Journal of Quantitative Description: Digital Media 4(2024), 1–53 10.51685/jqd.2024.009

Partisanship and Risk Talk on Twitter: A Multipronged Analysis of the Prominence, Targets, and Drivers of Risk-Related Expression by Democrats versus Republicans during the COVID-19 Pandemic

YINI ZHANG
University at Buffalo, USA

JODY CS WONG
RAND Corporation

ZIJIAN AN
MAXIMILIAN J. BRIMMER
KENNETH JOSEPH
JANET YANG
University at Buffalo, USA

There is abundant evidence that political polarization is associated with perceptions, beliefs, and behaviors related to the COVID-19 pandemic. This article turns to online expression and investigates the prominence, targets, and drivers of risk talk about the pandemic from Democrats and Republicans on Twitter/X, combining a user-level analysis, a content-level analysis, and manual coding of key accounts. We find that risk talk accounted for a greater share of pandemic-related discussion for Republicans compared to Democrats. Also, the specific targets of risk talk differed: Republicans focused more on risks related to public health

Yini Zhang: yzhang20@buffalo.edu
Date submitted: 2023-09-12

guidelines and policies, vaccines, science, and the information environment, whereas Democrats emphasized risks associated with the coronavirus, politicians, and the economy. Furthermore, risk talk on both sides was spearheaded by news media, politicians, and activists; yet, this tendency to retweet opinion leaders was stronger among Republicans. These findings reveal nuanced partisan differences in risk-related opinion expression regarding the pandemic and point to the critical role of media and political elites within partisan groups in driving such opinion expression. Methodologically, our findings demonstrate how a multipronged analytic approach can help us identify meaningful expression patterns on social media.

Keywords: partisanship, risk, social media, public opinion, COVID-19

In an era of deepening political polarization in the United States (Finkel et al., 2020; Iyengar et al., 2019), the COVID-19 pandemic represents the most recent and prominent example of the politicization of a non-political issue (Zhou et al., 2023). Using survey and behavioral data, social scientists have extensively documented how partisanship is entwined with the public's response to the COVID-19 pandemic. Compared to Republicans, Democrats were more likely to perceive risks from the COVID-19 virus, express skepticism toward politicians, criticize government response, and trust preventative measures and medical experts (Allcott et al., 2020; Barrios & Hochberg, 2020; Bruine de Bruin et al., 2020; Gratz et al., 2021; Kerr et al., 2021). These beliefs translated into differential behavioral responses to the pandemic by partisans, with Democrats being more likely to practice social distancing (Allcott et al., 2020; Barrios & Hochberg, 2020; Gollwitzer et al., 2020; Gratz et al., 2021), wear masks (Milosh et al., 2021), comply with stay-at-home orders (Clinton et al., 2021; Goldstein & Wiedemann, 2022), and receive COVID-19 vaccines (Ye, 2023). Poignantly, COVID-19 death rates in Republican counties became higher than those in Democratic counties as the pandemic progressed (Chen & Karim, 2022). These psychological and behavioral discrepancies (Gadarian et al., 2021) are closely related to information seeking and reasoning patterns in a partisan media environment (Barrios & Hochberg, 2020; Lorenz et al., 2023; Moon et al., 2022; Motta & Stecula, 2023; Peterson & Iyengar, 2022; Xu & Margolin, 2024).

With social media becoming a significant barometer of public opinion (Zhang et al., 2022), we turn our attention to expression patterns by vocal Democrats and Republicans on Twitter (rebranded as X in 2023). With access to all pandemic-related expression on Twitter, we quantify the relative emphasis that different partisan groups place on pandemic-related risks (i.e., the prominence of risk talk within all pandemic-related expression), which allows us to understand the relative importance of risk talk. We further unpack what kinds of risks warrant partisans' attention (i.e., the targets of risk talk) and which opinion leaders influence their risk concerns (i.e., the drivers of risk talk). Notably, in this context, a driver of risk talk is a Twitter/X account that is frequently retweeted by partisan users. These results provide nuanced understandings of the observed psychological and behavioral patterns among US partisans. Such visible public discourse on social media is also worth studying because it may further affect individual beliefs and attitudes, media coverage, and political responses (Bridgman et al., 2020; Nguyen & Nguyen, 2020). Ultimately, a deeper understanding of partisan differences in risk talk illuminates what influences partisans' attitudes and behaviors and prepares us to address the partisan divide more effectively in future public health crises.

By investigating differences in partisan discourse on various pandemic-related risks, from the virus itself to the economic impacts of shutdowns, our work further fills a gap in existing research by leveraging social media data to understand how both elite and rank-and-file partisans respond to the COVID-19 pandemic. Some studies focus on elite discourse on COVID-19, examining how politicians talked about COVID-19 in general (Green et al., 2020; Guntuku et al., 2021; Zhang et al., 2023), economic risks from the pandemic (Zhong & Broniatowski, 2023), mask-wearing (Shin et al., 2022), and vaccines (Zhou et al., 2023). Other studies concern public response to the pandemic, investigating general risk topics in public discourse (Park et al., 2021; Qiao et al., 2023; Wang et al., 2020). To date, however, no research has focused on understanding partisan disparities in discussions of risks related to the pandemic. By concentrating on partisan differences in risk talk, we aim to reveal the expression patterns that characterize the online partisan information ecology.

To better understand how partisans talked about pandemic-related risks, we adopt a multipronged analytic approach that combines user-level and content-level analyses. To our knowledge, only a few studies have examined partisans' social media discussions in terms of information sharing, echo chamber interaction, attitude toward science, and vaccine hesitancy (Haupt et al., 2021; Jiang, Ren & Ferrara, 2021; Jiang, Su, et al., 2021; Rao et al., 2021). We follow this approach while supplementing it with a manual coding of opinion leaders.

Our analysis of U.S. partisans' risk talk using a random 1% sample (over 10 million) of pandemic-related tweets consists of three steps: 1) applying two different machine learning methods to cluster users according to their retweeting patterns to identify partisans; 2) using another machine learning method with BERT embeddings to classify tweets into different risk types and targets; and 3) conducting manual coding of actors driving risk talk and running statistical analyses. In what follows, we first introduce our definition of risks and then discuss how the polarized political climate has contributed to differential perceptions and evaluations of risks related to the COVID-19 pandemic.

Literature Review

Risk Perception

Risk is the likelihood of something bad happening to people, entities, or society at large (Sjöberg et al., 2004). It comprises not only technical parameters and probabilistic estimates, but also psychological, social, and cultural contexts (Slovic, 2000). Risk perception is therefore molded by individual and social factors that influence how people react to risks (e.g., Wachinga et al., 2013). This well-established theoretical framework is detailed in the psychometric paradigm (Slovic, 1987), which postulates a taxonomy of hazards depicting how the lay public might perceive different risk types (Fischoff et al., 1978).

A basic premise of the psychometric paradigm is the significant relationship between the perceived need for regulation of a risk event or hazard and two primary types of risk—dread risk and unknown risk. Characteristics of dread risk include perceived uncontrollability, bearing a catastrophic or fatal outcome, and having an inequitable distribution of risks and benefits. Characteristics of the unknown factor include delayed manifestation of probable harm and novelty (Slovic, 1987). The public responds not only to the scientific assessments of a risk event or hazard, but also to the subjective features of risk (the dread and unknown risks) in ways that heighten or lessen their concern. Individuals want to see the risk attenuated, which motivates demands for stricter regulation (Clahsen et al., 2018; de Vries et al., 2019).

Overall, risks are tied to specific targets, such as vaccination and the economy during the pandemic, resulting in different concerns and priorities (Warren & Lofsteadt, 2021; Wong & Yang, 2021). People may identify different risk targets to form risk perception when evaluating the unfamiliar and immediate risk from COVID-19. Based on existing research, these risk targets may include virus transmission (e.g., Barrios & Hochberg, 2020), vaccination (e.g., Jiang, Su, et al., 2021), public health guidelines (e.g., Clinton et al., 2021; Milosh et al., 2021), the economy (e.g., Zhong & Broniatowski, 2023), pandemic-related misinformation (e.g., Zhang et al., 2023), and more.

Partisanship and Risk Talk

Partisanship in the United States has undergone dramatic changes over the past several decades. Since the 1970s, ideological orientation and partisan affiliation have converged, with liberal voters moving towards the Democratic Party and conservatives towards the Republican Party (Mason, 2015). With partisan-ideological sorting comes meaningful differentiation between these two parties across a variety of policy issues, ranging from abortion to presidential approval (Abramowitz & Saunders, 2008). Such ideological polarization is well-documented among political elites (McCarty et al., 2016) and the general public, especially among those with greater political sophistication (Claassen & Highton, 2009). Affective polarization is also observed, manifesting in how partisans view counter-partisans negatively and co-partisans positively (Iyengar et al., 2019). These trends, along with an increasingly partisan media ecosystem (Benkler et al., 2018), have culminated in political sectarianism—"the tendency to adopt a moralized identification with one political group and against another" (Finkel et al., 2020, p.533).

In this sense, partisanship has become a social identity and colors perceptions of risks, which may be evidenced by differences in their risk talk. According to social identity theory (SIT), one's self-concept not only hinges on personal identity, but also on the social contexts in which one is embedded (Tajfel & Turner, 1979). SIT argues that through the process of social categorization, we come to view those around us largely through the lens of "us" (i.e., in-group) versus "them" (i.e., out-group). This sense of belonging to a particular "us," or group membership, in turn, can fulfill a variety of goals: from an innate need to belong (Baumeister & Leary, 1995) to alternative goals, such as self-enhancement (Tajfel, 1982), coherence (Hogg & Abrams, 1988), and distinctiveness (Brewer, 1991). However, group membership also elicits natural cognitive biases, leading individuals to extend preferential treatment to their in-groups regarding perceptions, evaluations, and behaviors (Van Bavel & Pereira, 2018; Waytz & Epley, 2012). For instance, people may be more likely to practice sanitation routines such as washing hands during a pandemic if they know or presume that their in-group partakes in these behaviors (Goldring & Heiphetz, 2020).

Furthermore, when social identity is salient, members of a group can be significantly influenced by its leaders. As a result, expression can be shaped by opinion leadership dynamics. Within a given social group, "prototypical in-group members" emerge as leaders, who epitomize the group identity and are “perceived to be the most reliable source of normative information and thus effectively have disproportionate influence over the identity and behavior of group members” (Hogg et al., 2012, p.263). Though group members and norms can determine who the leaders are, leaders, with their popularity, trust, and legitimacy, can also influence group members by (re)constructing the group identity as “entrepreneurs of identity” (Reicher & Hopkins, 2001). For example, studies show that Donald Trump’s acknowledgment of the seriousness of the COVID-19 pandemic increased information seeking within Republican states (Xu & Margolin, 2024); messages from Republican leaders made Republicans more likely to comply with stay-at-home orders (Goldstein & Wiedemann, 2022); and overall people expressed greater policy support for in-group leaders than outgroup leaders (Cole et al., 2023).

Lastly, how conservatives and liberals prioritize different moral values (Haidt & Joseph, 2008) can also play a role in shaping risk evaluations and expressions. Moral foundations theory (MFT) suggests five dimensions of morality: care, fairness, loyalty, authority, and purity. Liberals tend to prioritize fairness, which includes self-interest and equality, while conservatives place more emphasis on authority, which involves following established societal hierarchies and traditions, and loyalty, such as standing with one's group, family, and nation (Ekins & Haidt, 2016; Graham et al., 2011). One recent study situated in a similar context to ours finds that Republican-leaning participants were more likely to see not getting vaccinated as morally permissible because they endorse and value loyalty to their in-group position, in contrast to their Democrat-leaning counterparts (Bruchmann & LaPierre, 2022). Likewise, another study demonstrates that during the pandemic political conservatism was associated with increased endorsement of MFT’s dimensions, such as authority, which was associated with fewer positive attitudes toward and reduced compliance with public health measures (Tarry et al., 2022).

