Original Research Article

Toxicity and verbal aggression on social media: Polarized discourse on wearing face masks during the COVID-19 pandemic

Paola Pascual-Ferrá[1] ⓘ, Neil Alperstein[1] ⓘ, Daniel J Barnett[2] ⓘ
and Rajiv N Rimal[3] ⓘ

Abstract
Medical and public health professionals recommend wearing face masks to combat the spread of the coronavirus disease of 2019 (COVID-19). While the majority of people in the United States support wearing face masks as an effective tool to combat COVID-19, a smaller percentage declared the recommendation by public health agencies as a government imposition and an infringement on personal liberty. Social media play a significant role in amplifying public health issues, whereby a minority against the imposition can speak loudly, perhaps using tactics of verbal aggression taking the form of toxic language. We investigated the role that toxicity plays in the online discourse around wearing face masks. Overall, we found tweets including anti-mask hashtags were significantly more likely to use toxic language, while tweets with pro-mask hashtags were somewhat less toxic with the exception of #WearADamnMask. We conclude that the tensions between these two positions raise doubt and uncertainty around the issue, which make it difficult for health communicators to break through the clutter in order to combat the infodemic. Public health agencies and other governmental institutions should monitor toxicity trends on social media in order to better ascertain prevailing sentiment toward their recommendations and then apply these data-driven insights to refine and adapt their risk communication messaging toward mask wearing, vaccine uptake, and other interventions.

Keywords
COVID-19, infodemic, social media, toxicity, verbal aggression

This article is a part of special theme on Studying the COVID-19 Infodemic at Scale. To see a full list of all articles in this special theme, please click here: https://journals.sagepub.com/page/bds/collections/studyinginfodemicatscale

Introduction

Face masks help combat the spread of the coronavirus disease of 2019 (COVID-19) (American Medical Association, 2020; American Public Health Association, 2020; Pandemic Action Network, 2020). Public health campaigns from the Centers for Disease Control and Prevention (CDC) in the United States and the World Health Organization (WHO) promote wearing face masks in public, paired with social distancing and frequent handwashing, as protective measures (Centers for Disease Control and Prevention, 2020; WHO, 2020a). Public compliance with face mask-wearing guidelines, however, remains a challenge in many settings. According to the Pew Research Center and Gallup, about 85% of people surveyed during the month of August 2020 in the United

[1]Department of Communication, Loyola University Maryland, Baltimore, MD, USA
[2]Department of Environmental Health & Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[3]Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Corresponding author:
Paola Pascual-Ferrá, Communication Department, Loyola University Maryland, 4501 N Charles St, Baltimore, MD 21210, USA.
Email: ppascualferra@loyola.edu

States reported wearing masks either "always" or "most of the time" when in stores and businesses, but only 47% did so outdoors (Kramer, 2020; Reinhart, 2020). They also found that mask-wearing habits differed by gender, level of education, geographic location, age, income, and political affiliation. According to a Gallup poll conducted in the United States from 29 June 2020 through 5 July 2020, about a third of men (34%), less than a third of Republicans (24%), about a third of people living in the U.S. Midwest region (33%), and 44% of people with annual household incomes over $90,000 surveyed reported "always" wearing masks outside the home (Brenan, 2020). In this study, we investigated the role that toxicity plays in the online discourse around wearing face masks.

Background on face mask controversy

Wearing face masks to slow the spread of COVID-19 has been controversial in the United States since the beginning of the pandemic. Initially, the science around the effectiveness of using face masks as a way to slow down the spread of COVID-19 by everyone other than essential health workers during daily routine activities was unclear and so was the advice given by public health organizations. First responders and medical service providers adopted the use of face masks early on as part of their personal protective equipment. The emerging pandemic led to a buying frenzy of hand sanitizers, disposable gloves, face masks, and toilet paper (Meyersohn, 2020). A shortage of N95 face masks, considered by health experts to be the most effective barrier against COVID-19, became a challenge for the medical community as they competed with nonessential workers and individual consumers for masks. At one point, the U.S. surgeon general pleaded on Twitter, "Seriously people—STOP BUYING MASKS!" (Cramer and Sheikh, 2020). Other public health officials and organizations emphasized that N95 masks should be reserved for frontline workers (Hill, 2020b; Nazaryan, 2020). Social media posts and news stories seemed to suggest face masks could do more harm than good (Gillespie, 2020) and create a false sense of security (Khan, 2020). One story focused on a car accident "believed to have resulted from the driver wearing an N95 mask for several hours and subsequently passing out behind the wheel due to insufficient oxygen intake/excessive carbon dioxide intake" (Carrega, 2020), which went viral on social media (Gillespie, 2020). The message by public health officials regarding the use of face masks by everyone else other than essential health workers was reversed as the science became clearer and the N95 face mask shortage crisis subsided. But the damage was already done; statements from public health officials and leading public health agencies casting doubt on the effectiveness of face masks were used by anti-mask groups as arguments for rejecting public health guidance surrounding face masks and helped fuel conspiracy theories.

Another reason for the controversy over wearing face masks had to do with the political environment in which this controversy played out. In many countries, governments and public establishments at first advised, and then mandated, the use of face masks. For many anti-establishment, anti-government groups, the mandate to wear face masks in public was interpreted as an affront to personal liberty. In the United States, with a highly contested general election taking place in November 2020, face masks became a symbol infused with political meaning (Aratani, 2020; Martinelli et al., 2021: 4). The Democratic presidential nominee, Joe Biden, frequently wore a face mask in public and advocated for its use, while the Republican presidential nominee, Donald Trump, chose to do the opposite and mocked Biden publicly for wearing masks so often: "Everytime you see him, he's got a mask" (Hill, 2020a; see also Bennett, 2020). While wearing face masks has been associated with kindness, empathy, and prosocial behavior (Biggers, 2020; Pfattheicher et al., 2020), anti-maskers scorned face mask-wearing guidelines and mandates as anti-democratic, a violation of individual freedoms, and unconstitutional (Mogelson, 2020; Stewart, 2020).

Use of hashtags

On social media, hashtags are often used as a communication tactic to increase public engagement with an issue, company, or organization. Most often, the use of a particular hashtag indicates where the author of the post stands on an issue. Hashtags also serve as markers of group identity. Rogers (2018) refers to the concepts of positioning and counterpositioning as ways in which actors on opposing sides of an issue insert themselves into the social media debate. From a research standpoint, a hashtag search can provide valuable information about public opinion on an issue (e.g. who are the most frequent posters, what is the sentiment, how often people post about it), as well as a measure of the public agenda (e.g. the extent to which a particular topic remains in the public consciousness). Tracking hashtags over time can provide a glimpse into how the focal issue's popularity is rising and falling.

In the United States, two prominent public health campaigns on social media that promoted the use of face masks during 2020 included the hashtag #MaskUp, promoted by the American Medical Association, and the hashtag #WearAMask, promoted by the CDC. These efforts were amplified by the voices

of public opinion leaders, public and private organizations, celebrities and other influencers who posted content on social media that included pro-mask-wearing messages and any of these or similar hashtags. Other hashtags on social media that were frequently used to promote the use of face masks included #WearAMaskPlease, #WearAMaskSaveALife, and #WearADamnMask, among others. Groups who resisted government guidelines and recommendations about wearing face masks frequently used the hashtags #NoMask, #BurnYourMask, #WeDoNotConsent, #WeWillNotComply, #MaskOff, and #MasksDontWork, among others.

Theoretical concepts and models

Several theoretical concepts and models inform our study of toxicity in comments made about face mask use on social media. In the context of online discussions, toxicity is defined as "a rude, disrespectful, or unreasonable comment that is likely to make you leave a discussion" (Jigsaw LLC, 2020a). Researchers have applied the concept of toxicity to the study of anti-social behavior on Reddit (Almerekhi et al., 2019; Gruzd et al., 2020; Mall et al., 2020), Twitter (Guberman et al., 2016), and YouTube (Salminen et al., 2020). Guberman et al. (2016) conceptualized toxicity as online harassment and a form of verbal violence; they adapted existing measures of cyber- and verbal aggression to develop a scale for measuring toxicity online that consists of four dimensions: anger, hostility, physical aggression, and verbal aggression.

