119e-05***
Moral foundation only			3.35833	7.461E-18***
R^2	0.000	0.000	0.000	0.006
Adjusted R^2	0.000	0.000	0.000	0.006

*p<0.05, **p<0.01, ***p<0.001

The general regression shown in Table 5[5], involving all three extremeness values, are done in addition to individual regressions. This resulting effect sizes in this broader analysis are, despite yielding good P-values, minute with tiny parameters, which we have anticipated prior to the approach. Therefore, we should focus on the individual results where each identifier's effect is better explained. In Table 2, the regression results using sentiment scores present the relatively more significant figures confirming the extremeness's positive impacts on the click statistics.

[5] Exact P-values can be found in Appendix 2

The individual analysis using sentiments still yields interpretable results while the other two fail to compile significant results. What's worth mentioning about the individual trial is applying the accounts-based model on the selected renowned accounts set. This regression does show readable figures congruent with the results of the main analysis. One must be reminded, though, that as Frimer et al. had shown, sentiment is a powerful tool but alone will confuse some non-extreme speeches for their strong emotions.

Generalizing from these findings, some more propositions can be suggested: posts using aggressive, rousing speech attain more attention on social media. Misclassifications do cause some models to function not so well on their own but the main analysis using all three generated scores performed suitably. Since none of the methods was successful in the Mturk validation experiment, we can only say that strong, negative sentiments are associated with social media popularity. Still, the profile score can meaningfully predict popularity in the Mturk dataset; though it failed to match our definition of extremism, under the notion of network position in Barbera et al. (2015) it is a robust extremism identifier and therefore provides evidence that extreme positions are related to popularity as well.

Discussion

The presence of extreme speeches is prevalent in today's social media sites and the above computational results suggest that more extreme and exaggerated materials are more likely to grab the attention of general internet users. Many past studies explored various methods to detect extreme language performance, which provided the foundation for answering our question of extremeness's association with popularity. Most notably we have developed the results in Wei,

Singh, and Martin (2016) and Hughes, McCabe et. al (2021) and provided a relatively straightforward conclusion: first we confirmed Wei, Singh, and Martin's religious extremism detection model's validity when applied to political extremism and then it was used to generate one of the main regression features that led to the final result; user profile based detection model in Hughes, McCabe et. al (2021), despite not providing similar levels of statistical importance in the main regression, showed some degree of relationship between Twitter users' political propensity and accounts frequented by them, and therefore at least reinforcing the proposition that their digital performance reflects their internal ideologies, on which the whole inference is based. Through this progression of precedent detection models, we attempted to fill the gap between detecting extremism and evaluating extremism's effect. Contrasting to the studies made when digital methods were unmatured like Blee and Creasap (2008), our findings suggest that extremism could be transmitted without a leading figure, organized network, or iterated political doctrines; some rhetorical devices in the texts are powerful enough to draw quite a few followers. On the other side, we took advantage of all prior efforts to quantify and digitalize the abstract concept of extremism, such as Barbera et al. (2015). Wei, Singh, and Martin (2016), Goldberg et Stein (2018), Conway, Scrivens, and Macnair (2019), and Hughes, McCabe et. al (2021), to construct scaffolding to explore the relationship between extreme performance and corresponding popularity. As for supporting arguments including McClosky and Chong (1985), and KhudaBukhsh et. al (2020), which enabled us to ignore the difference between the left-wing and right-wing extremists, our results could not further approve or withdraw their conclusions. Additional experiments focusing on this aspect of the question are required to make more validations.