While SIT, opinion leadership dynamics, and MFT all support the idea that partisans may talk about risks from the pandemic differently, it is less clear how these differences manifest in terms of how much emphasis Democrats and Republicans placed on pandemic-related risks and what kind of risks they were more concerned about. On the one hand, risk talk might figure more prominently in Democrats' COVID-19 discourse. A series of empirical studies have established that Democrats perceived COVID-19 as more severe and engage in risk mitigation behaviors more than Republicans (Allcott et al., 2020; Barrios & Hochberg, 2020; Bruine de Bruin et al., 2020; Clinton et al., 2021; Goldstein & Wiedemann, 2022; Gollwitzer et al., 2020; Gratz et al., 2021; Kerr et al., 2021; Milosh et al., 2021; Ye, 2023).

On the other hand, risk talk might account for a greater proportion of Republicans' COVID-19 discourse. Republicans/conservatives and Democrats/liberals have distinct moral palettes (Graham et al., 2009), potentially resulting in different risk evaluations: conservatives are more risk averse and prize traditional values, in-group loyalty, and obedience to authority, whereas liberals are more open to exploration and value equality and social justice (Choma et al., 2014; Jost, 2017). This suggests that Republicans might react more strongly to potential threats caused by COVID-19, particularly when mitigation practices such as mask-wearing and staying at home impinge on their existing way of life. This tendency might be strengthened by conservatives' limited media consumption (Faris et al., 2017) and the right-wing media ecosystem that circulates ideologically consistent information (Benkler et al., 2018). Furthermore, while Republicans might underestimate group risk, they are more responsive to individual risk (Kyung et al., 2022), suggesting they might be more sensitive to how COVID-19 brought about all kinds of risk to their personal lives.

Given that existing literature suggests opposing expectations about the proportion of risk talk in all pandemic-related expressions from Republicans and Democrats, we ask: Did risk talk (i.e., expression of dread and unknown risks) account for a higher proportion of expression related to the COVID-19 pandemic among Republicans or Democrats on social media? (**RQ1**)

Pandemic-related risk talk is further complicated by the drastically different risk targets that Republicans and Democrats have been shown to focus on. In terms of elite discourse on Twitter, Democratic members of Congress emphasized the risks of the pandemic to the public, whereas their Republican counterparts focused on risks related to China and businesses (Green et al., 2020). Another study shows that the dominant themes of Democratic legislators' tweets involved racial disparities in health care and insurance, COVID-19 testing, and public health guidelines, whereas dominant themes in Republican legislators' tweets included vaccine development, hospital resources, and equipment (Guntuku et al., 2021). Average Twitter users also exhibited ideological differences in discourse surrounding the COVID-19 vaccine: conservative Twitter users were more likely to focus on the side effects of vaccines, express distrust in medical professionals, and disseminate conspiracy theories than their liberal counterparts (Jiang, Su, et al., 2021). Given various risk targets, we pose another research question: How were Democrats' risk talk and Republicans' risk talk anchored in different risk targets? (**RQ2**)

As discussed above, leaders of social groups can play an outsized role in shaping attitudes, behaviors, and, in this research context, risk talk. Republicans who relied on Trump and conservative media for COVID-19 news were more likely to seek information sources that rated Trump's response highly (Deane et al., 2021). On social media, this dynamic may manifest as messages from a small fraction of people being widely shared, which is not only an act of imitation (Shah et al., 2015), but also a form of endorsement and amplification (Wojcieszak et al., 2022; Zhang et al., 2018). Since Republicans and Democrats exhibit different psychological make-ups (the former placing greater emphasis on group loyalty) and inhabit different media ecosystems (the former belonging to a media ecosystem that features more ideologically consistent information; Benkler et al., 2018), they might coalesce around different actors on social media via retweeting or reposting at varying degrees. Empirical evidence also reveals a tighter conservative social network structure, defined as fewer conservative elite accounts receiving most of the attention and conservatives amplifying their in-group leaders more frequently, which explains why conservative voices are more resounding than liberal voices in COVID-19 discourse on

Twitter (Haupt et al., 2021; Jiang, Ren & Ferrara, 2021; Wojcieszak et al., 2022). Thus, we investigate key drivers of risk talk on social media: Who drove the risk talk among Republicans and Democrats? (**RQ3**)

Methods

Data

Twitter's COVID-19 streaming endpoint provides researchers with free access to all pandemic-related conversations on Twitter in real time.[1] Twitter annotates related content with predefined parameters, and researchers can ingest the stream of tweets on a daily basis. Our research timeframe is from May 13, 2020, to March 19, 2021, covering multiple peaks of the pandemic and the early rollout of the COVID-19 vaccines.[2] In light of the vast volume of COVID-19 related expression on Twitter, we randomly sampled 1% of all data, which amounted to 10,267,941 tweets. Our data analysis included two parallel processes: clustering of partisans and classification of risk talk (Figure 1).

Clustering of Partisans: A User-Level Analysis

Given our interest in the partisan identities of people who engage in risk talk surrounding COVID-19 on Twitter, we first developed a novel and robust strategy to identify the likely partisan leanings of Twitter users based on retweet networks. Retweet networks have been widely used to detect clusters of similar users with high precision (Freelon et al., 2018; Guerrero-Solé, 2017) due to the high likelihood of Twitter users retweeting like-minded others (Conover et al., 2011). Previous research also shows how tightly partisans can coalesce around their leaders via retweeting (Zhang et al., 2018), providing further justification for this approach.

[1] https://developer.twitter.com/en/docs/twitter-api/tweets/covid-19-stream/overview

[2] Due to a server capacity issue, we were not able to collect data from the endpoint between December 23, 2020 and January 27, 2021. This gap should not affect our results significantly since it is relatively short within the entire time frame.

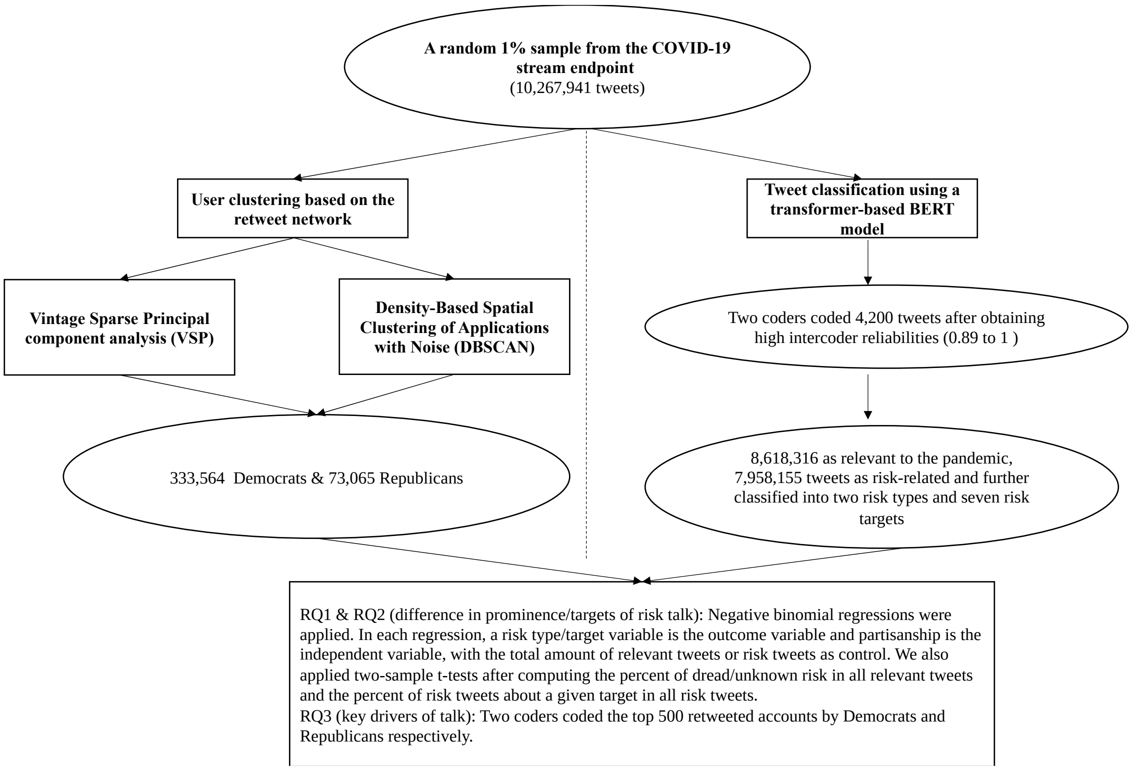

Figure 1. Data collection and analysis flow chart.

We applied, compared, and synthesized user clustering results from two different unsupervised machine learning methods to maximize the accuracy of results: an approach that uses Vintage Sparse Principal component analysis (VSP) (Rohe & Zeng, 2020), which was initially proposed for use on social media by Zhang et al. (2022), and an approach that uses the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm in conjunction with the Uniform Manifold Approximation and Projection (UMAP) algorithm for dimensionality reduction. This "DBSCAN plus UMAP" approach was originally proposed for use with Twitter data by Darwish et al. (2020). In both methods, we first built a bipartite matrix containing the users in our dataset (the rows of the matrix) and the accounts they retweeted (the columns), and then clustered the rows of this matrix based on similarities in their column-wise representations.

VSP is a spectral clustering technique that estimates latent factors in multivariate data in two steps: a low-rank singular value decomposition (SVD) and a varimax rotation on the singular vectors (Rohe & Zeng, 2020). To apply VSP to identify user partisanship, we first used a filter to include users (i.e., rows of the bipartite matrix) that retweeted at least two other accounts. Doing so induces a slight bias in our results towards more active users but is necessary to reduce noise in clustering results. In addition, we filtered out retweeted users (i.e., the columns of the bipartite matrix) that had been retweeted by fewer than 10 accounts, because they provide little information to the clustering while significantly increasing computational complexity. After filtering, we constructed a final bipartite network of who retweeted whom: 944,796 unique accounts retweeting 57,230 unique accounts. We set the number of clusters (k) at 50 for granular results (Appendix I). After applying VSP, each retweeting account had a probability of belonging to each of the 50 clusters, as did each account that was retweeted. Each account was assigned to the cluster it had the highest probability of belonging to. We interpreted each cluster of retweeting accounts based on 1) the most frequently retweeted accounts by each cluster and 2) the top words in the profile descriptions of the retweeted accounts by each cluster. See Appendix I for more information about the VSP method and results.

We then applied Darwish and colleagues' (2020) method. We first selected users (rows) who retweeted two or more other accounts and users being retweeted (columns) by 180 or more accounts (due to the high memory capacity imposed by this method). This resulted in a network of 944,796 unique accounts retweeting 5,202 unique accounts. Next, we conducted dimension reduction, transforming the network into a 2-dimension matrix using the UMAP algorithm. We then applied the DBSCAN clustering algorithm, which produced 22 clusters according to the density of the given 2D points. Finally, we interpreted the 22 clusters based on the same sets of information as we extracted in the VSP analysis. See Appendix II for more information about the results.

Classification of Risk Talk: A Content-Level Analysis

We applied a supervised machine learning method to classify all tweets into 11 variables based on our research hypotheses and questions. Two variables serve as filters: relevance (whether the content of the tweet is related to the COVID-19 pandemic) and non-U.S. (whether the tweet is exclusively about non-U.S. countries or global pandemic updates). Two variables focus on discrete risk types: dread risk (i.e., catastrophic potential, controllability, severity, voluntariness) and unknown risk (i.e., immediacy, whether knowledge is unknown to the public and/or known to scientists, novelty). In addition, after a first round of qualitatively coding a random sample of tweets, we developed seven variables encompassing different risk targets: 1) COVID (including cases, spread, diagnosis, & symptoms), 2) vaccines, 3) the economy, 4) public health (guidelines and policies such as mask mandates and social distancing), 5) politicians (including political figures and public institutions), 6) science (including individual scientists and the scientific community), and 7) information (including fake news and media manipulation).