Verbal aggression has been studied extensively by researchers in interpersonal communication. As a theoretical construct, *verbal aggressiveness* is conceptualized and measured as a trait and distinguished from verbal aggression (action). It is defined as "a personality trait that predisposes persons to attack the self-concepts of other people instead of, or in addition to, their positions on topics of communication" (Infante and Wigley, 1986: 61). Infante and Wigley distinguished between argumentativeness and verbal aggressiveness in their locus of attack; while the former focuses on the positions taken by others in an argument, the latter focuses on attacking the self-concept of others to make them feel badly about themselves. In other words, instead of debating the argument, the verbally aggressive person aims to "destroy" or "crush" the other person and his or her self-esteem. Infante et al. (1990: 367) identified 10 different types of verbally aggressive messages: character attacks, competence attacks, background attacks, physical appearance attacks, maledictions, teasing, ridicule, threats, swearing, and nonverbal emblems. Hamilton (2012: 6)

argued that verbal aggressiveness constitutes a threat to public health:

> . . . verbal aggression imperils public health. Verbal aggression and physical aggression harm society individually and collectively. Aggressive language, with its penchant to reverberate over social media, can damage the self-concept of its victims. Verbal aggression threatens to destroy civil discourse in groups and large organizations. It polarizes factions toward extremism, bringing strife and ultimately paralysis to institutions. Between cultures, verbal aggression can spiral out of control, leading to bloodshed or even full-scale war. In short, the incendiary effects of excessive verbal aggression represent an imminent danger to civilized society.

Hamilton et al. (2008) proposed a model in which verbal aggression leads to "social distancing" such as avoiding interactions or conversations with others (not to be confused with "social distancing" measures recommended in the context of COVID-19).

From the risk communication literature, we find that toxicity conceptually aligns with Peter Sandman's (1993) "Risk = Hazard + Outrage" model, in which he formulates outrage as a constituent element of risk perception. Sandman frames risk perception as the sum of a risk-laden situation (hazard) and the magnitude of people's responses to the given hazard as mediated by the extent to which these situations engender outrage among the public (outrage). Studies have applied Sandman's framework in the context of COVID-19 to improve risk communication (Malecki et al., 2021), analyze the amplification of social risk on the social media platform Weibo (Fan, 2020), and formulate communication strategies for family physicians (Ledford and Anderson, 2020), among others. Sandman (2020) himself has weighed in on the debate over face masks stating that,

> The mask issue is politicized on both sides. It's not just the right claiming that masks infringe unacceptably on their freedom; it's also the left claiming that failing to wear a mask infringes unacceptably on everybody else's safety. The middle position has few supporters.

Based on Sandman's "Risk = Hazard + Outrage" model, it is important for risk communicators and public health professionals to know about the nature of the environment they are entering when they send or receive communication about emerging public health issues, particularly when outrage is present. Sandman indicates that the public tends to focus more on outrage, while public health officials and institutions tend to focus more on the hazard. In the context of COVID-19, hazard refers to the objective harm caused by the

virus, while outrage refers to the effect of people's perception of COVID-19 as dangerous. Public perception about the hazard could match the real level of the hazard, undermine it or inflate it, leading to outrage. When left unmanaged, outrage could lead to undesirable outcomes such as panic and/or loss of credibility and authority of public health officials and institutions. Sandman (2015) warns that social media tends to exacerbate outrage and considers risk communication on social media platforms "a crucial outrage management tool – perhaps the most crucial." He offers that organizations should continuously monitor social media for outrage and be able to respond to it quickly and appropriately. When analyzing social media discourse, toxicity can serve as a measure of outrage.

Finally, in the social sciences, the concept of pluralistic ignorance (Allport, 1924; Prentice and Miller, 1996) is used to explain the finding that people systematically overestimate or underestimate the level of support there exists in society for a particular position on an issue, relative to their own support for the issue. For example, people often posit that, while they themselves hold anti-discriminatory attitudes, many others do not (Fields and Schuman, 1976). Similarly, college students routinely overestimate the level of alcohol consumption among their peers, a tendency stronger among drinkers (Perkins et al., 2005). One specific example of this phenomenon is the false-consensus effect (Marks and Miller, 1987; Ross et al., 1977), which refers to people's beliefs that support for their deviant behavior is more widespread than it actually is. Research has also found that those holding minority opinions are more vocal and that they "experience greater comfort and pride in expressing their opinions" (Miller and Morrison, 2009: 741). One means of expressing one's opinions more forcefully is through heightened toxicity and anti-social behavior online. Finally, studies have concluded that individual personality traits such as narcissism, when combined with a feeling of social rejection, are powerful predictors of aggressive behavior (Twenge and Campbell, 2003; see also Okada, 2010).

The current study

The WHO (2020b) defines infodemic as follows: "an overabundance of information – some accurate and some not – occurring during an epidemic." Most COVID-19 infodemic studies are focused on misinformation (incomplete or inaccurate) and disinformation (intentional falsehoods) (for representative studies, see Cuan-Baltazar et al., 2020; Hernández-García and Giménez-Júlvez, 2020; Sousa-Pinto et al., 2020). Some studies have investigated the toxicity of COVID-19-related Twitter discourse (Awal et al., 2020; Guerrero-Solé and Philippe, 2020; Majó-Vázquez et al., 2020),

with one in particular focusing on Sinophobic behaviors (Schild et al., 2020). However, no studies to our knowledge have focused on the toxicity of discourse on Twitter around face mask wearing. Yet, wearing face masks as a preventive measure is an important behavior from both a public health perspective and one around which so many politically charged sentiments are expressed.

Following the theoretical concepts and models discussed above, we believe that toxicity adds to the current infodemic in several ways. First, it creates a hostile environment that turns users away from online conversations about the issue and/or may distract them from acquiring factual, evidence-based information about face mask wearing as an effective measure to stop the spread of COVID-19. The use of toxic language leads to verbal distancing online and increased defensiveness among those engaged on this issue in the form of social distancing as proposed by Hamilton et al. (2008). Toxic language can impact the self-concept of those who are or feel targeted, as suggested in the work of both Infante and Wigley (1986) and Hamilton et al. (2008). Based on Infante and Wigley's work, we believe that the impact of toxic language on an individual's self-concept can result in the individual: (1) joining the fray in agreeing with what is being said and perhaps amplifying the toxicity; (2) withdrawing into a defensive posture; or (3) speaking back against what is being expressed. Second, those who hold anti-mask-wearing positions (in the minority group), and who may be feeding mis- and disinformation, are more likely to express their opinions more forcefully (Miller and Morrison, 2009: 741), perhaps using more toxic language than those who hold pro-mask-wearing positions. Third, because toxicity can serve as a measure of outrage (Sandman, 1993), it should be continuously monitored and managed by public health agencies and organizations tasked with fighting the infodemic. Hence, this study aimed to answer the following research questions (RQs):

> RQ1: How do toxicity scores compare between tweets using anti- vs. pro-mask sentiment hashtags?

> RQ2: How do the toxicity scores compare across tweets when analyzed by the individual hashtags?

Specifically, in addressing RQ1, our hypothesis (H) was that:

> H1: There would be a statistically significant difference in mean toxicity scores between tweets with anti- and pro-mask sentiment hashtags, with the former having a higher mean toxicity score. We believed this to be the case since the anti-mask stance is contrarian in nature, going against government recommendations and the views of the majority, and research shows that those

Table 1. Matrix for interpreting social media messages adapting Sandman's framework and Gruzd, Mai, and Vahedi's thresholds for high toxicity.

	Low outrage	High outrage
Low hazard	Messages that do not make reference to or undermine the danger or threat of COVID-19 and have low toxicity scores ($\leq$.30)	Messages that do not make reference to or undermine the danger or threat of COVID-19 and have high toxicity scores ($\geq$.70)
High hazard	Messages that make reference to or emphasize the danger or threat of COVID-19 and have low toxicity scores ($\leq$.30)	Messages that make reference to or emphasize the danger or threat of COVID-19 and have high toxicity scores ($\geq$.70)

COVID-19: coronavirus disease of 2019.

in the minority tend to be more vocal and prouder when expressing their opinions (Miller and Morrison, 2009: 741).

In addressing RQ2, our hypotheses were that:

H2: There would be a statistically significant difference in mean toxicity scores between tweets using #WearADamnMask and other pro-sentiment tweets, with the former having a higher mean toxicity score. We expected this to be the case given that the inclusion of the word "damn," considered a swear word, would increase toxicity scores. Because of this, we also expected this hashtag would attract users who felt most strongly about the issue and who would be more inclined to use toxic language in their tweets.

H3: There would be a statistically significant difference in mean toxicity scores between tweets using #WeDoNotConsent/#WeWillNotComply and other anti-mask sentiment tweets, with the former having a higher mean toxicity score. We believed this to be the case because, of all the different hashtags, these are directed against government policies and may be conflated with other issues and policies causing increased frustration and anger (outrage) among individuals.

H4: There would be no statistically significant difference in mean toxicity scores between tweets using #WeDoNotConsent/#WeWillNotComply and #WearADamnMask. We believed that those using these hashtags would be similar in the intensity of emotion expressed and level of toxicity.