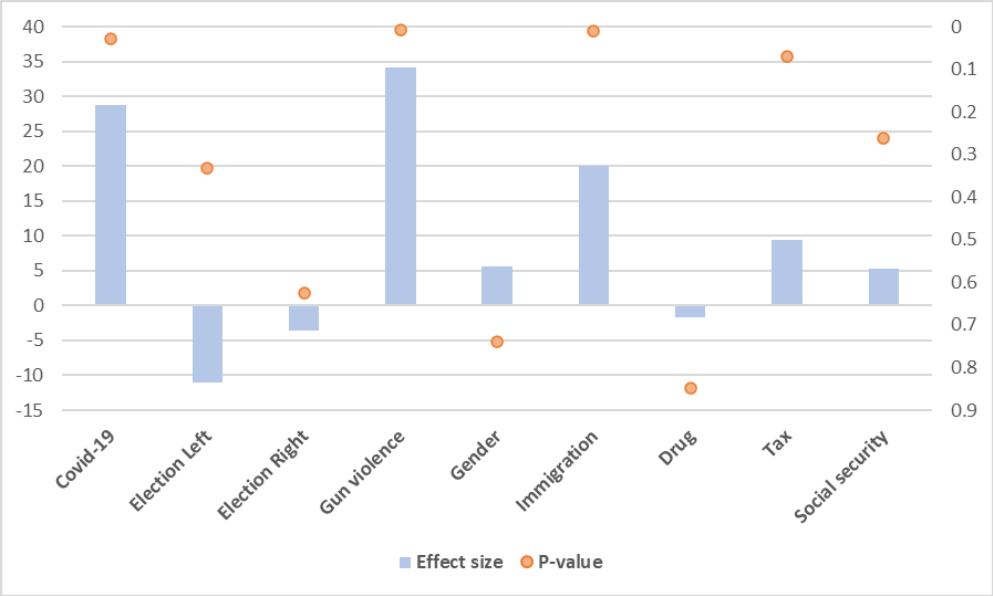
A few things are neglected in the study, some out of practicality. Most importantly, we will have to establish the connection more rigorously between such sentiments and extreme ideology. The shortage of the Mturk validation process should mainly be attributed to the insufficient survey scale: due to limited resources, for each text in our Mturk sample we were only able to assign five unique workers to determine its extremism scoring, making the survey result highly scattered and easily swayed by biases of individual workers. Should a larger crowdsourcing process is made possible, I think the more robust validation result should at least support the conclusions in Wei, Singh, and Martin (2016). Second, the practice of not discerning the difference between the extreme left and right, though supported by prior literature and making the analysis simpler, can be tailored with more details. Admittedly, the damage caused by ultra-left parties is nowhere near that of the ultra-right; notable left-wing violence often is protests and gatherings left unchecked and turned into street brawls and shop break-ins while ultra-rights, mostly against gun control and bearing arms themselves, can set up militias capable of mass shootings and even more uncontrollable accidents[6]. Right-wing online extremism more easily evolves into an event of destructive consequences in the real world, so it is justified to treat the extreme speeches of the right separately as greater physical threats. Second, the topics selected for extremism identification are from World's Smartest Political Quiz, and they enjoy different levels of attention on social media recently: COVID, which is still an unresolved public health setback, as well as the major political events like the Capitol attack hearings and 2022 Midterms, are a lot more discussed than somewhat outdated controversies of gender or slightly complicated policy issues of taxes and healthcare. Therefore, the dataset is unbalanced with a lot more observations on simple, recent events but not many serious debates on national policies.

[6] See the online news reference "Armed with AR-15s" on extremists and militias accumulating small arms and their threats.

Still, this should not impair the analysis since, as mentioned above, the left and right are large of the same sentiments and are only distinguishable by their stance towards particular affairs. This dataset composition improves the detection accuracy because when addressing these contentious issues people use more emotional and sentimental words detectable by our models while they appear more composed on complex matters. We made additional regressions using sentiment scores on each one of the data's subsets, as shown in Figure 2[7], to compare how tweets from different topics contribute to the overall effect sizes. Clearly, some topics yielded larger and more significant effect sizes and further explorations exquisitely studying specific topics and audience groups are possible. Another issue was that we could not base everything in the project on the most up-to-date materials and had to rely on several prior studies whose results do not necessarily reflect the latest post-COVID social media environment. Notably the relatively aged moral words lexicon from 2009, contains words and expressions too archaic for the social media scene, such as "wonton" or "profligate". Stewart and Morris (2021) proposed that group and individual level morals function differently when influencing ideological identities, and Kalimeri et al. (2019) complemented the six main foundations with the Schwartz human values (Schwartz, 1992), which they considered to better describe digital behaviors. Incorporating these more recent findings could help produce explicable results. Further expansion of this project will more robustly enhance the results and thus provide more sophisticated case-by-case solutions.