To construct a ground truth dataset, two authors independently coded 200 randomly sampled tweets for all the variables at each round for six rounds, for a total of 1,200 tweets coded. A subset of authors, all of whom have experience with qualitative analysis, convened after each round of coding to resolve discrepancies, refine coding definitions, and create mutual understanding of definitions to promote consensus on subsequent iterations of coding. On the sixth round of coding, the two authors reached high intercoder reliabilities for each variable (Krippendorff’s alpha ranging from 0.89 to 1, mean alpha = 0.95), with the number of tweets that had agreed-upon codes ranging from 193 to 200 (96.5% to 100.0%, mean = 98.5%). Afterward, they coded 1,500 tweets each. See Appendix III for detailed information about the coding scheme and intercoder training process. With the 4,200 coded tweets, we conducted standard preprocessing, including removing “RT”, removing “@user,” removing URLs, and converting all text to lowercase, and then applied a transformer-based BERT model (Devlin et al., 2018), which is a state-of-the-art natural language processing model for text classification. To optimize model performance, we chose a BERT model that has been shown to work best with COVID-19-

related tweets (technically speaking, it was “pre-trained” with COVID-19-related tweets; Müller et al., 2020). For each of the risk targets, we used the simpletransformers package in Python to fine-tune each classifier individually.

To analyze the difference in the amounts and targets of risk talk by Democrats and Republicans (RQ1 and RQ2), we applied negative binominal regressions, a method appropriate for over-dispersed count outcome variables. We also applied two-sample t-tests to corroborate the findings. To analyze the key drivers of risk talk (RQ3), we used manual coding and focused on the top 500 retweeted accounts by Republicans and Democrats, respectively. Two authors independently coded the top retweeted accounts into seven categories: “politician, political affiliate, or government,” “political media,” “the academic and scientific community,” “activist,” “entertainment,” “Twitter influencer,” and “other” (See the details of the codebook in Appendix IV). Differences between the authors were resolved after extensive discussion.

Results

User Clustering and Tweet Classification

The VSP method produced user clusters with high resolution, including those based in India, Sri Lanka, the Philippines, South Africa, Nigeria, Australia, U.K., Canada, and the U.S. Given the scope of this paper, we focused on two U.S. Democrat clusters and two U.S. Republican clusters. The U.S. Democrat clusters retweeted Democratic politicians like Joe Biden, Bernie Sanders, and Alexandria Ocasio-Cortez, and progressive media personalities; words like “resist” and “blm” appeared most frequently in the profiles of most retweeted accounts. In contrast, the U.S. Republican clusters retweeted Donald Trump and other conservative personalities, and the profile descriptions of the accounts retweeted by those clusters featured words like “trump,” “maga,” and “conservative”.

Altogether, this method identified 155,961 Republicans and 458,051 Democrats.[3] The UMAP with DBSCAN clustering method produced 22 clusters, including clusters based in the U.S., U.K., Canada, and India. This method identified 78,312 Republicans and 575,738 Democrats.

To evaluate results generated from the two methods, we introduced the Jaccard index to measure the similarity of two sample sets, with a higher value in the Jaccard index indicating greater similarity between the two samples. For Democrats, VSP and UMAP-DBSCAN results share 333,564 overlapping users, with the Jaccard index being 0.48. For Republicans, there are 73,065 overlapping users, with the Jaccard index being 0.45. To prioritize precision, we kept only users classified into the same category by both methods.[4]

The BERT-based supervised machine learning approach classified 8,618,316 tweets among all tweets as relevant to the pandemic. Within all relevant tweets, 6,667,549 (77.4%) expressed dread risk, while a much lower number of tweets (1,290,606; 15.0%) were about unknown risk. Among all risk-related tweets (i.e., about either dread or unknown risks, 7,958,155), the most significant risk target was COVID-19 cases, spread, diagnosis, and symptoms (2,684,358, 33.7%), followed by public health (1,393,236, 17.5%), politicians (712,394, 9.0%), science (174,320, 2.2%), information (170,979, 2.1%), economy (111,207, 1.4%), and the COVID-19 vaccine (19,674, 0.2%). See Appendix V for the performance metrics and Appendix VI for detailed descriptive statistics.

Partisan Differences in Risk Talk

[3] Given that our unsupervised methods allow discovery instead of classification, we cannot make the fine distinction between a Republican cluster and a conservative cluster (or a Democrat cluster vs a liberal cluster). However, this should not be a concern given how partisan-ideological sorting has resulted in the convergence of partisan identity and ideological identity.

[4] Overall, the two methods did a relatively consistent job of identifying Republicans and Democrats (Appendix I & Appendix II). Selecting the overlap between the two sets of results provided a more stringent and accurate estimate of the talk by strong Republicans and Democrats.

RQ1 asks whether Republicans or Democrats emphasized risks more (including dread & unknown risks). We computed for a given Republican or Democrat user the percent of dread risk tweets and the percent of unknown risk tweets in all relevant tweets and plotted the distributions in Figure 2.

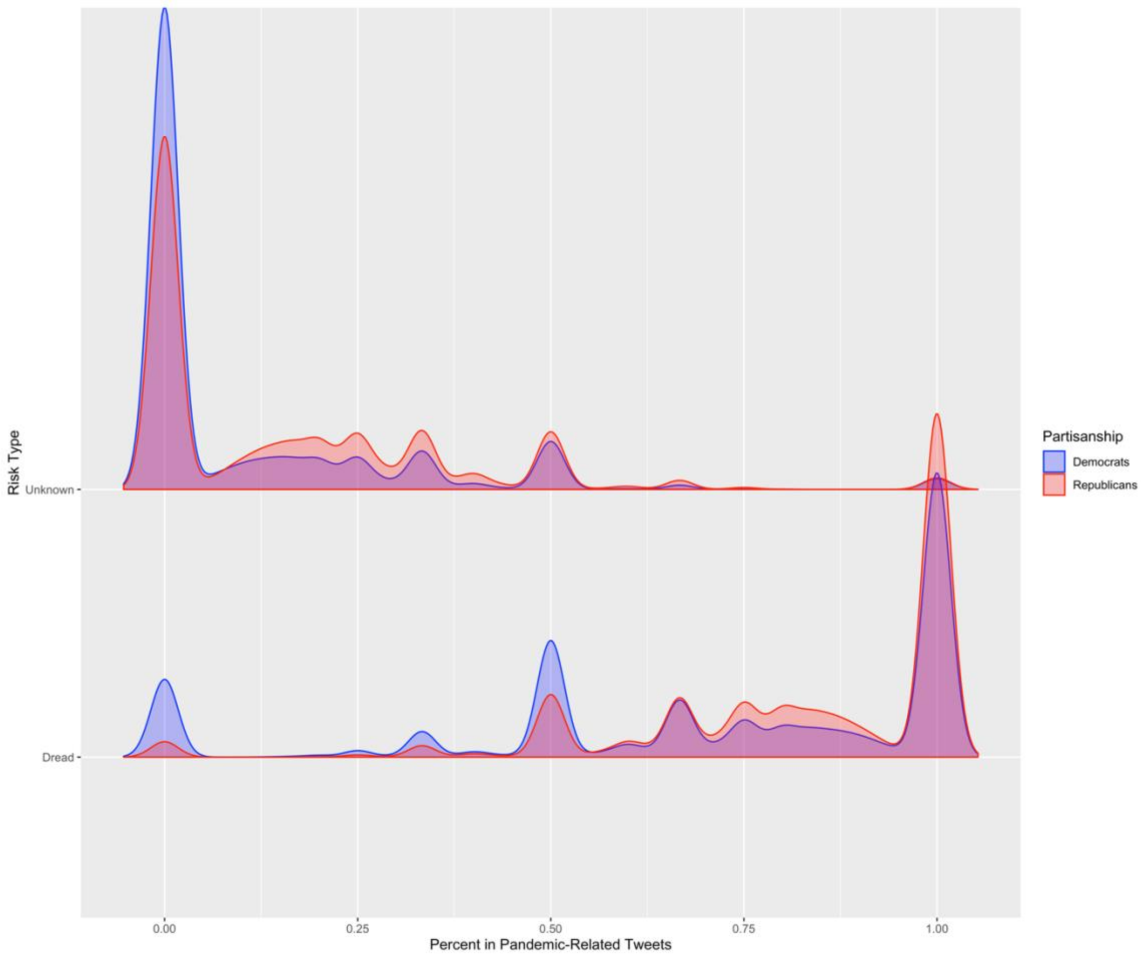

Figure 2. Distribution of percent of dread/unknown-risk tweets in pandemic-related tweets for each user by partisan identity.

According to negative binomial regressions, compared to Democrats, Republicans in our sample produced 1.1 times as many dread risk related tweets ($p < .001$), and 1.4

times as many unknown risk related tweets ($p < .001$). Welsch's two-sample t-tests yield identical findings: Republicans tended to post a higher percentage of dread risk tweets in relevant tweets in comparison to Democrats ($t(151087) = 120.5$, $p < .001$); on average, Republican users posted a higher percentage of unknown risk tweets in relevant tweets than Democrats ($t(101967) = 65.2$, $p < .001$).

RQ2 asks how Republicans' risk talk and Democrats' risk talk were anchored in distinct risk targets. For a given user identified as a Republican or Democrat, we computed the percent of tweets concerning a given risk target in all risk tweets and visualized the distributions in Figure 3. Compared to Democrats, Republicans on average generated 0.7 times as many tweets about COVID-19 cases, spread, diagnosis, and symptoms, 0.7 times as many tweets about political figures and public institutions, and 0.5 times as many tweets about the economy; however, they produced an average of 1.3 times as many tweets about public health guidelines and policies, 5.5 times as many tweets about vaccines, 1.8 times as many tweets about science and the scientific community, and 7.4 times as many tweets about information. Welsch's two-sample t-tests corroborate this pattern. In our sample, Democrats had the tendency to talk more about risks related to COVID-19 cases, spread, diagnosis, and symptoms ($t(126657) = 83.0$, $p < .001$), express more concerns over political figures and public institutions ($t(116471) = 10.2$, $p < .001$), and discuss more economy-related risks ($t(150469) = 33.6$, $p < .001$). However, Republicans, on average, talked more about all other risk targets, including vaccines ($t(85008) = 26.9$, $p < .001$), public health guidelines and policies ($t(122608) = 14.3$, $p < .001$), science and the scientific community ($t(89564) = 58.9$, $p < .001$), and information ($t(75819) = 122.9$, $p < .001$).