H5: There would be a statistically significant difference in mean toxicity scores between #NoMask and #WearAMask, with the former having a higher mean toxicity score. Again, we expected users espousing minority opinions to be more vocal and use more toxic language than those espousing the majority opinion.

In addition, we were interested in looking at the views expressed in the tweets themselves and seeing if and

how those fit within Sandman's (1993) "Risk = Hazard + Outrage" framework. To that end, we created a matrix for interpreting toxicity scores, adapting Sandman's framework, Gruzd et al.'s (2020) $\geq$.70 threshold for high toxicity, and the proposed $\leq$.30 threshold for low toxicity (see Table 1). In terms of Sandman's framework, tweets expressing incredulity about the reality or danger of COVID-19 but using toxic language regarding face mask measures would be considered as low hazard/high outrage (e.g. a tweet that includes a combination of hashtags such as #MaskHoax and #MaskOff in the same tweet). We would expect anti-mask sentiment tweets in that category to have high toxicity scores of $\geq$ .70, based on the postulation that people are angry because of what they deem to be an unnecessary imposition on and/or infringement on their individual liberties. Similarly, tweets confirming COVID-19 as a real and dangerous threat and expressing anger towards those who resist wearing face masks in public would be considered as high hazard/high outrage (e.g. a tweet that includes a combination of hashtags such as #COVIDisreal and #WearADamnMask) and also expected to have high toxicity scores of $\geq$.70.

On the other hand, we expected tweets coming from the accounts of verified public health organizations and agencies to fall mostly under the high hazard/low outrage category (toxicity scores $\leq$.30), expressing the real danger and threat of the virus but tempering the tone of tweets to be more informational and educational regarding the use of face masks and other preventive measures. We expected those tweets to have low toxicity scores of $\leq$ .30. Finally, we did not expect to find many tweets falling under the first quadrant, given that few motivations to post would exist when one perceives the situation to be without hazard and not worthy of eliciting outrage.

Methods

For this study, we investigated the toxicity of messages regarding face mask wearing in the context of COVID-19 on Twitter. We were interested in identifying anti-

vs. pro-mask sentiment around public communication events that we believed would create conversations and debate over face masks. We computed the toxicity level of messages that expressed a specific position (for or against face masks) as evidenced by the hashtag used. First, we used Hashtagify (CyBranding Ltd., 2020) to identify the most popular hashtags used to express pro- and anti-face mask sentiment on Twitter. The website allows researchers and marketing professionals to track hashtags over time and compare their popularity and use with similar hashtags. Hashtagify defines popularity as frequency of use and assigns a score ranging from 0 (not used at all) to 100 (most used or most popular), relative to the most popular hashtag on Twitter at the time (CyBranding Ltd., 2020). Second, we used Netlytic, a web-based software that collects publicly available data from social media sites such as Twitter and helps researchers build, visualize, and analyze communication networks using social network analysis, to build our corpus of tweets for analysis (Gruzd, 2016).

We focused on tweets posted around four key time periods related to the COVID-19 pandemic: (a) change of advice by public health organizations regarding the use of face masks to slow the spread of the virus (T1, 18–24 July), (b) the first United States presidential debate followed soon after by President Donald Trump's announcement over Twitter that he had contracted the coronavirus and his stay and treatment at Walter Reed National Military Medical Center (T2, 29 September–5 October), (c) the U.S. 2020 general election and its immediate aftermath (T3, 3–11 November), and (d) the U.S. Thanksgiving 2020 holiday with the increased warnings and restrictions regarding travel and gatherings (T4, 25–28 November). For each of the data collection periods (i.e. T1–T4), we parsed the datasets into anti- and pro-mask sentiment based on the hashtags identified. We drew equal sample sizes from each of the different datasets before comparing groups across time periods. We calculated the sample size for this based on the desired confidence level (95%) and statistical power (.80 or above). We found that $N=$ 3000 was more than enough to meet the criteria established. Thus, random samples of 3000 tweets were drawn from the anti- and pro-mask datasets for each of the four key time periods, for a total of eight separate datasets. To keep the analysis for the individual hashtags manageable, we focused on the last time period (T4) and created separate datasets for each of the hashtags, for a total of eight separate datasets.

Fourth, we used Communalytic to analyze the toxicity of tweets for each of the 16 datasets separately. Because coding big datasets manually can be costly and laborious, web-based analytic tools such as Communalytic (Gruzd and Mai, 2020) allow researchers to conduct toxicity analyses of social media discourse using Jigsaw and Google's Perspective application programming interface (API). Perspective is an API created by Jigsaw and Google's Counter Abuse Technology team as part of a research initiative called Conversation-AI. According to their website, they "open source experiments, tools, and research data to explore the strengths and weaknesses of ML [machine learning] as a means to combat online toxicity and harassment" (Jigsaw LLC, 2020a). The analysis produces toxicity scores for each post in a dataset, as well as scores for specific message attributes such as *identity attack*, *insult*, and *profanity*, set by the Perspective API (see Jigsaw LLC, 2020b for a definition of all attributes). We found that the measures for *toxicity*, *severe toxicity*, *identity attack*, *insult*, and *profanity* were comparable and included many of the 10 types of verbally aggressive messages identified by Infante et al. (1990). For example, Perspective assigns an *identity attack* score to a post based on how many "negative or hateful comments targeting someone because of their identity" it contains, while a post's *insult* score is based on "insulting, inflammatory, or negative comment towards a person or a group of people." Scores in these measures range from 0 to 1.0, with the latter representing the highest toxicity. We tested our hypotheses through tests of difference using SPSS. We inspected the data for normality and homogeneity of variance before testing for statistical differences in mean scores using the Levene's test. While independent samples *t*-tests and analysis of variance (ANOVA) tests are considered robust against the normality assumption (Blanca et al., 2017), alternative tests are required when the homogeneity of variance assumption cannot be assumed. When normality and homogeneity of variance could not be assumed, we used the Welch–Satterthwaite method (Zimmerman, 2004) to correct for this violation and followed that up with the Games-Howell post hoc test.

Finally, we applied Sandman's model and adapted the thresholds suggested by Gruzd et al. (2020) to create the matrix for interpreting social media messages (Table 1). In this paper, we use the $\geq .70$ threshold to indicate high toxicity and propose $\leq .30$ as the threshold for low toxicity.

Results

Identifying anti- and pro-face mask sentiment hashtags

Figure 1 shows the popularity of the most used face mask-related hashtags on Twitter at different times during 2020 according to the website Hashtagify. The #Mask hashtag in Figure 1 serves as a baseline, as it contains all tweets including #mask. Use of the hashtag #WearAMask, which is mostly used to express support of wearing face masks, is the only one well above the
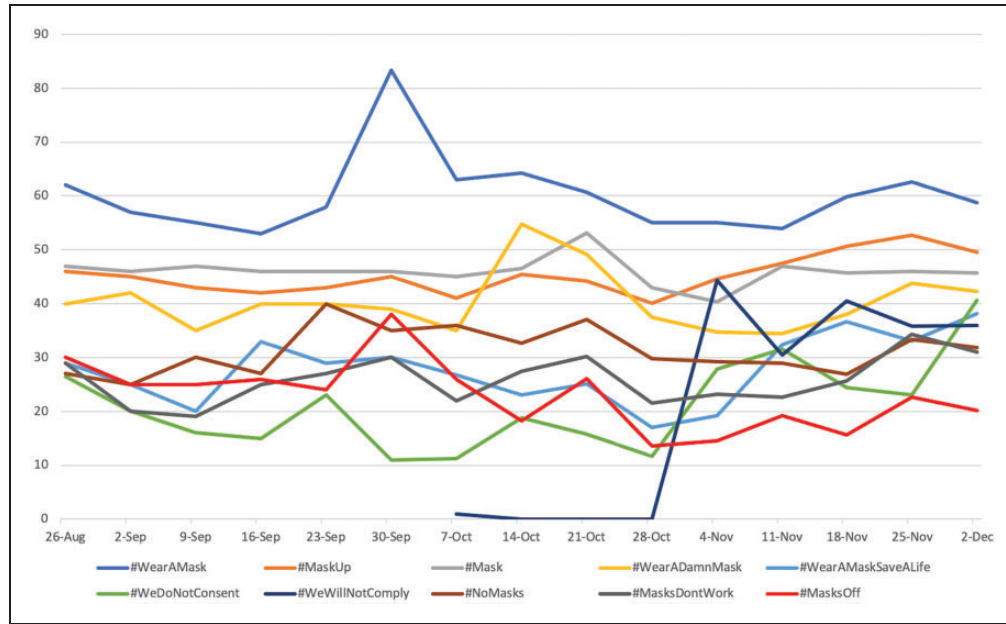
Figure 1. Popularity of face-mask-related hashtags on Twitter from 26 August 2020 to 2 December 2020. The data for this figure was collected using Hashtagify.

baseline. According to the sample data from Hashtagify, #WearAMask was the most popular pro-mask sentiment hashtag used in tweets from late August to early December 2020, while #NoMasks was the most used anti-mask sentiment hashtag during the same time period. Hashtagify estimates that from 27 September to 6 December 2020, #WearAMask was tweeted an average of over 11,000 times per day, by 10,900 average daily users, and had a total of 13.1 billion impressions with an average of 62% level of engagement throughout the time period. On the other hand, Hashtagify estimates that #NoMask was tweeted an average of 431 times per day, by 394 average daily users, and had a total of 151.3 million impressions with an average of 73% level of engagement from 4 October to 6 December 2020. Based on the data from Hashtagify, we identified the following hashtags for our analysis: #MasksDontWork, #MaskOff, #MasksOff, #NoMask, #NoMasks, #WeDoNotConsent, and #WeWillNotComply to identify and categorize anti-mask sentiment tweets, and #MaskUp, #WearADamnMask, #WearAMask, and #WearAMasksavealife to identify and categorize pro-mask sentiment tweets.