Figure 3. Result using sentiment scores on each topic.

[7] Also see Appendix 2c for exact numbers of these coefficients



Additional questions can be asked based on this result's commonplace social impacts. For instance, we can observe for some time that journalism has been using extreme language in news titles and headlines in a similar fashion to these extreme digital performances to attract clicks and subscriptions; perhaps journalists at traditionally prestigious national and international newspapers and agencies would at times regain conscience and shun such behaviors, but reporters employed solely to further political parties' interests and numerous self-made columnists aiming to only cash in media influence continue the practice with no restraint. While the decentralized digital media continue to expand, traditional media do not simply fade away; our study so far has demonstrated the effect on individual social media posts which mostly have no frequent viewers, but one can easily imagine it being tremendously amplified when utilized by large multi-media institutions with numerous subscribers. Therefore, a more resourceful extension of the project could be focusing on known news agencies on multiple media platforms. The follow-up question is whether journalism should be held up to more social responsibilities than individual posts and thus face more supervision when exploiting extremeness's popularity.

The study may also be further extended on the aspect of mechanism with approaches of psychology. Social network analysis can be useful to trace how a particular explosive post is transmitted rampantly across many individuals and psychological surveys can be conducted to examine the thinking behind a single receiver's decision to ignore, approve, or spread these posts. This generation of social media audience seems to defy Duncan Black's characterization of the typical median voter (Duncan 1948), which has been a compelling proposition for political and economic thinking for a lengthy period. According to the proposition, as well as a larger portion of the US's election history, candidates would usually approach a compromising middle-ground complying with the majority's interests to win the most supporters possible, but the present reality sees political figures at both wings content with rallying smaller packs of more staunch proponents. Supposedly, this violence of the proposition could still be coherent with the underlying principle of minimum differentiation by Harold Hotelling (1929): within the opposing political camps, there still exist "medians" to draw as many followers as possible, as one can observe in the proximity between ideologies of Trump and his competitors within the Republican Party with the likes of John McCain losing hold. Specifically designed surveys and experiments directed by trained psychologists may be carried out to explain the individual and environmental reasons behind this mental transition from "median voters" to more extreme voters and will greatly assist in explicating the phenomenon summarized in this project.

Lastly, that the ordinary internet audience welcomed extreme materials shall not justify their rampant existence. The results showed that people are drawn to these posts for the content without the interference of leading figures. Charismatic leaders are not required for a diaspora of extremism to flourish; the decentralized yet polarized internet environment is an arid woodland where dropping small sparks causes raging wildfires. Therefore, targeting and silencing iconic

figures of the right-wing is only counterproductive, for they can be easily substituted; it may even irritate the followers and further exacerbate their hatred. For the ultra-left, it is simply difficult to name one such leading figure. In extremism movements, the common follower is also one's leader, consuming and producing extreme speeches at the same time. Defendants may invoke the Hegelian "was wirklich ist, das ist vernünftig" (that which is real is reasonable) and argue that they should be incorporated into today's way of life. But as the adverse effects continue to surface, causing both domestic social discord and international isolationism, I think every relevant institution from government regulators and legislators to the social media tech companies and user communities should take awareness and measures to combat this tendency. The previous actions by Twitter and Facebook to ban specific accounts of leading figures will not have lasting effects due to the extreme speeches' decentralized, vulgar nature enabling their circulation at the common people's level requiring no leaders. An upgrade to the automated web censors using the extremism identifying models similar to those in this study should effectively remove these speeches but will not only severely impact the spirit of freedom of speech but also prolong the information circulation process, discouraging users and significantly making social media maintenance more expensive; therefore, the companies are not expected to take such a drastic measure, which itself is an embodiment of censorship and extremism. Even with an impartial machine identifier, it is far from convincing to simply call someone's post inflammatory and thus issue a ban. The hope may lie in encouraging rather than forbidding specific tactics: in a spirit similar to "subsidiary over tax", bonuses can be set up to reward users with continuous good communal behavior, which should discourage the more numerous group of people with vague political inclinations from participating in extreme content creation. Contracted content authors and video makers on social media earning profits through the

platforms should be obliged to produce materials promoting a tolerant, healthy internet atmosphere. To demand all these efforts from major social media companies would require the participation of regulators and legislative support.