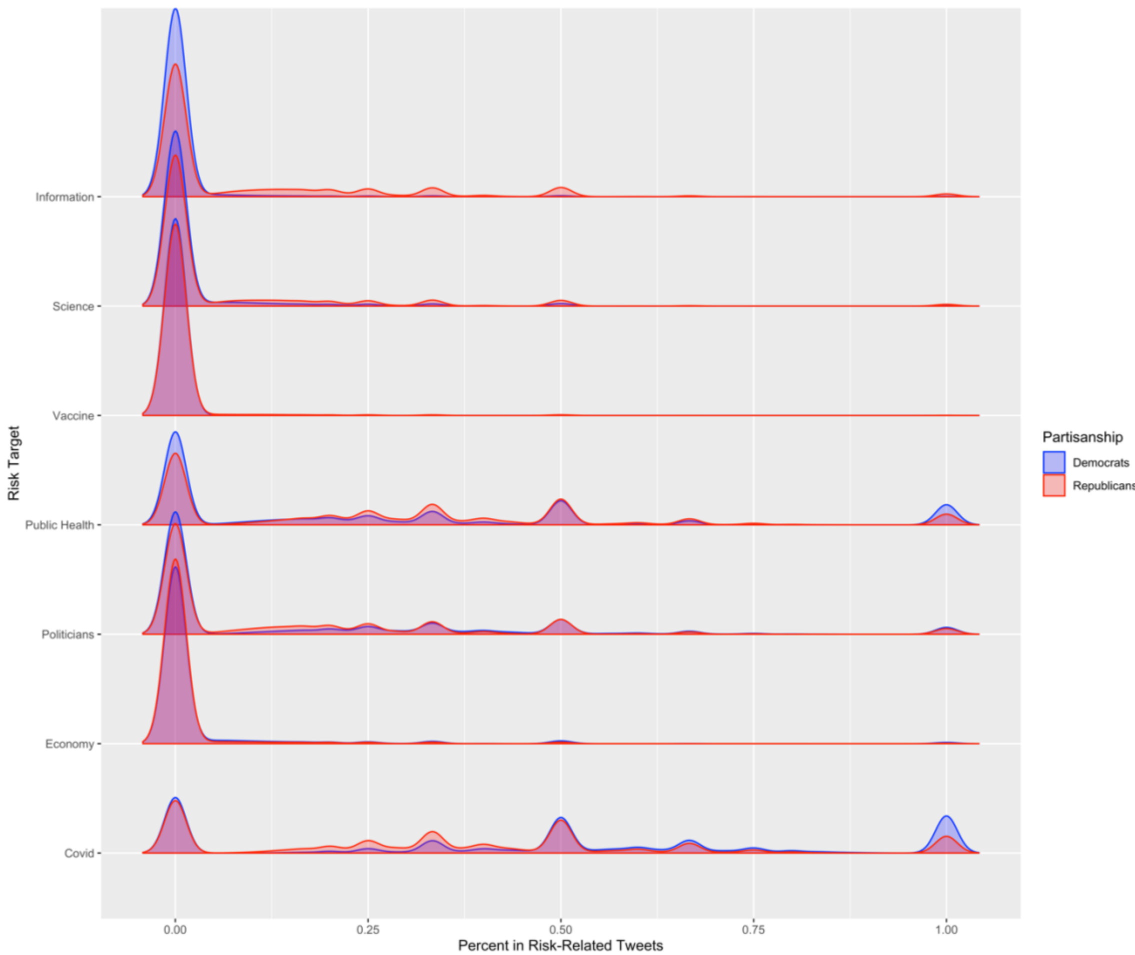

Figure 3. Distribution of percent of tweets about a given risk target in all risk tweets for each user by partisan identity.

Drivers of Risk Talk

Retweets accounted for the overwhelming majority of all tweets by Republicans (89.2%) and Democrats (89.6%). The distribution of the number of retweets received per retweeted account roughly followed a power-law distribution, with the top retweeted accounts receiving outsized attention (see the distributions in Appendix VII). However, a partisan asymmetry can also be observed. Among all the dread and unknown risk talk by Democrats in our sample, retweets of the 20 most retweeted accounts made up 13.6% of all risk-related retweets; for Republicans, this proportion was nearly doubled at 24.3%. Similarly, the 500 most retweeted accounts were responsible for 47.6% of all risk-related tweets for Democrats; meanwhile, the top 500 retweeted accounts provided 69.4% of all risk-related tweets among Republicans. This pattern indicates that the top retweeted accounts commanded greater attention and popularity among Republicans in our sample than among Democrats.

In November 2021, we used Twitter's Search API to query the user information for the top 500 retweeted accounts by Republicans and Democrats, respectively. We found that 166 accounts (33%) among the top 500 retweeted accounts by Republicans were suspended by Twitter, more than six times the number of suspended accounts retweeted by Democrats (25 accounts, 5%). This suggests that a far greater share of those key drivers of risk talk among Republicans violated Twitter's community guidelines. This discrepancy is possibly due to the suspension of far-right accounts on Twitter following the January 6th Capitol Attack.[5] Table 1 summarizes the percentages of the different types of accounts in the unsuspended accounts. Political media accounted for retweeted content the most on both the left and the right, with 33.3% of the top retweeted accounts by Democrats and 40.4% of top retweeted accounts by Republicans belonging to established or grassroots political media (e.g., Kyle Griffin, a producer at MSNBC; Laura Ingraham with Fox News). Politicians (e.g., Joe Biden among Democrats and Donald Trump among Republicans) appeared in Republican retweets (19.5% of the top retweeted accounts by Republicans) more than the Democrat retweets (12.8%). In fact, media, politicians, and activists together accounted for 76.4% of the unsuspended top accounts retweeted by Republicans, and 64.2% by Democrats. Interestingly, 9.9% of the top retweeted accounts by Democrats were academics and scientists, though only 4.8% of those top accounts retweeted by Republicans belonged to this group.

[5] See "Twitter has suspended 70,000 QAnon accounts since U.S. Capitol riot" from CNBC.

Table 1. Distribution of different types of (unsuspended) top retweeted accounts by Republicans and Democrats.

	Republicans		Democrats	
	Count	Percent	Count	Percent
Political Media	135	40.4%	158	33.3%
Politician, Political Affiliate, or Government	65	19.5%	61	12.8%
Activist	55	16.5%	86	18.1%
The Academic and Scientific Community	16	4.8%	47	9.9%
Entertainment	27	8.1%	54	11.4%
Twitter Influencer	20	6.0%	23	4.8%
Other	16	4.8%	46	9.7%
Total	334	100%	475	100%

Discussion

This article examines the interplay between partisanship and risk talk on Twitter during the early stages of the COVID-19 pandemic. This study complements existing research by describing which partisans gave more emphasis to risk talk, what risk targets partisans focused on, and which individual accounts were central to their risk talk. We present rich and nuanced patterns concerning the prominence, targets, and drivers of partisans' risk talk. The results suggest that not only did Republicans talk about pandemic-related risks with greater intensity on Twitter, but they also focused on a distinct set of risk targets and followed the cues of media and political leaders more than their Democratic counterparts. These results provide valuable insights into U.S. partisan identities and their manifestations in risk-related expression on social media. Our methodological approach attends to user characteristics, content features, and key actors in users' expression patterns, providing a powerful template for harnessing unsupervised and supervised machine learning as well as manual coding to discover online expression dynamics.

Though dread risk is much more prevalent than unknown risk for both Republicans and Democrats in our sample, Republicans talked about dread risk and unknown risk in their tweets at higher rates than Democrats. This finding corroborates prior evidence showing the greater level of activity and engagement of conservative Twitter users in general COVID-19 discourse (Jiang, Ren & Ferrara, 2021). It indicates that, although Republicans in general might report lower risk perception, as survey-based research shows, Republicans on Twitter were rather vocal about risks related to the pandemic. This discrepancy between survey and social media results might be attributable to the different demographics included in survey and social media data (Barberá & Rivero, 2015) and/or the different social norms in survey settings and on social media (Joseph et al., 2021). In other words, Republicans who have made up their minds about risks during the pandemic may be more compelled to speak their minds than Democrats.

The uneven sensitivity to the risk targets among partisans on social media that we documented adds more nuance to this Republican outspokenness about risks. To summarize, compared to Democrats, Republicans tended to focus more on the risks related to public health guidelines and policies, media and information, science and the scientific community, and vaccines (a similar finding about vaccines can be found in Jiang, Su, et al., 2021). In contrast, Democrats, on average, expressed more concerns over COVID-19 cases, spread, diagnosis, and symptoms, political figures and public institutions, and the economy. These patterns suggest that the overall heightened outspokenness on the Republican side might be rooted in conservatives' distinct psychological orientation. The catastrophic potential, uncontrollability, severity, and involuntariness of the pandemic may be particularly dreadful for Republicans/conservatives in light of their sensitivity to threat against order (Jost, 2017), thus motivating action in the form of tweeting. Specifically, their risk-related expression may reflect their grievances against public health guidelines and policies and the vaccines, likely viewed as an infringement on their personal freedom and liberty, as well as American cultural traditionalism. Contributing to risk perception theories, these results suggest that deeply rooted political and psychological orientations may influence the extent to which different risk attributes impact risk perception.

Furthermore, the patterns observed in our data align with Republicans' rising distrust of the government, science, and news media (Krause et al., 2019; Lee, 2010). Rush Limbaugh, the deceased, influential, right-wing talk show host, offered a prototype of this tendency among Republicans when he called government, academia, science, and media the "Four Corners of Deceit." This claim from a core conservative political figure appears to have found its expression in how Republicans think and talk about risks during the COVID-19 pandemic. This ethos is also consistent with the pattern observed within our sample regarding key actors frequently retweeted and amplified by partisans: Republicans retweeted academics about half as frequently as Democrats.

In addition, our analysis of the top retweeted accounts by Republicans and Democrats implies that risk talk among partisans mostly entailed retweeting content from political media, politicians, and political activists, demonstrating how politics is deeply intertwined with partisans' risk perception and expression. For example, Democrats' greater emphasis on risks posed by politicians might be attributed as much to Donald Trump's unsatisfactory pandemic response as to him being the president throughout much of the time period covered by our dataset. These results underscore the politicization of the pandemic and the seeping of a deep partisan divide into an ongoing public health crisis, which should be factored into risk communication about science and public health guidance.

The fact that a relatively small number of political media accounts, politicians, and political activists were retweeted more in risk talk by Republicans (as compared to Democrats) underscores how the ideologically oriented conservative information ecosystem might be especially powerful in steering conservatives' beliefs and expressions. This finding is consistent with research showing that conservative networks on Twitter are more centralized and produce more consistent messages (Haupt et al., 2021; Jiang, Ren & Ferrara, 2021). As Donald Trump, Republican governors, and Fox News alike have unanimously and persistently downplayed COVID-19 risks (Simonov et al., 2020) and undermined the scientific community and mainstream media, it is not surprising to see that

Republicans in our sample expressed concerns over risks associated with coronavirus at a lower rate than Democrats, yet voiced more concerns about vaccines, science, and media.

These findings also point to the persuasive power residing in elites during times of division. Existing research considers how both message and source characteristics can influence behaviors during public health crises (e.g., Bokemper et al., 2021; Moniz, 2020; Sanchez & Dunning, 2021). Our research speaks in concert with such studies by finding real and meaningful partisan differences in both message content (in terms of the volume and targets of risk) and source (in terms of which actors are opinion leaders) characteristics in social media discourses surrounding the COVID-19 pandemic-related risks. Since these groups hold on to their partisan identities, opinion leaders within their groups are more likely to be the sources explaining their attitudes and behaviors.

Methodologically, this article applies unsupervised and supervised machine learning methods and manual coding to conduct a thorough investigation of "who said what" and "who served as opinion leaders" on social media. The stark and consistent partisan differences noted by our findings demonstrate the value of this multipronged analytic approach that examines users, expression patterns, and focal users of social media discourses. We call on future research to use complementary methods to attend to the multiple dimensions of social media discourses. Since our research is descriptive and situated in one context within one country, future studies should examine how partisanship seeps into risk talk in different countries and test what messaging strategies can temper the biasing effects of partisan identities on risk perceptions and expression.

Overall, our study provides valuable insights into the pandemic-related risk talk of Democrats and Republicans in the U.S., pointing to the huge challenges faced by American society in times of crisis. Deep political division and overpowering partisan identity can hinder collective efforts in response to societal risks. However, our results also suggest that the solution may rest within leveraging social identity: leaders of various social groups might be the last resort to help partisans rise above group differences and recognize our many shared social identities and our ultimate bond as humanity (Levendusky, 2023).