Twitter data

Over 4.8 million mask-related tweets were collected from Twitter via Netlytic between 18 July and 2 December 2020. We narrowed down our analysis of tweets to the four key time periods mentioned above:

T1 ($N = 847,090$), T2 ($N = 731,474$), T3 ($N = 305,285$), and T4 ($N = 216,397$). We observed from the tweets collected that pro-mask sentiment tweets significantly outnumbered anti-mask sentiment tweets throughout the four key time periods. For example, for the last data collection during the U.S. Thanksgiving holiday (T4), 34,955 (19%) of the tweets that remained after removing duplicates included one or more of the hashtags of interest in the current study. Of those tweets, only 3144 or about 9% of the data included anti-mask sentiment hashtags.

Figure 2 shows the DrL visualization of the named network of tweets ($N = 34,955$), including anti- and pro-mask hashtags for T4 done using Netlytic. The network is large, decentralized, and loosely connected. The network size is 105 in diameter, and it represents 7487 posters with 28,978 ties, including self-loops. The modularity measure ($N = 0.91$) is high, while density ($N = 0.00$), reciprocity ($N = 0.02$), and centralization ($N = 0.03$) are all low. Approximately 9% of tweets ($N = 3144$) had anti-mask sentiment hashtags; the overwhelming majority of messages included pro-mask hashtags ($N = 31,873$). A small number ($N = 62$) of messages included both anti- and pro-mask sentiment hashtags. The rectangles in Figure 2 denote areas of the network graph where anti-mask sentiment hashtag activity is concentrated. The circle marks the area of the network graph with tweets mentioning former president Donald Trump (@realdonaldtrump), which included a mix of both anti- and pro-mask sentiment hashtags.
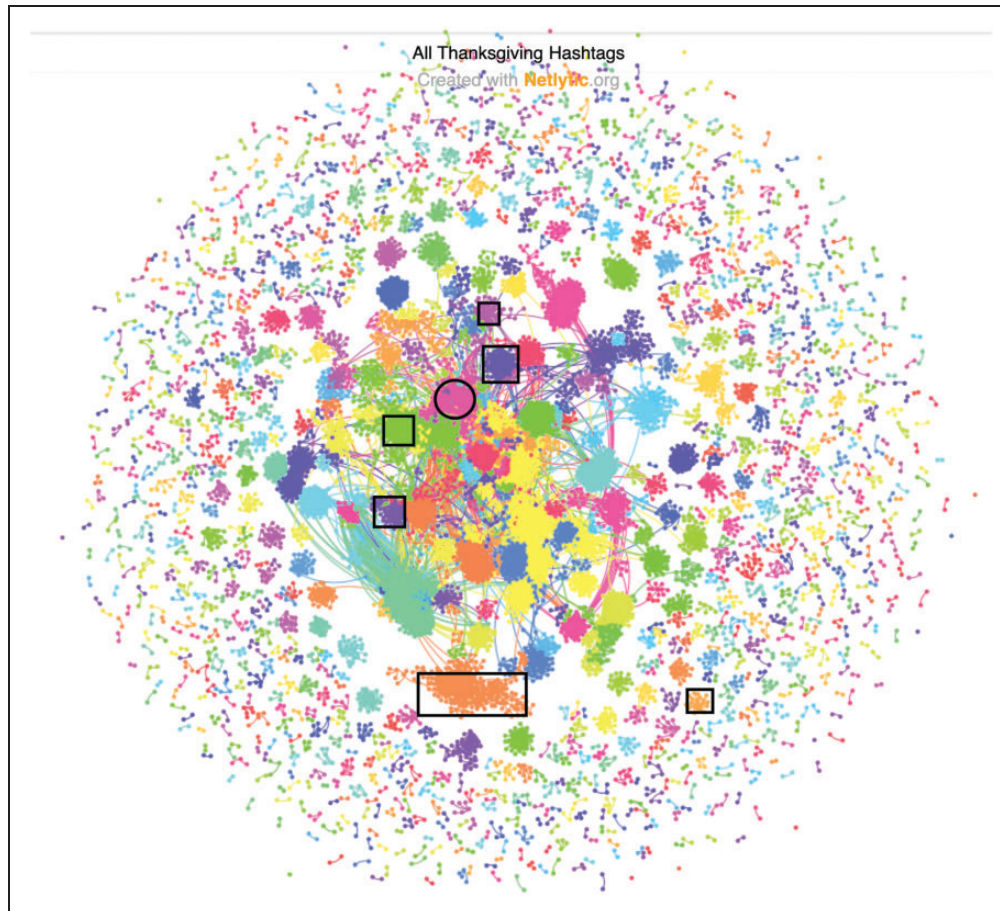
Figure 2. Name network (DrL) for tweets with anti- and pro-mask sentiment hashtags collected from 25 November 2020 to 28 November 2020. This network visualization was done using Netlytic.

Toxicity analysis

To answer RQ1, we started by comparing the mean toxicity scores of anti- and pro-mask sentiment tweets.

Figure 3 is a visual representation of mean toxicity scores for each time period (i.e. T1–T4). Standard deviation (SD) values are shown in parentheses next to each mean toxicity score. As evident in the graph, mean toxicity scores were very different between the anti- and pro-mask hashtag data, as well as for the different time periods studied. The only measure for which scores approximated a normal distribution was the *flirtation* measure. For all other measures, the data was either heavily skewed or followed a well curve or bimodal distribution. The Levene's test for equality of variances was significant at the $p = .00$ level for all comparisons done. Therefore, to test H1, we conducted independent samples *t*-tests using the Welch–Satterthwaite method (Zimmerman, 2004) to correct for this violation. Overall, we found that mean toxicity scores for tweets with anti-face mask hashtags were significantly higher than mean toxicity scores for tweets with pro-face mask hashtags for each of the time periods studied.

Therefore, H1 was supported. Larger differences in mean toxicity scores were observed in the data from the Thanksgiving period (T4), *Welch's t*(4863)= 25.61, $p < .001$, 95% confidence interval (CI) [0.15, 0.18], followed by the data collected in July (T1), *Welch's t*(5160)= 18.09, $p < .001$, 95% CI [0.11, 0.14], while the smallest differences in mean toxicity scores were observed in the period following the first presidential debate (T2), *Welch's t*(5981)= 9.36, $p < .001$, 95% CI [0.04, 0.06]. The differences in mean toxicity scores observed in the data from the week of U.S. 2020 general election (T3) fell in between, *Welch's t*(5929)= 16.95, $p < .001$, 95% CI [0.07, 0.09].