References

Advocates for Self-government quiz for political identities, The World's Smallest Political Quiz, Advocates for Self-government, www.theadvocates.org/quiz/

Aldera, S. , Emam, A. , Al-Qurishi, M. , Alrubaian M. and Alothaim, A. , "Online Extremism Detection in Textual Content: A Systematic Literature Review," in IEEE Access, vol. 9, pp. 42384-42396, 2021, doi: 10.1109/ACCESS.2021.3064178

Alizadeh, M., Weber, I., Cioffi-Revilla, C. et al. Psychology and morality of political extremists: evidence from Twitter language analysis of alt-right and Antifa. EPJ Data Sci. 8, 17 (2019), doi: 10.1140/epjds/s13688-019-0193-9

Barberá, Pablo. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." Political Analysis 23, no. 1 (2015): 76–91, doi: 10.1093/pan/mpu011

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? Psychological Science, 26(10), 1531–1542, doi: 10.1177/0956797615594620

Black, Duncan. "On the Rationale of Group Decision-Making." Journal of Political Economy 56, no. 1 (1948): 23–34, www.jstor.org/stable/1825026.ttps://doi.org/10.2307/2224214.

Blee, Kathleen M., and Kimberly A. Creasap. "Conservative and Right-Wing Movements." Annual Review of Sociology 36 (2010): 269–86, www.jstor.org/stable/25735078

Bötticher, Astrid. "Towards Academic Consensus Definitions of Radicalism and Extremism." Perspectives on Terrorism 11, no. 4 (2017): 73–77, www.jstor.org/stable/26297896

Conway, Maura, Ryan Scrivens, and Logan Macnair. "Right-Wing Extremists' Persistent Online Presence: History and Contemporary Trends." International Centre for Counter-Terrorism, 2019, www.jstor.org/stable/resrep19623

Doxsee, Catrina, Seth G. Jones, Jared Thompson, Grace Hwang, and Kateryna Halstead. "Pushed to Extremes: Domestic Terrorism amid Polarization and Protest." Center for Strategic and International Studies (CSIS), 2022, www.jstor.org/stable/resrep41416

Drake, Richard. "Left-Wing Populism Meets 'La Grande Crisi.'" New Labor Forum 27, no. 2 (2018): 56–63, www.jstor.org/stable/26503625

Frimer, Jeremy A., Mark J. Brandt, Zachary Melton, and Matt Motyl. "Extremists on the Left and Right Use Angry, Negative Language." Personality and Social Psychology Bulletin 45, no. 8 (August 2019): 1216–31, doi: 10.1177/0146167218809705

Gandesha, Samir. "Understanding Right and Left Populism." In Critical Theory and Authoritarian Populism, edited by Jeremiah Morelock, 9:49–70. University of Westminster Press, 2018, doi: 10.2307/j.ctv9hvtcf.7

Goldberg, Amir, and Sarah K. Stein. "Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation." American Sociological Review, vol. 83, no. 5, 2018, pp. 897–932. JSTOR, https://www.jstor.org/stable/48588676. Accessed 20 Apr. 2023

Gounari, Panayota. "Authoritarianism, Discourse and Social Media: Trump as the 'American Agitator.'" In Critical Theory and Authoritarian Populism, edited by Jeremiah Morelock, 9:207–28. University of Westminster Press, 2018, doi: 10.2307/j.ctv9hvtcf.13

Graham, Jesse, Haidt, Jonathan, Nosek, BA. "Liberals and conservatives rely on different sets of moral foundations." J Pers Soc Psychol. 2009 May;96(5):1029-46, doi: 10.1037/a0015141

Hegel, G. W. F., *Grundlinien der Philosophie des Rechts* (Frankfurt am Main: CreateSpace Independent Publishing Platform) 2013, 24.

Hotelling, Harold. "Stability in Competition." The Economic Journal 39, no. 153 (1929): 41–57.