Acknowledgements

We thank Dhavan Shah for sharing data from Twitter's COVID-19 stream endpoint.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Abramowitz, A. I., & Saunders, K. L. (2008). Is polarization a myth? The Journal of Politics, 70(2), 542-555. https://doi.org/10.1017/S0022381608080493

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. Journal of public economics, 191, 104254.

Barberá, P., & Rivero, G. (2015). Understanding the political representativeness of Twitter users. Social Science Computer Review, 33(6), 712-729.

Barrios, J. M., & Hochberg, Y. (2020). Risk perception through the lens of politics in the time of the covid-19 pandemic (No. w27008). National Bureau of Economic Research.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. Psychological Bulletin, 117(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Benkler, Y., Faris, F., & Roberts, H. (2018). Network propaganda: Manipulation, disinformation, and radicalization in American politics. Oxford University Press. https://doi.org/10.1093/oso/9780190923624.001.0001

Bokemper, S. E., Huber, G. A., Gerber, A. S., James, E. K., & Omer, S. B. (2021). Timing of COVID-19 vaccine approval and endorsement by public figures. Vaccine, 39(5), 825-829. https://doi.org/10.1016/j.vaccine.2020.12.048

Brewer, M. B. (1991). The social self: On being the same and different at the same time. Personality and Social Psychology Bulletin, 17(5), 475–482. https://doi.org/10.1177/0146167291175001

Bridgman, A., Merkley, E., Loewen, P. J., Owen, T., Ruths, D., Teichmann, L., & Zhilin, O. (2020). The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media. Harvard Kennedy School Misinformation Review, 1(3).

Bruchmann, K., & LaPierre, L. (2022). Moral foundations predict perceptions of moral permissibility of COVID-19 public health guideline violations in United States university students. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.795278

Bruine de Bruine, W., Saw, H-W., & Goldman, D. P. (2020). Political polarization US residents' COVID-19 risk perceptions, policy preferences, and protective behaviors. Journal of Risk and Uncertainty, 61, 177-194. http://doi.org/ https://.1007/s11166-020-09336-3

Chen, H. F., & Karim, S. A. (2022). Relationship between political partisanship and COVID-19 deaths: future implications for public health. *Journal of public health*, *44*(3), 716-723.

Choma, B. L., Hanoch, Y., Hodson, G., & Gummerum, M. (2014). Risk propensity among liberals and conservatives: The effect of risk perception, expected benefits, and risk domain. *Social Psychological and Personality Science, 5*(6), 713-721. https://doi.org/10.1177/1948550613519682

Clahsen, S. C. S., van Kamp, I., Hakkert, B. C., Vermeire, T. G., Piersma, A. D., & Lebret, E. (2018). Why do countries regulate environmental health risks differently? A theoretical perspective. Risk Analysis, 39(2), 439-461.

https://doi.org/10.1111/risa.13165

Claassen, R. L., & Highton, B. (2008). Policy polarization among party elites and the significance of political awareness in the mass public. Political Research Quarterly, 62(3), 538-551. https://doi.org/10.1177%2F1065912908322415

Clinton, J., Cohen, J., Lapinski, J., & Trussler, M. (2021). Partisan pandemic: How partisanship and public health concerns affect individuals' social mobility during COVID-19. *Science advances*, *7*(2), eabd7204.

Cole, J. C., Flores, A., Jiga-Boy, G. M., Klein, O., Sherman, D. K., & Van Boven, L. (2023). Party over pandemic: Polarized trust in political leaders and experts explains public support for COVID-19 policies. *Group Processes & Intergroup Relations*, *26*(7), 1611-1640.

Conover, M. D., Ratkiewicz, J. R., Francisco, M., Goncalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on Twitter. Proceedings of the Fifth International Conference on Weblogs and Social Media, 89-96. http://www.cse.cuhk.edu.hk/~cslui/CMSC5734/Conover_PoliticalPolarizationTwitter.pdf

Darwish, K., Stefanov, P., Aupetit, M., & Nakov, P. (2020). Unsupervised user stance detection on twitter. Proceedings of the International AAAI Conference on Web and Social Media, 14, 141-152.

de Vries, M., Ciaassen, L., Mennen, M., Timen, A., te Wierik, M. J. M., & Timmermans, D. R. M. (2019). Public perceptions of contentious risk: The case of rubber granulate in the Netherlands. International Journal of Environmental Research and Public Health, 16(12), 2250. https://doi.org/10.3390/ijerph16122250

Deane, C., Parker, K., & Gramlich, J. (2021, June 28). U.S. public opinion on the coronavirus pandemic. Pew Trust. https://www.pewtrusts.org/en/trust/archive/spring-2021/us-public-opinion-on-the-coronavirus-pandemic

Density-Based Spatial Clustering of Applications with Noise (DBSCAN). (2021, n.d.). DBSCAN - Density-based spatial clustering of applications with noise. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. PsyARXiv. https://arxiv.org/abs/1810.04805

Ekins, M. & Haidt, J. (2016). "Donald Trump supporters think about morality differently than other voters. Here's how." Cato Institute, February 5. https://www.cato.org/publications/commentary/donald-trump-sup-porters-think-about-morality-differently-other-voters-heres

Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2017). Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election. Berkman Klein Center Research Publication, 6.

Finkel, E. J., Bail, C., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, I., Mcgrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C., & Druckman, J. N. (2020). Political sectarianism in America. Science, 370(6516), 533-536. https://doi.org/10.1126/science.abe1715

Freelon, D., McIlwain, C., & Clark, M. (2018). Quantifying the power and consequences of social media protest. new media & society, 20(3), 990-1011.

Gadarian, S. K., Goodman, S. W., & Pepinsky, T. B. (2021). Partisanship, health behavior, and policy attitudes in the early stages of the COVID-19 pandemic. Plos one, 16(4), e0249596.

Goldring, M. R., & Heiphetz, L. (2020). Sensitivity to ingroup and outgroup norms in the association between commonality and morality. Journal of Experimental Social Psychology, 91. https://doi.org/10.1016/j.jesp.2020.104025

Goldstein, D. A., & Wiedemann, J. (2022). Who do you trust? The consequences of partisanship and trust for public responsiveness to COVID-19 orders. *Perspectives on Politics*, *20*(2), 412-438.

Gollwitzer, A., Martel, C., Brady, W. J., Pärnamets, P., Freedman, I. G., Knowles, E. D., & Van Bavel, J. J. (2020). Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic. *Nature human behaviour*, *4*(11), 1186-1197.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. Journal of personality and social psychology, 96(5), 1029.

Graham, J., Nosek, B. A., Iyer, H. R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. Journal of Personality and Social Psychology, 101(2), 366-385. https://doi.org/10.1037/a0021847.

Gratz, K. L., Richmond, J. R., Woods, S. E., Dixon-Gordon, K. L., Scamaldo, K. M., Rose, J. P., & Tull, M. T. (2021). Adherence to social distancing guidelines throughout the COVID-19 pandemic: The roles of pseudoscientific beliefs, trust, political party affiliation, and risk perceptions. *Annals of Behavioral Medicine*, *55*(5), 399-412.

Green, J., Edgerton, J., Naftel, D., Shoub, K., & Cranmer, S. J. (2020). Elusive consensus: Polarization in elite communication on the COVID-19 pandemic. Science Advances, 6(28). https://doi.org/10.1126/sciadv.abc2717

Guerrero-Solé, F. (2017). Community detection in political discussions on Twitter: An application of the retweet overlap network method to the Catalan process toward independence. Social science computer review, 35(2), 244-261.

Guntuku, S. C., Purtle, J., Meisel, Z. F., Merchant, R. M., & Agarwal, A. (2021). Partisan differences in Twitter language among US legislators during the COVID-19 pandemic: Cross-sectional study. Journal of Medical Internet Research, 23(6). https://doi.org/10.2196/27300

Haidt, J., & C. Joseph. (2008). The Moral Mind: How Five Sets of Innate Intuitions

Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules." In The Innate Mind Vol. 3. Foundations and the Future, edited by P. Carruthers, S. Laurence, and S. Stich, 367–391. Oxford, GB: Oxford University Press.

Haupt, M. R., Jinich-Diamant, A., Li, J., Nali, M., & Mackey, T. K. (2021). Characterizing twitter user topics and communication network dynamics of the "liberate" movement during COVID-19 using unsupervised machine learning and social network analysis. Online Social Networks and Media, 21, 100114.

Hogg, M. A., & Abrams, D. (1988). Social identifications: A social psychology of intergroup relations and group processes. Taylor & Frances/Routledge.

Hogg, M. A., van Knippenberg, D., & Rast III, D. E. (2012). The social identity theory of leadership: Theoretical origins, research findings, and conceptual developments. European Review of Social Psychology, 23(1), 258-304. https://doi.org/10.1080/10463283.2012.741134

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. Annual Review of Political Science, 22, 129-146. https://doi.org/10.1146/annurev-polisci-051117-073034

Joseph, K., Shugars, S., Gallagher, R., Green, J., Mathé, A. Q., An, Z., & Lazer, D. (2021). (Mis) alignment Between Stance Expressed in Social Media Data and Public Opinion Surveys. arXiv preprint arXiv:2109.01762.

Jost, J. T. (2017). Ideological asymmetries and the essence of political psychology. Political psychology, 38(2), 167-208.

Jiang, J., Ren, X., & Ferrara, E. (2021). Social media polarization and echo chambers in the context of COVID-19: Case study. JMIRx med, 2(3), e29570.

Jiang, X., Su, M. H., Hwang, J., Lian, R., Brauer, M., Kim, S., & Shah, D. (2021). Polarization over vaccination: Ideological differences in Twitter expression about

COVID-19 vaccine favorability and specific hesitancy concerns. *Social Media+ Society*, *7*(3), 205630512110484.

Kerr, J., Panagopoulos, C., & van der Linden, S. (2021). Political polarization on COVID-19 pandemic response in the United States. Personality and Individual Differences, 179. https://doi.org/10.1016/j.paid.2021.110892

Krause, N. M., Brossard, D., Scheufele, D. A., Xenos, M. A., & Franke, K. (2019). Trends—Americans' trust in science and scientists. Public Opinion Quarterly, 83(4), 817-836.

Kyung, E., Thomas, M., & Krishna, A. (2022). How political identity influences COVID-19 risk perception: A model of identity-based risk perception. *Journal of the Association for Consumer Research*, *7*(3), 316-324.

Lee, T. T. (2010). Why they don't trust the media: An examination of factors predicting trust. American behavioral scientist, 54(1), 8-21.

Levendusky, M. (2023). Our common bonds: Using what Americans share to help bridge the partisan divide. University of Chicago Press.

Lorenz, A., Schmitt, C., McGregor, S., & Malmer, D. (2023). "CNN Can Kiss My As $": A Novel Description of Hyperpartisan US News Consumers. Journal of Quantitative Description: Digital Media, 3.

Mason, L. (2015). "I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. American journal of political science, 59(1), 128-145.

McCarty, N., Poole, K. T., & Rosenthal, H. (2016). Polarized America: The Dance of Ideology and Unequal Riches. MIT Press.

Milosh, M., Painter, M., Sonin, K., Van Dijcke, D., & Wright, A. L. (2021). Unmasking partisanship: Polarization undermines public response to collective risk. *Journal of Public Economics*, *204*, 104538.

Moniz, P. (2020). How Bad is it? Elite Influence and the Perceived Seriousness of the Coronavirus Pandemic. Journal of Experimental Political Science, 1-9. https://doi.org/10.1017/XPS.2020.45

Moon, W. K., Atkinson, L., Kahlor, L. A., Yun, C., & Son, H. (2022). US political partisanship and COVID-19: Risk information seeking and prevention behaviors. *Health Communication*, *37*(13), 1671-1681.

Motta, M., & Stecula, D. (2023). The effects of partisan media in the face of global pandemic: how news shaped COVID-19 vaccine hesitancy. *Political Communication*, *40*(5), 505-526.