We also found significant differences for the other types of toxicity scores. Figure 4 shows the mean differences for each of the other seven types of toxicity scores reported by Communalytic. As evident in the graph, mean differences in the other types of toxicity scores peaked during the Thanksgiving period (T4), except for mean differences in *threat* and *flirtation* scores, which both peaked during the U.S. general election week. We then followed the thresholds established by Gruzd et al.
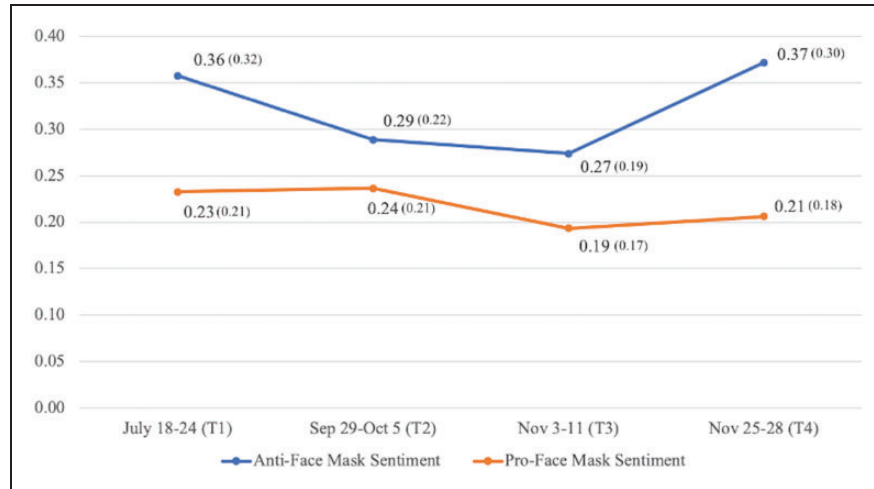
Figure 3. Average toxicity scores for anti- and pro-face mask hashtag tweets. Toxicity scores were generated using Communalytic.
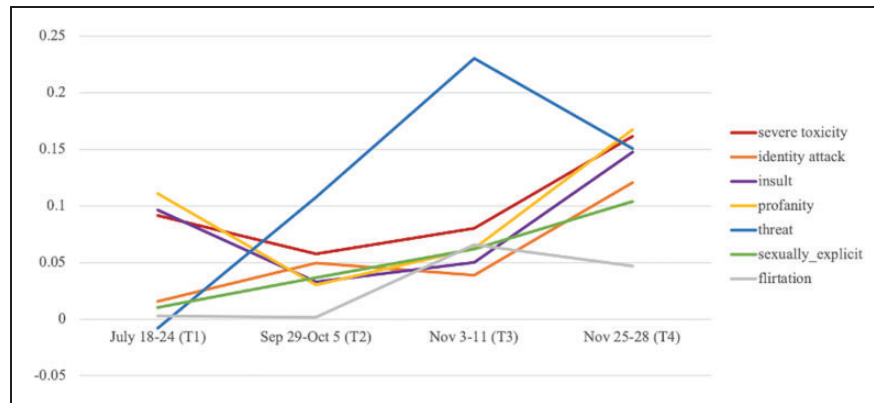

Figure 4. Mean differences in seven types of toxicity scores.

(2020) to calculate the percentage of the data with high toxicity scores.

Figure 5 shows the percentage of tweets in each dataset that scored $\geq .70$ for each type of toxicity at T1 (18–24 July) and T4 (25–28 November). For example, almost 22% of anti-mask tweets at T1 scored $\geq .70$ in the toxicity measure, while less than 6% of pro-mask tweets did during the same time period. The percentage of anti-mask tweets scoring $\geq .70$ for the insult measure *increased* from 5.5% in T1 to 17% in T4, while the percentage of pro-mask tweets scoring $\geq .70$ *decreased* from 4.8% in T1 to 3.4% in T4. Severe toxicity and threat scores also increased significantly among anti-mask tweets from T1 to T4, while profanity scores increased but less markedly. *Threat* and *flirtation* scores both increased among pro-mask tweets from T1 to T4, though again less markedly than the increase in some of the scores observed for the anti-mask tweets.

To answer RQ2, we analyzed the toxicity score for each of the eight hashtags of interest. Table 1 in the

Supplemental Material includes the means and SDs for toxicity scores per hashtag. We found a statistically significant difference in mean toxicity scores for tweets when grouped by anti-mask sentiment and pro-mask sentiment hashtags, *Welch's* $t(9811)=31.34$, $p = .00$, with mean toxicity scores for anti-mask sentiment hashtag being significantly higher. Therefore, H1 was supported. Among anti-face mask sentiment hashtags, #MasksDontWork had the highest mean scores in measures of *toxicity*, *severe toxicity*, *identity attack*, *insult*, *profanity*, and *sexually explicit*, while the combined #WeDoNotConsent/#WeWillNotComply had the highest mean scores in measures of *threat* and *flirtation*. The hashtag #WearAMask had the lowest mean scores for all measures except *flirtation*. Among pro-face mask sentiment hashtags, #WearADamnMask had the highest mean scores for all measures except for *flirtation*. This is not surprising given that the hashtag itself includes a swear word. We inspected the data for normality and homogeneity of variance before
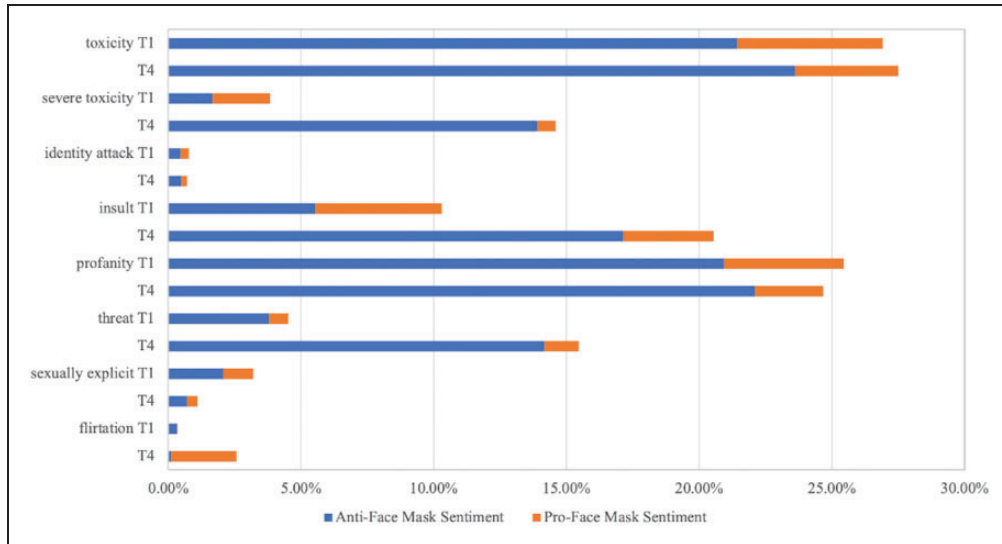
Figure 5. Percentages of messages in dataset with scores $\geq$ .70 for each type of toxicity at T1 and T4.

testing for statistical differences in mean scores. Similar to the first analysis, *flirtation* scores approximated a normal distribution for most hashtags. For most other measures, however, the data was either heavily skewed or followed a well curve or bimodal distribution. The Levene's test for equality of variances was significant at the $p = .00$ level. Since homogeneity of variance could not be assumed, we used the Welch's adjusted F alternative to ANOVA to test our remaining hypotheses and found statistically significant differences in mean scores between tweets when grouped by individual hashtags, Welch's $F(8, 3534)=437.81$, $p = .00$. Post hoc comparisons were conducted using the Games-Howell post hoc procedure to determine which hashtags differed significantly (see Table 2 in the Supplemental Material).

In terms of H2, we found a statistically significant difference in mean toxicity scores between pro-mask sentiment tweets when compared by hashtags, Welch's $F(3, 3618)=767.81$, $p = .00$. The mean toxicity score for tweets containing #WearADamnMask was significantly higher than the mean toxicity scores for all other pro-mask sentiment hashtags tweets, with the greatest mean difference occurring between #WearADamnMask and #WearAMask ($MD = 0.17$, $p = .00$, 95% CI [0.16, 0.18]. H2 was supported in that we expected that the mere inclusion of the word "damn" in the hashtag would yield higher toxicity scores. However, when we re-ran the analysis replacing the word "damn" in the hashtag for "A12," the adapted hashtag yielded significantly lower mean scores (see results for #WearA12Mask in Table 1 of the Supplemental Material). When we excluded the word "damn" from the hashtag, tweets containing

#MaskUp had the highest mean toxicity score among pro-mask sentiment hashtags for all toxicity measures except *threat*, for which tweets containing #WearAMaskSaveALife had the highest mean score. In other words, while we were correct in expecting the word "damn" in the hashtag to drive toxicity scores up for tweets using #WearADamnMask, we did not find that the language of the rest of the tweets beyond the hashtag was significantly more toxic than the language used in tweets with other pro-mask sentiment hashtags. We discuss the implications of this finding in the limitations section.

We also found a statistically significant difference in mean toxicity scores between anti-mask sentiment tweets when compared by hashtags, Welch's $F(3, 3618)=767.81$, $p = .00$. However, in terms of H3, we found that while there were some statistically significant differences between tweets containing #WeDoNotConsent/#WeDoNotComply and some (not all) of the other tweets containing anti-mask sentiment hashtags, the direction was not what we expected; in fact, almost the opposite was true. The mean toxicity score for tweets containing #WeDoNotConsent/#WeDoNotComply was significantly lower than those containing #MasksDontWork and #MaskOff and not significantly different from those containing #NoMask, which received the lowest mean toxicity score. Tweets containing the hashtag #MasksDontWork had the highest mean toxicity score among anti-mask sentiment tweets. Therefore, H3 was not supported.