Hughes, Adam G, McCabe, Stefan D, Hobbs, William R, Remy, Emma, Shah, Sono and Lazer, David M J, "Using Administrative Records and Survey Data to Construct Samples of

Tweeters and Tweets", Public Opinion Quarterly, Volume 85, Issue S1, 2021, Pages 323–346, doi: 10.1093/poq/nfab020

Kalimeri, Kyriaki, Beiró, Mariano G. , Delfino, Matteo, Raleigh, Robert and Cattuto, Ciro, "Predicting demographics, moral foundations, and human values from digital behaviours", Computers in Human Behavior, Volume 92, 2019, Pages 428-445, doi: 10.1016/j.chb.2018.11.024

KhudaBukhsh, Ashiqur R, Rupak Sarkar, Mark S. Kamlet and Tom M. Mitchell, "We don't speak the same language: Interpreting Polarization through machine translation", arxiv.org/pdf/2010.02339.pdf

Klikauer, Thomas, "Review of Germany's New Populist Party: The AfD, by Marcus Bensmann, Stephan Grigat, Michael Wildt, and Gerd Wiegel", German Politics & Society 36, no. 4 (2018): 78–97, www.jstor.org/stable/48561517

Leeker, Martina, et al. "Performativity, Performance Studies and Digital Cultures." Performing the Digital: Performance Studies and Performances in Digital Cultures, edited by Martina Leeker et al., Transcript Verlag, 2017, pp. 9–18, JSTOR, www.jstor.org/stable/j.ctv1xxsxb.3

Manovich, Lev, *Cultural Analytics*, MIT Press, 2020.

McClosky, Herbert, and Dennis Chong. "Similarities and Differences between Left-Wing and Right-Wing Radicals." British Journal of Political Science 15, no. 3 (1985): 329–63, www.jstor.org/stable/193697

Prier, Jarred. "Commanding the Trend: Social Media as Information Warfare." Strategic Studies Quarterly 11, no. 4 (2017): 50–85, www.jstor.org/stable/26271634

Schroeder, Ralph. "Digital Media and the Rise of Right-Wing Populism." In Social Theory after the Internet: Media, Technology, and Globalization, 60–81. UCL Press, 2018, doi: 10.2307/j.ctt20krxdr.6

SentimentDetector information, Spark NLP - Annotators, John Snow Labs, https://nlp.johnsnowlabs.com/docs/en/annotators#sentimentdetector

Schwartz, Shalom H. , “Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries”, Advances in Experimental Social Psychology, Volume 25, 1992, Pages 1-65, ISSN 0065-2601, doi: 10.1016/S0065-2601(08)60281-6

Sloman, Sabina J. , Oppenheimer, Daniel M. and DeDeo, Simon, “Can we detect conditioned variation in political speech? two kinds of discussion and types of conversation”, PLoS ONE 16(2): e0246689, doi: 10.1371/journal.pone.0246689

Spruyt, Bram, Gil Keppens, and Filip Van Droogenbroeck. “Who Supports Populism and What Attracts People to It?” Political Research Quarterly 69, no. 2 (2016): 335–46, www.jstor.org/stable/44018014

Stewart, Brandon D. and Morris, David S. M, “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, doi: 10.3389/fpsyg.2021.579908

Twitter API structrue, Tweepy documentation, https://docs.tweepy.org/en/stable/api.html?highlight=search

Twitter ideology and user behavior, in data repositories by Stefan McCabe used for Hughes et al. (2021), Estimating Ideological Positions with Twitter Data, Barberá, Pablo, https://github.com/sdmccabe/twitter_ideology

Torregrosa, J., Bello-Orgaz, G., Martínez-Cámara, E. et al. A survey on extremism analysis using natural language processing: definitions, literature review, trends and challenges. J Ambient Intell Human Comput (2022), doi: 0.1007/s12652-021-03658-z

Wei, Y. , Singh L. and Martin, S. , “Identification of extremism on Twitter,” 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2016, pp. 1251-1255, doi: 10.1109/ASONAM.2016.7752398

Word2vec information, Word2vec embeddings, Gensim, https://radimrehurek.com/gensim/models/word2vec.html