Müller, M., Salathé, M., & Kummervold, P. E. (2020). COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on Twitter. PsyArXiv. https://arxiv.org/abs/2005.07503

Nguyen, H., & Nguyen, A. (2020). Covid-19 misinformation and the social (media) amplification of risk: A Vietnamese perspective. Media and Communication, 8(2), 444-447.

Park, S., Han, S., Kim, J., Molaie, M. M., Vu, H. D., Singh, K., ... & Cha, M. (2021). COVID-19 discourse on twitter in four asian countries: Case study of risk communication. Journal of medical Internet research, 23(3), e23272.

Peterson, E., & Iyengar, S. (2022). Partisan reasoning in a high stakes environment: Assessing partisan informational gaps on COVID-19. *Harvard Kennedy School Misinformation Review*.

Qiao, S., Li, Z., Liang, C., Li, X., & Rudisill, C. (2023). Three dimensions of COVID-19 risk perceptions and their socioeconomic correlates in the United States: A social media analysis. Risk Analysis, 43(6), 1174-1186.

Rao, A., Morstatter, F., Hu, M., Chen, E., Burghardt, K., Ferrara, E., & Lerman, K. (2021). Political partisanship and antiscience attitudes in online discussions about COVID-19: Twitter content analysis. Journal of medical Internet research, 23(6), e26692.

Reicher, S., & Hopkins, N. (2001). Psychology and the end of history: A critique and a proposal for the psychology of social categorization. Political Psychology, 22(2), 383–407. https://doi.org/10.1111/0162-895X.00246

Rohe, K., & Zeng, M. (2020, April 20). Vintage factor analysis with varimax performs statistical inference. https://arxiv.org/abs/2004.05387

Sanchez, C., & Dunning, D. (2021). The anti-scientists bias: The role of feelings about scientists in COVID-19 attitudes and behaviors. Journal of Applied Social Psychology, 51, 461-473. https://doi.org/10.1111/jasp.12748

Shah, D. V., Culver, K. B., Hanna, A., Macafee, T., & Yang, J. (2015). Computational approaches to online political expression: Rediscovering a 'science of the social'. In Handbook of digital politics. Edward Elgar Publishing.

Shin, J., Yang, A., Liu, W., Min Kim, H., Zhou, A., & Sun, J. (2022). Mask-wearing as a partisan issue: Social identity and communication of party norms on social media among political elites. *Social Media+ Society*, *8*(1), 20563051221086233.

Sjöberg, L., Moen, B. E., & Rundmo, T. (2004). Explaining risk perception. An evaluation of the psychometric paradigm in risk perception research, 10(2), 1-33.

Slovic, P. (1987). Perception of risk. Science, 236(4799), 280-285. https://doi.org/10.1126/science.3563507

Slovic, P. (Ed.). (2000). Risk, society, and policy series. The perception of risk. Earthscan Publications.

Tajfel, H. (1982). Social psychology of intergroup relations. Annual Review of Psychology, 33, 1–39. https://doi.org/10.1146/annurev.ps.33.020182.000245

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.), The social psychology of intergroup relations (pp.

33-37). Monterey, CA: Brooks/Cole.

Tarry, H., Vezina, V., Bailey, J., & Lopes,L. (2022). Political orientation, moral foundations ,and COVID-19 social distancing. PloS ONE, 17(6). https://doi.org/10.1371/journal.pone.0267136

Simonov, A., Sacher, S. K., Dubé, J. P. H., & Biswas, S. (2020). The persuasive effect of fox news: non-compliance with social distancing during the covid-19 pandemic (No. w27237). National Bureau of Economic Research.

Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP). (2018, n.d.). UMAP: Uniform manifold approximation and projection for dimension reduction. https://umap-learn.readthedocs.io/en/latest

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. Trends in cognitive sciences, 22(3), 213-224.

Wachinga, G., Renn, O., Begg, C., & Kuhlicke. (2013). The risk perception paradox—Implications for governance and communication of natural hazards. Risk Analysis, 33(6), 1049-1065. https://doi.org/10.1111/j.1539-6924.2012.01942.x

Wang, J., Zhou, Y., Zhang, W., Evans, R., & Zhu, C. (2020). Concerns expressed by Chinese social media users during the COVID-19 pandemic: content analysis of Sina Weibo microblogging data. Journal of medical Internet research, 22(11), e22152.

Warren, G. W., Lofstedt, R. (2021). COVID-19 vaccine rollout risk communication strategies in Europe: A rapid response. Journal of Risk Research, 24(3-4), 369-379. https://doi.org/10.1080/13669877.2020.1870533

Waytz, A., & Epley, N. (2012). Social connection enables dehumanization. Journal of Experimental Social Psychology, 48(1), 70–76. https://doi.org/10.1016/j.jesp.2011.07.012

Wojcieszak, M., Casas, A., Yu, X., Nagler, J., & Tucker, J. A. (2022). Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity. Science advances, 8(39), eabn9418.

Wong, J. CS., & Yang, J. Z. (2021). Comparative risk: Dread and unknown characteristics of the COVID-19 pandemic vs. COVID-19 vaccines. Risk Analysis. Advance online publication. https://doi.org/10.1111/risa.13852

Xu, Y., & Margolin, D. (2024). Collective Information Seeking During a Health Crisis: Predictors of Google Trends During COVID-19. Health Communication, 39(2), 388-402.

Ye, X. (2023). Exploring the relationship between political partisanship and COVID-19 vaccination rate. *Journal of Public Health*, *45*(1), 91-98.

Zhang, Y., Chen, F., & Rohe, K. (2022). Social Media Public Opinion as Flocks in a Murmuration: Conceptualizing and Measuring Opinion Expression on Social Media. Journal of Computer-Mediated Communication, 27(1), https://doi.org/10.1093/jcmc/zmab021

Zhang, Y., Chen, F., & Lukito, J. (2023). Network amplification of politicized information and misinformation about COVID-19 by conservative media and partisan influencers on Twitter. *Political Communication*, *40*(1), 24-47.

Zhang, Y., Wells, C., Wang, S., & Rohe, K. (2018). Attention and amplification in the hybrid media system: The composition and activity of Donald Trump's Twitter following during the 2016 presidential election. *New Media & Society*, *20*(9), 3161-3182.

Zhong, W., & Broniatowski, D. A. (2023). Economic risk framing increases intention to vaccinate among Republican COVID-19 vaccine refusers. *Social Science & Medicine*, *317*, 115594.

Zhou, A., Liu, W., & Yang, A. (2023). Politicization of science in COVID-19 vaccine communication: Comparing US politicians, medical experts, and government agencies. Political Communication, 1-23.

Supplementary Materials

Appendix I. VSP clustering results

A Scree plot of the magnitude of the eigenvalue (y-axis) for each eigenvector (x-axis) in the matrix decomposition generated by the VSP algorithm. Eigenvalues can be interpreted as the percentage of variance explained by the relevant eigenvector assuming the model from the algorithm. The Scree plot below shows the variance explained by different Ks. As can be seen, the explained variance stops decreasing significantly after K = 26. Therefore, we picked K =50, a relatively large K, so that unique clusters can be preserved, and similar clusters can be merged.

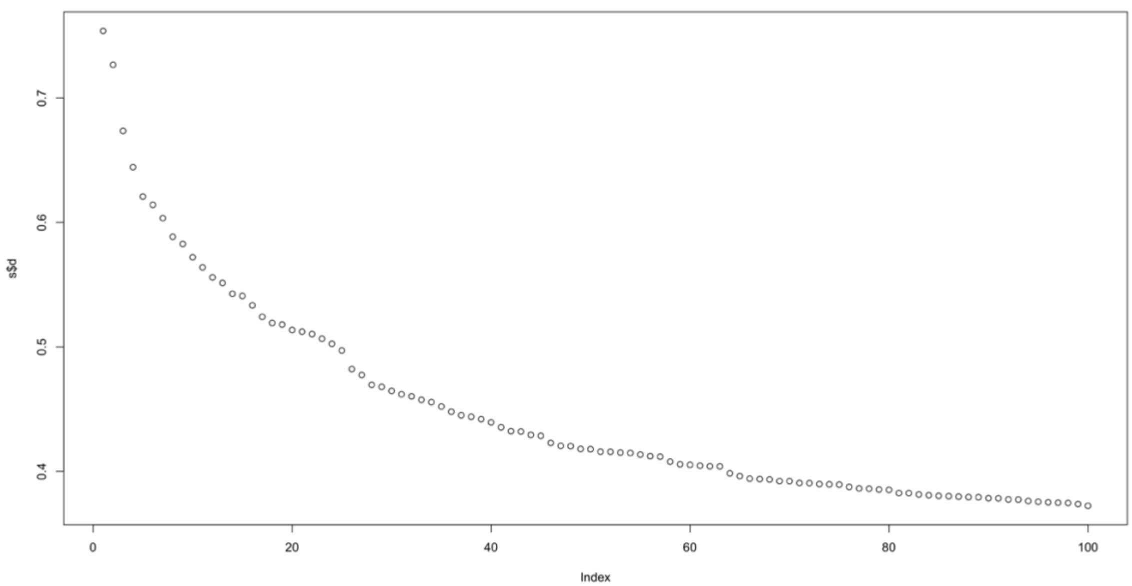

Figure S1. Scree plot for VSP clustering.

The table below presents the percent of Republican vs Democratic Senators, Congressmen, and Congresswomen within the top 500 accounts retweeted by each cluster. The lists of Senators, Congressmen, and Congresswomen and their Twitter handles come from the following papers and data repositories compiled by a researcher at NYU:

- Lasser, J., Aroyehun, S. T., Carrella, F., Simchon, A., Garcia, D., & Lewandowsky, S. (2023). From alternative conceptions of honesty to alternative facts in communications by US politicians. Nature human behaviour, 7(12), 2140-2151.
- https://triagecancer.org/congressional-social-media
- https://ucsd.libguides.com/congress_twitter
- https://github.com/leilasaoud/congress-kmeans-clustering/tree/master/data
- https://github.com/r-congress/congress116

These two metrics allow us to evaluate the performance of our VSP-based clustering method, as well as the UMAP-based approach (see Appendix II below). A Republican cluster should have more Republican politicians in the top 500 retweet accounts than a Democrat cluster does, and the same should hold for Democratic clusters (i.e. retweeting more Democrats than Republicans). As can be seen in the table below, all clusters of US-Democrats and US-Republicans follow this pattern.

Table S1. Clustering results using VSP-based clustering.