In terms of H4, we found a statistically significant difference between mean toxicity scores of tweets containing #WeDoNotConsent/#WeWillNotComply and

Table 2. Summary table of hypotheses and results.

Hypothesis	Result
H1: There would be a statistically significant difference in mean toxicity scores between tweets with anti- and pro-mask sentiment hashtags and pro-mask sentiment hashtags, with the former having a higher mean toxicity score.	Supported
H2: There would be a statistically significant difference in mean toxicity scores between tweets using #WearADamnMask and other pro-sentiment tweets, with the former having a higher mean toxicity score.	Supported
H3: There would be a statistically significant difference in mean toxicity scores between tweets using #WeDoNotConsent/#WeWillNotComply and other anti-mask sentiment tweets, with the former having a higher mean toxicity score.	Not supported
H4: There would be no statistically significant difference in mean toxicity scores between tweets using #WeDoNotConsent/#WeWillNotComply and #WearADamnMask.	Not supported
H5: There would be a statistically significant difference in mean toxicity scores between tweets using #NoMask and #WearAMask, with the former having a higher mean toxicity score.	Supported

those containing #WearADamnMask Welch's t (1821)=10.97, $p < .0001$. The mean toxicity score for #WearADamnMask was significantly higher than the mean toxicity score for #WeDoNotConsent/ #WeWillNotComply. Therefore, H4 was not supported. However, when we re-ran the analysis like we did for H2, replacing "damn" from the hashtag with "A12," tweets with #WeDoNotConsent/ #WeWillNotComply ($M = 0.29$) had a significantly higher mean toxicity score than those with #WearA12Mask ($M = 0.26$), Welch's t(2054)=3.97, $p < .0001$. We address the implications of this finding in the limitations section.

In terms of H5, we found a statistically significant difference in mean toxicity scores between tweets using the #NoMask and those using the #WearAMask hashtags, Welch's t(358)=6.05, $p < .0001$, with the mean toxicity score for #NoMask being significantly higher. Thus, H5 is supported. Table 2 summarizes our study's hypotheses and results.

Finally, we looked at tweets from T4 to see how they mapped on the matrix we created linking Sandman's framework to the toxicity thresholds. Looking at the random sample of tweets with anti-mask sentiment hashtags for T4 ($N = 3000$), 24% scored $\geq.70$ in toxicity, while 60% scored $\leq .30$ in toxicity. In comparison, looking at the random sample of tweets with pro-mask sentiment hashtags ($N = 3000$), less than 4% of tweets scored $\geq.70$ in toxicity, while 79% scored $\leq .30$ in toxicity. Table 3 shows examples of tweets from the datasets that help illustrate the concepts of outrage and toxicity, with the four different combinations adapted from Sandman's framework.

The high hazard/high outrage example is a #WearADamnMask tweet that emphasizes COVID-19 as life-threatening and has a high toxicity score. That tweet was accompanied by a visual showing two

images side by side—a mother and child wearing a mask on the left and a man on a respirator on the right—with the text "Which mask do you prefer? #WearADamnMask" superimposed on the images. We believe the images, if analyzed, would score high in toxicity. However, the Perspective API does not analyze the toxicity of images or multimedia content in a tweet.

Discussion

The findings from our toxicity analysis map well onto Sandman's conceptualization of risk perception as the sum of hazard plus outrage. As reflected in our results, social media—in this case, Twitter—is a readily quantifiable vehicle for examining communication of outrage-based sentiments in anti- and pro-mask contexts. A key finding of this research is that, overall, anti-mask tweets are more toxic than pro-mask tweets for all hashtags studied in this paper, except #WearADamnMask. Consistent with the matrix for interpreting toxicity scores (see Tables 1 and 3), many anti-mask tweets are reflective of low perceived risk of COVID-19 and high outrage. This appears inconsistent given that if one didn't perceive risk, then there would be no need to express outrage in the form of toxic tweets. Therefore, we think something else is at work here that is reflective of characteristics of anti-maskers. We offer two explanations.

First, we note that anti-maskers had low ratings of risk and higher ratings of toxicity. It may well be that because they believed risks of getting COVID-19 were low, they also believed that mask wearing was unnecessary. This could have led to the belief that being forced to do so was an infringement into their rights, something to which they reacted with a toxic response precisely *because* they deemed the mandates to be

Table 3. Examples of tweets to illustrate the different combinations of hazard and outrage, with toxicity scores in parentheses.

	Low outrage	High outrage
Low hazard	Happy Thanksgiving! Big Thanksgiving feast today. The more the merrier! No masks allowed! #HappyThanksgiving #Thanksgiving2020 #Trump2020 #Trump2020Landslide #StopTheCheating #StopTheSteal #StopTheSteal2020 #NoMasks #nomandates (toxicity score = 0.18)	████████ you need to stop with these ridiculous mask commercials!! All for a virus with a 99.9% survival rate. This non sense has gone too far and y'all need to STOP encouraging this type of behavior with the stupid masks. #Takeoffyourmask #MasksDontWork (toxicity score = 0.92)
High hazard	Want a sobering reality? About 16000 Americans who just celebrated #Thanksgiving will NOT LIVE TO SEE CHRISTMAS.......#covid19 #WearAMaskSaveALife (toxicity score = 0.12)	Christmas has come early, people. There is no room in the inn. Don't expect to get a bed after exercising your damn "right" to "freedom" n turkey. You'll be chillin on ice for days to get in the morgue. Too morbid for ya? Then #WearADamnMask n stop acting like a whiny lil bitch ‼️ (toxicity score = 0.85)

unnecessary and unwarranted given their perception of low risk. This is something that we expected to find given the literature's suggestion that those espousing minority views tend to do so louder and prouder (Miller and Morrison, 2009: 741).

The second explanation centers around in-group/ out-group membership. It may well be that communicating through toxic tweets is a requirement of group membership as well as a marker of personal identification, the goal of which is to entertain, please, or upset others. In other words, toxicity in this sense is more likely related to issues of personal freedom and identity and group membership discussed earlier in this paper. The overall toxicity among the anti-mask group, the research found, increased at peak periods, like Thanksgiving, when the issue of mask wearing was more salient because of impending visits with friends and family. The data suggests that over time there is a need among anti-maskers to reinforce their beliefs, with increasing strength (toxicity) regarding anti-mask behavior. Conversely, it might be that they really don't believe in anti-mask behavior but feel the need to support the political/ideological position of the group with whom they identify.

A specific example from the literature that provides an explanation of the phenomenon under study is the false-consensus effect (Marks and Miller, 1987; Ross et al., 1977) in which people believe their deviant behavior is more widely supported than it actually is. For example, those in the minority, like those associated with the anti-mask group, tend to be more vocal about their position on the issue and the outcome is one of "greater comfort and pride in expressing their opinions" (Miller and Morrison, 2009: 741). Applying these findings, we believe that those individuals who

take anti-mask-wearing positions on social media platforms, while quite vocal, are actually in the minority, at least based on the number of tweets associated with anti-mask hashtags. As a result, they are more likely to express their opinions in a more forceful manner than those individuals posting pro-mask-wearing hashtags through the use of toxic language that may in the view of the individual bring primacy to her or his position regarding mask wearing. While our study does not look into personality, the study by Twenge and Campbell (2003) concludes that narcissism paired with social rejection serve to predict aggressive behavior. Future research may be able to establish a connection between an individual's personality, tweets, and toxicity online.

We were surprised to find out that #MasksDontWork had the highest mean toxicity score among all the anti-mask sentiment hashtags. We expected tweets using this hashtag to carry most of the mis- and dis-information. In other words, we expected tweets with this hashtag to be more informational and appeal more to reason showing facts and/or evidence that would contradict claims regarding the effectiveness of masks as a way to plant doubt in readers. Instead, what we found was that, while some of the tweets we analyzed did in fact include images, links, and information from research studies discrediting the effectiveness of masks, most references were informal, as prima facie, without providing scientific evidence to support the claim that #MasksDontWork. This is an important finding for the study of infodemic because it shows that, at least in the context of face mask wearing as a preventive measure to combat COVID-19, users on social media seem to be less attracted to posts citing scientific information and

more driven to posts that emphasize emotion, anecdotes, and those that reflect on their own personal experiences.

As people communicate in a heated manner on social media and mainstream media echoes the polarization on- and offline, the extremes that emerge have the potential to become more toxic to both emphasize a particular position regarding the efficacy or lack thereof of mask wearing and to reinforce the communicator's beliefs. At the same time, toxicity online has a spillover effect on society at large. What is happening as evidenced by the tweets in this study is planting seeds of uncertainty, and such uncertainty is important as it extends the issues related to the infodemic to go beyond concern for misinformation or disinformation to consider the role of doubt. In her essay, "Did Media Literacy Backfire," Danah Boyd (2017) argues that doubt is a useful tool, because doubt isn't based on information but based on emotions, and one can't really argue with emotions. As there is no middle ground on the issue, doubt emerges as an internal response as a result of which the individual either moves (in their beliefs) toward one of the positions in the binary or they exist in no-person's land—a worldview based on doubt.