Zhaksylyk, K. , Batyrkhan, O. and Shynar, M. , "Review of violent extremism detection techniques on social media," 2021 16th International Conference on Electronics Computer and Computation (ICECCO), 2021, pp. 1-5, doi: 10.1109/ICECCO53203.2021.9663813

Online News:

Allam, Hannah, "Armed with AR-15s, extremist and militia groups anticipate civil unrest", The Washington Post, March 27, 2023, https://www.washingtonpost.com/nation/interactive/2023/ar-15-armed-extremist-militia-groups/

Goodwyn, Wade, "Alt-Right, White Nationalist, Free Speech: The Far Right's Language Explained", *WBEZ Chicago*, NPR, June 4, 2017, https://www.npr.org/2017/06/04/531314097/alt-right-white-nationalist-free-speech-the-far-rights-language-explained

Kleinfield, Rachel, "The Rise in Political Violence in the United States and Damage to Our Democracy", Carnegie Endowment for International Peace, March 31, 2022, https://carnegieendowment.org/2022/03/31/rise-in-political-violence-in-united-states-and-damage-to-our-democracy-pub-87584

St. Aubin, Christopher and Stocking, Galen, "Key facts about Gab", *Pew Research Center*, January 24, 2023, https://www.pewresearch.org/short-reads/2023/01/24/key-facts-about-gab/

"The Left and the Right Speak Different Languages—Literally", Wired, https://www.wired.com/story/left-right-speak-different-languages-literally/

Thompson, Stuart A. and Goldstein, Matthew, "Truth Social's Influence Grows Despite Its Business Problems", *The New York Times*, November 1, 2022,

https://www.nytimes.com/2022/11/01/technology/truth-social-conservative-social-app.html

Replication materials for this project:

https://github.com/PAHADRIANUS/sm_extremism_shuyi_yang

Appendix

1. The Amazon Mechanical Turk process, including the operating interface and the detailed instruction for the Mturk workers.

We drew a random sample of 200 tweets from our scraped dataset and directed 5 workers to rate each tweet. A budget was spent to pay $0.05 each for the 1000 workers, with additional operating fees charged by Amazon. We requested that workers have to be Twitter account holders.

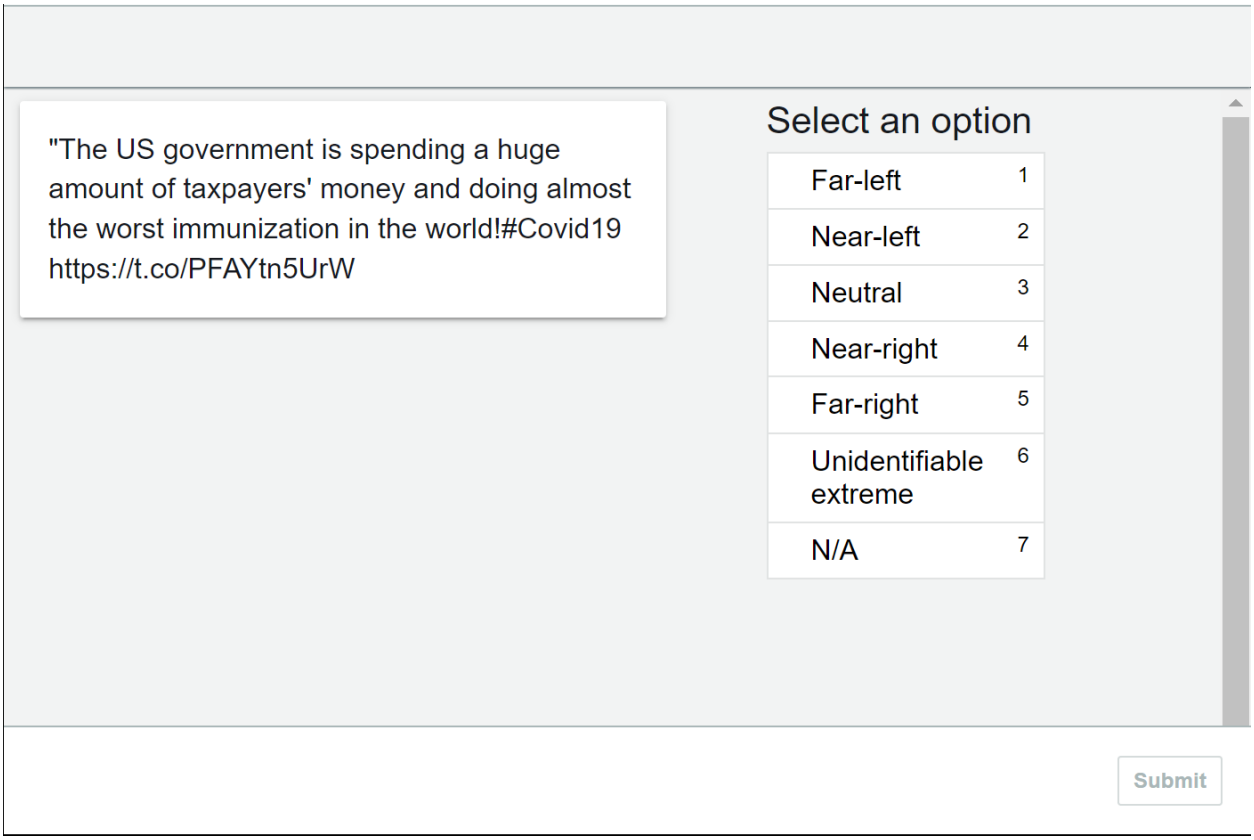
The batched results return the labels that workers consider to best represent the given text, and in the subsequent data cleaning we transform the labels into linearly distributed values between 0 and 1 to match calibers of other scores (sentiment, profile and moral words) already generated. Same magnitude of political inclination (such as far-left and far-right) has the same value. The final value for each text is the average of the results from the 5 workers.