Cluster ID	Cluster Label	Percent of Republican politicians within the top 500 retweeted accounts (%)	Percent of Democratic politicians within the top 500 retweeted accounts (%)
1	India	0.19	0.00
2	**US-Democrats**	**0.00**	**8.12**
3	**US-Republicans**	**4.11**	**0.00**
4	India	0.19	0.64
5	Iran	0.75	0.43
6	mixed	0.00	0.00
7	mixed	0.37	1.28
8	India	0.19	0.00
9	UK-liberals	0.00	1.07
10	mixed	0.00	0.00
11	UK-conservatives	0.19	0.00

Cluster ID	Cluster Label	Percent of Republican politicians within the top 500 retweeted accounts (%)	Percent of Democratic politicians within the top 500 retweeted accounts (%)
12	mixed	0.37	0.21
13	positivity	0.19	3.21
14	India	0.19	0.21
15	mixed	0.00	0.21
16	Philippines	0.19	0.64
17	bots	0.00	0.64
18	mixed	0.00	0.43
19	mixed	0.00	0.00
20	India	0.19	0.64
21	India	0.19	0.00
22	India	0.19	0.00
23	India	0.19	0.21
24	Australia	0.00	1.07
25	UAE & India	0.00	0.00
26	Canada	0.93	0.00
27	India	0.00	0.00
28	news	0.56	2.99
29	India	0.00	0.00
30	mixed	0.00	0.21
31	**US-Democrats**	**0.00**	**1.92**
32	mixed	0.19	0.00
33	South Africa	0.19	0.21
34	India	0.19	0.85
35	Canada	0.19	2.35
36	India	0.19	0.21
37	India	0.19	0.21
38	news	0.00	3.21
39	India	0.19	0.85

Cluster ID	Cluster Label	Percent of Republican politicians within the top 500 retweeted accounts (%)	Percent of Democratic politicians within the top 500 retweeted accounts (%)
40	India	0.19	0.21
41	**US-Republicans**	**3.74**	**0.21**
42	Nigeria	0.19	0.43
43	India	0.19	1.07
44	mixed	0.00	0.21
45	Srilanka	0.19	0.64
46	UK-conservatives	0.37	0.21
47	India	0.19	0.43
48	mixed	0.00	1.50
49	news	0.19	2.14
50	UK-liberals	0.00	1.07

Appendix II. UMAP-DBSCAN clustering results

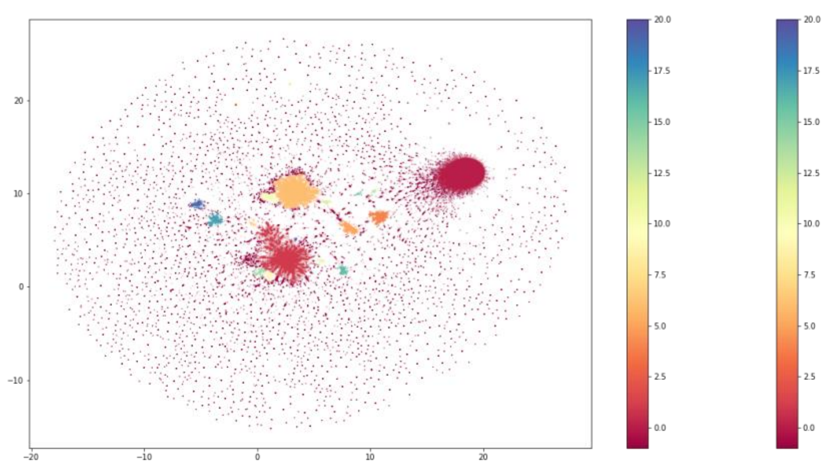

Figure S2. UMAP-DBSCAN clusters.

The table below presents the percent of Republican vs Democratic Senators, Congressmen, and Congresswomen within the top 500 accounts retweeted by each cluster. The table below shows a similar pattern as the table for VSP clustering above, suggesting the effectiveness of the clustering method.

Table S2. Clustering results using UMAP-DBSCAN clustering.

Cluster ID	Cluster Label	Percent of Republican politicians within the top 500 retweeted accounts (%)	Percent of Democratic politicians within the top 500 retweeted accounts (%)
-1	out_of_sample	0.56	4.27
0	**US-Democrats**	**1.12**	**5.77**
1	**US-Democrats**	**0.19**	**7.91**
2	out_of_sample	0.19	0.64
3	**US-Republicans**	**2.24**	**0.00**
4	UK-liberals	0.00	0.85
5	UK-conservatives	0.19	0.00
6	**US-Republicans**	**4.30**	**0.00**
7	out_of_sample	0.37	2.35
8	**US-Democrats**	**0.00**	**1.28**
9	**US-Democrats**	**0.00**	**6.62**
10	**US-Republicans**	**4.67**	**0.64**
11	**US-Democrats**	**0.00**	**5.77**
12	Canada	0.93	0.00
13	**US-Democrats**	**0.00**	**6.41**
14	**US-Democrats**	**0.00**	**5.77**
15	**US-Democrats**	**0.00**	**5.13**
16	India	0.00	0.21
17	India	0.19	0.21
18	India	0.00	0.00
19	India	0.00	0.00

Cluster ID	Cluster Label	Percent of Republican politicians within the top 500 retweeted accounts (%)	Percent of Democratic politicians within the top 500 retweeted accounts (%)
20	**US-Democrats**	**0.19**	**4.70**

Appendix III. Coding scheme and intercoder training process

Coding scheme

All tweets were coded sequentially; tweets that fit a category were coded as 1, and those that did not fit were coded as 0. The coding process was as such: First, the coders determined the relevance of the tweet (relevant = 1; non-relevant = 0). Relevant tweets were then identified as either US or non-US (US = 1; non-US = 0), and if they constituted risk (risk = 1; non-risk = 0). Next, only risk tweets were categorized into whether they were dread risk (1 = dread risk, 0 = non-dread risk) and/or unknown risk (1 = unknown risk, 0 = non-unknown risk). A tweet may comprise both dread and unknown risk. Lastly, the tweets were categorized into seven targets of risk concern in which a single tweet may fit into more than a single category of concern. Below are specific instructions the coders followed.

General notes:

- When a tweet is ambiguous regarding a variable. Code 0.
- Focus on the literal meaning of the text. Do not infer what's not been said. Only code based on what's being said literally.
- Code based on the whole sentence, don't get hung up on part of it. Focus on the "main clause" of the sentence.
- In the case of a quote tweet: USER A: X RT@USER B: Y. For V1 RELEVANCE, if X OR Y is relevant, code 1. For all the other variables, both X and Y must be considered. If Y is about risk, but X counters it, code 0 for V2 RISK.
- Consider only words and hashtags.

  - Ignore URLs. Do not look at them nor click them. They will be removed in the machine learning process.
  - Ignore all @ mention handles.
- As long as the majority of a tweet is in English, treat it as English tweets. A few unknown words should not be the criterion to code it as 1 for Non-US.
- If a tweet is truncated, don't infer the truncated part. But in the case of a truncated question like "how many people have …" code it as "unknown" because "how many" is all that we need to know it's a question.

V1.1 Relevance (only code relevant tweets for the other variables)
About any aspect of pandemic, such as the virus, vaccine, public health guidelines, school closure, reopening etc. See the specific categories in the V3 variables.
Mentioning covid in passing also counts
Even though the tweet is not related to covid, code it 1 for relevance if it uses words from the "covid" vocabulary.

V1.2 Non-US
Primarily in non-English words (if a tweet contains only a couple of non-English words, code based on the context)
Explicitly talking about subject matters occurring <u>exclusively</u> in other countries
Even if a US public figure is present, as long as the subject matter is not about US, code non-US

V2 Risk (V2 Risk = V4.1 dread + V4.2 unknownness)
The main message is explicitly about threat, harm, hazard and uncertainty, pertaining to the V3 variables: the virus and its variants, vaccines, public health guidelines/policies, economy, public figures and institutions, science and scientific community, and media and information.

- Mention risk in passing should be risk: "the side effects of vaccines" should be risk
- A "factual" statement with a clear valence should be coded based on the valence.

- A statement about other people expressing concern should be risk. “RT @GroveKilosdad73: So, the Pitbull heard Governor Cuomo saying he didn't want the vaccine for Coronavirus and he didn't trust it because”
- Tongue-in-cheek expression is not risk
- When the user themselves is certain about another user’s lack of knowledge/uncertainty, code 0 for risk. (e.g., “I know Fauci doesn’t know what he’s talking about”)

V3.1. cases, spread, diagnosis & symptoms
cases, testing, deaths, spread, diagnosis & symptoms; ways of “measuring” the scale of the pandemic problem in some fashion.

V3.2. vaccine
safety, efficacy, rollout, and development

V3.3. economy
Jobs & business, unemployment, prices, etc.

V3.4. public health guidelines and policies
lockdown, reopening, school closure, mask mandate, individual compliance

V3.5. political figures and public institutions (can double code with V3.4)
local, state, and federal gov. office holders and candidates.
Government units and offices

V3.6. science and the scientific community (can double code with V3.4)
CDC, WHO, individual public health experts

V3.7. media and information
Biased news media, fake news, misinformation, rumors

V4.1 dread

Dread risk usually constitute:

- Risks that are catastrophic, uncontrollable, and potentially fatal
- Associated with negative emotions such as fear, anxiety, and dread
- Examples: Nuclear accidents, terrorist attacks, natural disasters, epidemics, pandemics

Tweets should contain:

- Diagnosis and cases, unless there's explicitly positive expression
- Something should have been done or should be done; someone should have done or do something (lack of control)
- Negative emotions: anxiety/fear/anger/defiance/dissatisfaction/indignation
- Public health mandates: lockdown, mask, school/restaurant closure (suggesting uncontrollability)
- Expressions that suggest a catastrophic outcomes, and are potentially fatal (e.g., deaths)
- Examples of dread risk tweets:
  - Joining @AC360 with @drsanjaygupta: There is good news with #covid19 daily infections & hospitalizations finally declining, but January was the deadliest month with over 90,000 deaths. We should be very concerned with more contagious variants—expediting vaccine rollout is key https://t.co/Q5ih1KOg8j
  - "“We are going to see cases go up. The virus is going to continue spreading.” Professor @devisridhar says lifting the coronavirus lockdown is “pointless” if we do not have the proper infrastructure to detect and trace. #bbcqt https://t.co/cycMG5q5gG"

V4.2 Unknownness

Unknown risk usually constitute:

- Risks often associated with emerging technologies or phenomena not well understood or have not been experienced before
- Often accompanied by high degree of uncertainty and ambiguity, thus people lean on heuristics to make judgements about unknown risks
- Examples: Genetically modified organisms, nanotechnology, and emerging infectious diseases

Tweets should contain:

- Bad things "could" happen (projection) (e.g., - A 113% higher risk of hospitalization)
- Bad things happen very fast (e.g., 100 more cases in 1 day, a week; mention of a certain timeframe)
- The "tweeter" not sure about something (using words like if, possible, maybe, likely, seems, etc.)
- Things that are new and unfamiliar in an unpleasant way
- Questions about risk. However, if the user asks a question and then answers it, code 0. If a question is truncated but it's clearly a question, code 1 for unknownness. "How many lives have died…"
- The "tweeter" frames the object in tweet as not well understood or something that has not been experienced before
- Examples of unknown risk tweets:
    - Dr. Peter Hotez: 'We're going to move towards 2,000 deaths a day … Within a few weeks, COVID-19 will be the single leading cause of death on a daily basis in the United States.' https://t.co/T4TMiKBCCp
    - If it's true (but I doubt it) that Dumb Dumb Donald has COVID-19 I don't know why he's going to the hospital since he only had to inject bleach in his veins in order to be cured! https://t.co/N5dAncyH0v

Intercoder training process

The intercoder training process started in late July 2021 and ended in late August 2021. Two authors (graduate students with experience in qualitative analysis) completed six rounds of coding. During each round, they independently coded 200 randomly sampled tweets. A meeting was held after each round to resolve any disagreements between the two authors. A high intercoder reliability was achieved (Krippendorff's alpha ranging from 0.89 to 1, mean alpha = 0.95) during the final session (Round 6), and the coders then subsequently coded 1500 tweets each independently. Below we report the Krippendorff's alphas from all the six rounds of coder training:

Round 1:

Variable	Krippendorff's alpha	Number of tweets the coders agreed on
V1. relevance	0.67	170
V2. risk	0.65	124
V3.1. disease & symptoms	0.53	167
V3.2. vaccine	0.53	193
V3.3. economy	0.39	192
V3.4. public health guidelines	0.66	155
V3.5. political figures	0.56	173
V3.6. media	-0.02	193
V3.7. science and public health	0.5	187
V3.8. misc	-0.05	108
V4.1. dread	0.48	129
V4.2. unknownness	0.44	121
V5.1. virus characteristics	0.8	197
V5.2. virus spread	0.27	195
V5.3. virus origins	0.5	197
V5.4. vaccine	0	198
V5.5. public authority	0.5	197
V5.6. misc	1	199

Round 2:

Variable	Krippendorff's alpha	Number of tweets the coders agreed on

V1.1 relevance	0.88	187
V1.2 non-US	0.69	184
V2.1 risk	0.67	139
V2.2 affect	0.43	192
V4.1. dread	0.62	135
V4.2. unknownness	0.3	124
V3.1. cases, spread, diagnosis & symptoms	0.65	158
V3.2. vaccine	0.69	194
V3.3. economy	0.56	184
V3.4. public health guidelines	0.68	155
V3.5. political figures	0.77	179
V3.6. media	0.29	190
V3.7. science and public health	-0.02	189
V5.1. virus characteristics	-0.01	197
V5.2. virus spread	1	200
V5.3. virus origins	0.32	195
V5.4. vaccine	0	199
V5.5. public authority	1	200

Round 3:

Variable	Krippendorff's alpha	Number of tweets the coders agreed on
V1.1 relevance	0.85	188
V1.2 non-US	0.75	173
V2.1 risk	0.67	131
V4.1. dread	0.65	131
V4.2. unknownness	0.22	157

Round 4:

Variable	Krippendorff's alpha	Number of tweets the coders agreed on

V1.1 relevance	0.85	191
V1.2 non-US	0.78	190
V2.1 risk	0.67	167
V4.1. dread	0.64	164
V4.2. unknownness	0.37	171

Round 5:

Variable	Krippendorff's alpha	Number of tweets the coders agreed on
V1.1 relevance	0.95	197
V1.2 non-US	0.8	190
V2.1 risk	0.91	191
V4.1. dread	0.9	190
V4.2. unknownness	0.79	195

Round 6:

Variable	Krippendorff's alpha	Number of tweets the coders agreed on
V1.1 relevance	0.94	195
V1.2 non-US	0.96	198
V2.1 risk	0.93	193
V4.1. dread	0.93	193
V4.2. unknownness	0.89	197
V3.1. cases, spread, diagnosis & symptoms	1	200
V3.2. vaccine	0.95	199
V3.3. economy	1	200
V3.4. public health guidelines and policies	0.9	195
V3.5. political figures and public institutions	0.95	198
V3.6. science and the scientific community	1	200
V3.7. media and information	1	200

Appendix IV. Coding of the top 500 retweeted accounts by Democrats and the top 500 retweeted accounts by Republicans.

Table S3. Types of accounts in the 500 retweeted accounts.

Account Type	Coding Number	Account Type Description
Politician, Political Affiliate, or Government	1	Current and former officeholders and candidates for elected office, as well as those appointed by such persons (ex: cabinet member, spokesperson). Includes elected roles such as president, vice president, governor, senator or representative (both state and federal), judges, etc. Also includes campaign staff (campaign managers, political strategists, leaders of political party organizations) and accounts of government agencies (ex: the Department of Justice)
Political Media	2	Producers, editors, journalists, hosts, anchors, columnists, correspondents, etc. who work or produce content for politically-oriented media outlets, such as newspapers, radio and TV news stations, online-only news media, YouTube video series, and podcasts. Also includes accounts of media outlets themselves (ex: @MSNBC), as well as authors of politically-oriented books.
The Academic and Scientific Community	3	Users who identify as professors or university-affiliated researchers, or else those with some kind of doctroral degree in a specialized field (including MDs and lawyers, who generally must earn a juris doctor degree to practice law).
Activist	4	Users who self-identify as “activists” or “organizers”, those involved with political activism through social movements, advocacy groups, lobbying, PACs, NGOs, etc., and anyone with an activist hashtag (ex: #BLM, #NRA, #StopAsianHate) or political statement/claim in their profile description. Simply listing one's ideology (ex: "proud

Account Type	Coding Number	Account Type Description
		conservative") does not qualify someone as an activist.
Entertainment	5	Non-political persons of prominence, including athletes, actors, musicians, non-political authors, artists, comedians, etc. Further includes persons and groups who produce satirical political news, since their sole purpose is to entertain rather than inform.
Twitter Influencer	6	Any users who would be categorized as "Other" for not meeting the criteria for the above categories, but who have over 100,000 followers currently.
Other	7	Any users who do not fit into the categories outlined above. This would include those accounts of more-or-less regular people and those without any profile description text or supplemental information on a Wikipedia page.
Suspended	8	Any users whose profile information can't be retrieved via Twitter API as of Nov 12. This suggests that they have been suspended or have deleted/deactivated their accounts.

Table S4. Coding Rules & Procedures.

Coding Rules & Procedures
1) The general pattern of coding users for their categorical description will be to give priority for the first valid identifier that appears in the field being assessed. If the first identifier describes a previously-held role (ex: by indicators such as "former" or "retired"), then code for the first valid identifier that signals a currently-

held role. If all indicators describe previously-held roles, then code for the first valid identifier mentioned by the user. Users may only be assigned one single categorical label.
2) Begin by assessing the user's account description (rt_description) that appears in our dataset. If the account can be categorized based on the text that appears in this field, no further digging is required. Include information contained in hashtags (ex: "#BLM") or @ mentions (ex: "Editor at @NBCNews"); conducting Google searches of these hashtags and @ mentions is allowed in order to provide a categorization if the coder is unsure what the user is referencing.
3) If the user's account description provides insufficient information to codify them, look next to their username for identifiers (ex: Dr., MD, Rep, Senator, etc.) and code accordingly.
4) Failing a categorization so far, we will next Google search the user (based on the name provided in the "rt_uname" field of the dataset) and assess their Wikipedia page, if one exists. Should a Wikipedia page not exist for a user, checking to see if they have a personal website will be the next step. Code based on the practices outlined in step 1 for the first available information that appears on the user's Wikipedia page or personal website.
5) If we are still unable to categorize a user at this point, they will be coded under the "Other" or "Twitter/Social Media Influencer" categories depending upon how many followers their account has.
6) Take note of any awards the user lists in their account (ex: "Pulitzer Prize winner", "Nobel Laureate") and categorize based on this information if applicable. For instance, Nobel Laureates are typically awarded to academics, which should place a user in the "The Academic and Scientific Community" category.
7) Role-relevant labels (ex: "governor", "editor", "musician") override role-irrelevant labels (ex: "fisherman", "mother").

Appendix V

Table S5. BERT supervised machine learning performance metrics (best output performance).

	label	precision	recall	f1-score	accuracy
relevance	0	0.91	0.80	0.86	0.95
	1	0.96	0.98	0.97	

	label	precision	recall	f1-score	accuracy
non-US	0	0.99	0.93	0.96	0.94
	1	0.75	0.95	0.84	
dread risk	0	0.95	0.79	0.86	0.88
	1	0.85	0.97	0.90	
unknown risk	0	0.96	0.94	0.95	0.91
	1	0.57	0.67	0.62	
covid	0	0.97	0.90	0.93	0.90
	1	0.72	0.89	0.80	
vaccine	0	0.98	1.00	0.99	0.97
	1	0.94	0.19	0.31	
economy	0	0.96	0.99	0.98	0.96
	1	0.81	0.37	0.51	
public health	0	0.94	0.97	0.96	0.93
	1	0.84	0.72	0.78	
politicians	0	0.94	0.98	0.96	0.93
	1	0.79	0.58	0.67	
science	0	0.99	0.99	0.99	0.99
	1	0.64	0.63	0.64	
information	0	0.98	0.99	0.99	0.97
	1	0.74	0.54	0.63	

Appendix VI

Table S6. Summary statistics based on BERT supervised machine learning.

	US Democrats	US Republicans
number of classified users	333,564	73,065
number of classified relevant tweets	1,704,359	429,215
number of non-US related tweets (percentage of all relevant tweets)	93,880 (6%)	32,441 (8%)
number of classified "dread-risk" tweets (percentage of all relevant tweets)	1,318,093 (77.3%)	355,398 (82.8%)
number of classified "unknown-risk" tweets (percentage of all relevant tweets)	204,941 (12.0%)	71,141 (16.6%)
number of classified "covid" tweets (percentage of all risk tweets)	697,806 (51.8%)	130,327 (35.8%)
number of classified "vaccine" tweets (percentage of all risk tweets)	1699 (0.1%)	2530 (0.7%)
number of classified "economy" tweets (percentage of all risk tweets)	40,196 (3.0%)	5266 (1.4%)
number of classified "public health" tweets (percentage of all risk tweets)	290,273 (21.5%)	98,852 (27.1%)
number of classified "politicians" tweets (percentage of all risk tweets)	306,124 (22.7%)	51,713 (14.2%)
number of classified "science" tweets (percentage of all risk tweets)	45,468 (3.4%)	22,280 (6.1%)
number of classified "information" tweets (percentage of all risk tweets)	16,622 (1.2%)	33,026 (9.1%)

Appendix VII

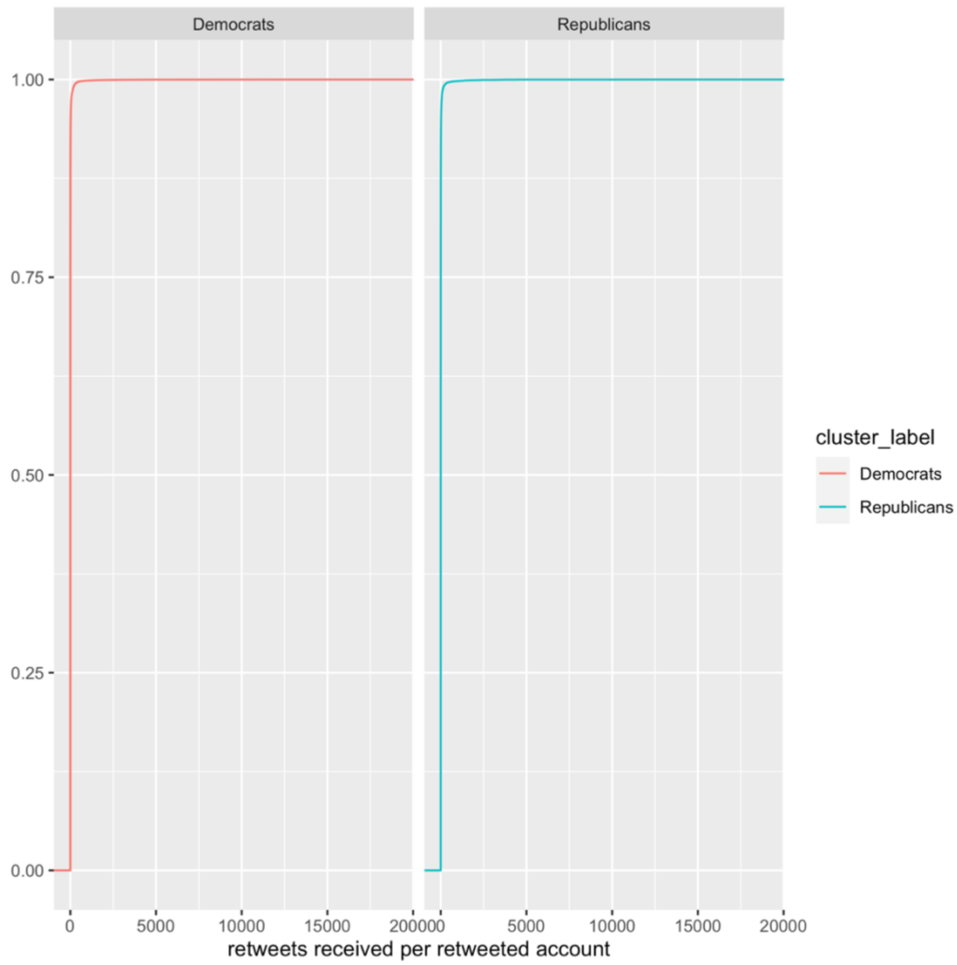

Figure S3. Cumulative Frequency Distribution of Retweets Per Retweeted Account.