Therefore, from a public health perspective, the question is not only about correcting or misinformation or countering disinformation as an antidote to the infodemic. Health communicators need to deal with the emotions associated with the issue to reduce doubt and uncertainty. Sandman's suggestion to mitigate outrage around the issue of face masks is to use bandwagon strategies to build empathy among those reluctant to wear them, as opposed to finger-wagging (Sandman, 2020). In regard to countering doubt, Boyd suggests changing "how we make sense of information." The point here is not about changing information but about the role that media—both mainstream and social media—play as a sense-making mechanism. Instead, we need to examine how people utilize social media to work through issues and make sense of their world, or at least try to; otherwise, they are left to grapple with doubt.

Limitations

There are several limitations to our study. Our data collection is only a fraction of the Twitter universe, which is about 500 million tweets per day, or about 6000 tweets every second (Internet Live Stats, 2020). Each instance of data extraction from Twitter using Netlytic is capped at 1000 tweets at a time. While we used sampling techniques that would help reduce bias, the reality is that the universe of data is so much larger than what we analyzed that we would caution against making strict generalizations from our results. Having said that, we were able to collect enough data across time to observe repeated toxicity patterns between anti- and pro-mask sentiment hashtag tweets that we believe to be meaningful. Also, our data comes from one platform, Twitter, which has been accused of having a liberal bias and censoring conservative voices (Guynn, 2016), leading many pro-Trump supporters to migrate to other platforms such as Parler, Gab, Rumble, and MeWe (Bomey, 2020).

Another limitation in our research has to do with machine learning itself and the use of APIs to analyze Big Data. Some studies have noted inaccuracies in the detection of anti-social sentiment by the Perspective API (Awal et al., 2020; Gruzd et al., 2020). Awal et al. developed a basic and an extended lexicon set to identify anti-social sentiment in 40 million COVID-related tweets taken from publicly available datasets. They then cross-validated their results with results from the Perspective API. They found that false positives occurred when using both annotation methods and in some cases contradicted each other. While manually coding large datasets can be costly, laborious, and, depending on the size of the dataset (such as 40 million), nearly impossible, their study does punctuate the limitations of current automated methods focused on singling out keywords as a measure of toxicity and anti-social behavior. Also, the difference that we found when we re-ran the analyses after removing the word "damn" from the #WearADamnMask hashtag portion of the tweets underscores the impact that a single word can have on toxicity scores assigned by the API and how removing just one word can lead to significantly different results. We believe this is an important consideration for researchers in this area to address the relationship between the toxicity scores of the hashtag(s) used and the rest of the tweet or text of the content posted. Furthermore, as Gruzd et al. (2020) noted, automation is not yet at a point where it can interpret words in context. While the keyword may be present, it may be used in a sarcastic or other nuanced way that the machine algorithm cannot detect. Moreover, as we saw in the example for the high hazard/high outrage condition, social media messages many times include multimedia content that carries a lot of valuable information that is lost on keyword/lexicon-based content analysis.

Finally, in this study, we refer to toxicity as a measure of generalized outrage without taking into consideration different types of outrage that may exist directed at different actors and aspects of an issue. Taking a more nuanced view of outrage into consideration would have had an impact on the kind of hypotheses we tested, how we collected data, how we conducted our analysis, and the results of our study.

For example, Peter Sandman refers to two different types of outrage on social media: acute outrage versus chronic outrage (Sandman, 2015), and to at least 12 different components of outrage (Sandman, 1991). Future research in this area looking at toxicity and outrage would benefit from taking into consideration different taxonomies of outrage and include them in the study design and data collection.

Conclusion

In this paper, we propose that toxicity as a form of verbal aggression and as an expression of outrage creates an additional and powerful barrier for individuals exposed to it on social media and is just as dangerous as mis- and dis-information. Furthermore, we believe that toxic discourse online has a spillover effect offline that threatens the successful promotion of COVID-19-related risk communication. Specifically, toxicity of discourse around face mask wearing: (a) indicates the presence of public outrage, which in turn threatens the success of risk communication efforts around an issue; (b) alienates individuals from the discussion, leading to an increase in public disengagement with the issue; and (c) breeds doubt, which effectively erodes trust in public health authorities and guidance. If toxicity results in people leaving a discussion, then toxic discourse could create large numbers of people who do not want to engage with a particular topic. This could present a hurdle for public health and risk communicators who are competing with other sources of information for the attention of audiences, who at the same time might grow more disappointed by the level of discourse, and thus disengaged. Despite these challenges, our findings nonetheless point to the importance of public health agencies' monitoring toxic communication on social media during pandemics in order to inform data-driven adjustments as needed to these agencies' risk communication efforts.

ORCID iDs

Paola Pascual-Ferrá https://orcid.org/0000-0001-8511-6798
Neil Alperstein https://orcid.org/0000-0003-3906-3474
Daniel J Barnett https://orcid.org/0000-0003-1382-6197
Rajiv N Rimal https://orcid.org/0000-0003-1413-9305

Supplemental Material

Supplementary material for this article is available online.

References

Allport FH (1924) *Social Psychology*. Boston: Houghton Mifflin.

Almerekhi H, Kwak H, Jansen BJ, et al. (2019) Detecting toxicity triggers in online discussions. In: *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, Hof University, Germany, 17–20 September 2019, pp.291–292. New York: Association for Computing Machinery.

American Medical Association (2020) *AMA, AHA, ANA: Open letter on mask use, stopping spread of COVID-19*. Available at: www.ama-assn.org/delivering-care/public-health/ama-aha-ana-open-letter-mask-use-stopping-spread-covid-19 (accessed 10 December 2020).

American Public Health Association (2020) *Stop the spread and show you care: Wear a mask to protect others from COVID-19* [press release] 30 July. Available at: www.apha.org/news-and-media/news-releases/apha-news-releases/2020/wear-a-mask (accessed 10 December 2020).

Aratani L (2020) How did face masks become a political issue in America? *The Guardian*, 29 June. Available at: www.theguardian.com/world/2020/jun/29/face-masks-us-politics-coronavirus (accessed 3 May 2021).

Awal MR, Cao R, Mitrovic S, et al. (2020) On analyzing anti-social behaviors amid COVID-19 pandemic. Available at: https://arxiv.org/abs/2007.10712 (accessed 10 December 2020).

Bennett JT (2020) Trump mocks Biden for wearing a face mask: 'What was the plastic surgery for?' *The Independent*, 23 September. Available at: www.independent.co.uk/news/world/americas/us-election/url-trump-biden-covid-masks-plastic-surgery-b540606.html (accessed 10 December 2020).

Biggers L (2020) Wear a face mask to protect each other. *Duke Health Blog*. Available at: www.dukehealth.org/blog/wear-face-mask-protect-each-other (accessed 10 December 2020).

Blanca MJ, Alarcón R, Arnau J, et al. (2017) Non-normal data: Is ANOVA still a valid option? *Psicothema* 29(4): 552–557.

Bomey N (2020) Parler, MeWe, Gab gain momentum as conservative social media alternatives in post-Trump age. *USA Today*, 11 November. Available at: www.usatoday.com/story/tech/2020/11/11/parler-mewe-gab-social-media-trump-election-facebook-twitter/6232351002/ (accessed 10 December 2020).

boyd d (2017) Did media literacy backfire? In: *Data & Society: Points* [blog]. Available at: https://points.datasociety.net/did-media-literacy-backfire-7418c084d88d (accessed 10 December 2020).

Brenan M (2020) Americans' face mask usage varies greatly by demographics. *Gallup*, 13 July. Available at: https://news.gallup.com/poll/315590/americans-face-mask-usage-varies-greatly-demographics.aspx (accessed 10 December 2020).

Carrega C (2020) Driver in crash may have passed out from wearing N95 mask too long: Police. *ABC News*, 25 April. Available at: https://abcnews.go.com/US/driver-crash-passed-wearing-n95-mask-long-police/story?id = 70346532 (accessed 3 May 2021).

Centers for Disease Control and Prevention (2020) *CDC Calls on Americans to Wear Masks to Prevent COVID-19 Spread* [press release] 14 July. Available at: www.cdc.gov/media/releases/2020/p0714-americans-to-wear-masks.html (accessed 10 (December 2020).