Assigning values to crowd-sourced labels

Label	Score
Unidentifiable extreme	1

Far-left	1
Far-right	1
near-left	0.5
near-right	0.5
neutral	0
NA	discarded

The survey interface



Political Extremism Identification Task Instructions

Far-left texts describe opinions demanding social reforms based on some form of socialist ideology and display strong resentment towards the established elites. In the US context, please select this if the text is associated with various radical progressive movements such as feminist groups, Antifa, etc. and shows some signs of violence.

Near-left texts describe opinions resembling mainstream Democrats' agendas, such as stronger government, tax reforms, climate change policies, etc.

Neutral texts show neither positive nor negative attitudes towards events, merely stating a fact.

Near-right texts describe opinions resembling mainstream Republicans' agendas, such as weaker government interventions, tax reductions, and cultural traditionalism etc.

Far-right texts describe opinions supporting conservative policies such as trade barriers, xenophobia, adherence to religion, etc., and display resentment against feminism, abortion, minority groups, and immigrants. In the US context, select anything that supports MAGA, Trump 2020 victory, anti-vaccination, etc.

Unidentifiable extreme is picked when you cannot decide the direction.

N/A: when the text cannot be understood

Note 1: When the text's opinions are mixed and you can not be sure, just select neutral and there is no need to discern carefully between near-left or near-right. These two options are spares.

Note 2: If the text display signs of both far-left and right signs, or simply general populism, just select unidentifiable extreme.

2. Extended regression results

a). Result for predicting post popularity in a general regression using Mturk data.

Detection method	SentimentDetector	User profile	Moral foundation
coefficient	17.6086	20.3102	-295.3069
P-value	0.658	0.014	0.300

b). Each method's result for predicting post popularity when only one method is used in the regression.

Detection method	SentimentDetector	User profile	Moral foundation
coefficient	21.7541	18.0872	70.7928
P-value	0.025	0.567	0.295

c). Result for predicting post popularity in a general regression including all methods.

Detection method	SentimentDetector	User profile	Moral foundation
coefficient	1.459e-18	5.119e-05	7.461e-18
P-value	0.000	0.000	0.000

d). Result for predicting post popularity in a general regression including all methods.

Topic	Effect size	P-value
Covid-19	28.7267	0.028
Election Left	-10.9772	0.331
Election Right	-3.5509	0.626
Gun violence	34.1276	0.008
Gender	5.6761	0.739
Immigration	19.9874	0.011
Drug	-1.739	0.848
Tax	9.4933	0.069
Social security	5.2589	0.262