Cramer M and Sheikh K (2020) Surgeon general urges the public to stop buying face masks. *The New York Times*, 29 February. Available at: www.nytimes.com/2020/02/29/health/coronavirus-n95-face-masks.html (accessed 10 December 2020).

Cuan-Baltazar JY, et al. (2020) Misinformation of COVID-19 on the Internet: Infodemiology study. *JMIR Public Health and Surveillance* 6(2): e18444.

CyBranding Ltd. (2020) Hashtagify find guide. Available at: https://hashtagify.me/orientation/guide_find (accessed 10 December 2020).

Fan Y (2020) The social amplification of risk on Weibo-Take the COVID-19 epidemic in China as an example. In: *Proceedings of the 2020 Conference on Artificial Intelligence and Healthcare*, Taiyuan, China, 23 October 2020, pp.91–95. New York: Association for Computing Machinery.

Fields JM and Schuman H (1976) Public beliefs about the beliefs of the public. *Public Opinion Quarterly* 40(4): 427–448.

Gillespie C (2020) Does wearing a face mask reduce oxygen—and can it increase CO_2 levels? Here's what experts say. *Health*, 13 May. Available at: www.health.com/condition/infectious-diseases/coronavirus/does-wearing-face-mask-increase-co2-levels (accessed 10 December 2020).

Gruzd A (2016) Netlytic: Software for automated text and social network analysis. Available at: http://Netlytic.org (accessed 10 December 2020).

Gruzd A, Mai P and Vahedi Z (2020) Studying anti-social behaviour on Reddit with Communalytic [pre-print]. Available at: https://advance.sagepub.com/articles/Studying_Anti-Social_Behaviour_on_Reddit_with_Communalytic/12453749 (accessed 10 December 2020).

Gruzd A and Mai P (2020) Communalytic: An online community analyzer for studying anti-social behaviour. Available at: https://Communalytic.com (accessed 10 December 2020).

Guberman J, Schmitz C and Hemphill L (2016) Quantifying toxicity and verbal violence on Twitter. In: *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, San Francisco, CA, 27 February 2016, pp.277–280. New York: Association for Computer Machinery.

Guerrero-Solé F and Philippe O (2020) La toxicidad de la política española en twitter durante la pandemia de la COVID-19. *Hipertextnet* 21: 133–139.

Guynn J (2016) Twitter suspends alt-right accounts. *USA Today*, 16 November. Available at: www.usatoday.com/story/tech/news/2016/11/15/twitter-suspends-alt-right-accounts/93943194/ (accessed 10 December 2020).

Hamilton MA (2012) Verbal aggression: Understanding the psychological antecedents and social consequences. *Journal of Language and Social Psychology* 31(1): 5–12.

Hamilton MA, Buck R, Chory RM, et al. (2008) Individualistic and cooperative affect systems as determinants of aggressive or collaborative message choice. In: Beatty MJ, McCroskey JC and Floyd K (eds) *Advances in Communication Research: Biological Perspectives*. Cresskill, NJ: Hampton Press, pp. 227–250.

Hernández-García I and Giménez-Júlvez T (2020) Assessment of health information about COVID-19 prevention on the Internet: Infodemiological study. *JMIR Public Health and Surveillance* 6(2): e18717.

Hill C (2020a) Trump mocks Biden for wearing masks: 'Every time you see him, he's got a mask'. *Yahoo! News*, 29 September. Available at: https://news.yahoo.com/trump-mocks-biden-for-wearing-masks-every-time-you-see-him-hes-got-a-mask-022635582.html (accessed 10 December 2020).

Hill DG (2020b) From the Frontlines: The truth about masks and COVID-19. In: *Each Breath* [blog of the American Lung Association]. Available at: www.lung.org/blog/covid-masks (accessed 10 December 2020).

Infante DA, Sabourin TC, Rudd JE, et al. (1990) Verbal aggression in violent and nonviolent marital disputes. *Communication Quarterly* 38(4): 361–371.

Infante DA and Wigley IIC (1986) Verbal aggressiveness: An interpersonal model and measure. *Communication Monographs* 53(1): 61–69.

Internet Live Stats (2020) Twitter usage statistics. Available at: www.internetlivestats.com/twitter-statistics/ (accessed 10 December 2020)

Jigsaw LLC (2020a) Perspective API FAQs: What is perspective? Available at: https://support.perspectiveapi.com/s/about-the-api-faqs (accessed 10 December 2020).

Jigsaw LLC (2020b) Perspective API: Attributes & languages. Available at: https://support.perspectiveapi.com/s/about-the-api-attributes-and-languages (accessed 10 December 2020).

Khan A (2020) 'Wearing a mask can give a false sense of security': What the experts say about wearing face masks. *Yahoo! News*, 30 March. Available at: https://news.yahoo.com/facemask-proper-use-necessary-221717984.html (accessed 10 December 2020).

Kramer S (2020) More Americans say they are regularly wearing masks in stores and other businesses. *Pew Research Center Fact Tank*, 27 August. Available at: https://www.pewresearch.org/fact-tank/2020/08/27/more-americans-say-they-are-regularly-wearing-masks-in-stores-and-other-businesses/ (accessed 10 December 2020).

Ledford C and Anderson L (2020) Communication strategies for family physicians practicing throughout emerging public health crises. *Family Medicine* 52(5): 48–50.

Majó-Vázquez S, Nielsen R, Verdú J, et al. (2020) *Volume and patterns of toxicity in social media conversations during the COVID-19 pandemic*. Reuters Institute. Available at: https://reutersinstitute.politics.ox.ac.uk/volume-and-pat terns-toxicity-social-media-conversations-during-covid-19-pandemic (accessed 10 December 2020).

Malecki K, Keating JA and Safdar N (2021) Crisis communication and public perception of COVID-19 risk in the era of social media. *Clinical Infectious Diseases* 72(4): 697–702.

Mall R, Nagpal M, Salminen J, et al. (2020) Four types of toxic people: Characterizing online users' toxicity over time. In: *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences*, Tallin, Estonia, 25 October 2020, pp.1–11. New York: Association for Computing Machinery.

Marks G and Miller N (1987) Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin* 102(1): 72–90.

Martinelli L, Kopilaš V, Vidmar M, et al. (2021) Face masks during the COVID-19 pandemic: A simple protection tool with many meanings. *Frontiers in Public Health* 8: 606635.

Meyersohn N (2020) Americans are panic shopping for hand sanitizer and face masks. *CNN Business*, 3 March. Available at: www.cnn.com/2020/03/03/business/panic-buying-coronavirus-retail-shopping/index.html (accessed 10 December 2020).

Miller DT and Morrison KR (2009) Expressing deviant opinions: Believing you are in the majority helps. *Journal of Experimental Social Psychology* 45(4): 740–747.

Mogelson L (2020) The militias against masks. *The New Yorker*, 17 August. Available at: www.newyorker.com/magazine/2020/08/24/the-militias-against-masks (accessed 10 December 2020).

Nazaryan A (2020) CDC warns: Save respirator masks for health care workers. *Yahoo! News*, 28 February. Available at: https://news.yahoo.com/cdc-warns-save-respirator-masks-for-healthcare-workers-162144855.html (accessed 10 December 2020).

Okada R (2010) The relationship between vulnerable narcissism and aggression in Japanese undergraduate students. *Personality and Individual Differences* 49(2): 113–118.

Pandemic Action Network (2020) *World Mask Week Aims to Inspire Global Movement to Wear Face Coverings in Public to Help Stem Exponential Spread of COVID-19* [press release] 6 August. Available at: www.prnewswire.com/news-releases/world-mask-week-aims-to-inspire-global-movement-to-wear-face-coverings-in-public-to-help-stem-exponential-spread-of-covid-19-301107357.html (accessed 10 December 2020)

Perkins HW, Haines MP and Rice R (2005) Misperceiving the college drinking norm and related problems: A nationwide study of exposure to prevention information, perceived norms and student alcohol misuse. *Journal of Studies on Alcohol* 66(4): 470–478.

Pfattheicher S, Nockur L, Böhm R, et al. (2020) The emotional path to action: Empathy promotes physical distancing and wearing of face masks during the COVID-19 pandemic. *Psychological Science* 31(11): 1363–1373.

Prentice DA and Miller DT (1996) Pluralistic ignorance and the perpetuation of social norms by unwitting actors. *Advances in Experimental Social Psychology* 28: 161–209.

Reinhart RJ (2020) U.S. face mask usage relatively uncommon in outdoor settings. *Gallup*, 7 August. Available at: https://news.gallup.com/poll/316928/face-mask-usage-rela tively-uncommon-outdoor-settings.aspx (accessed 10 December 2020).

Rogers R (2018) Digital traces in context| Otherwise engaged: Social media from vanity metrics to critical analytics. *International Journal of Communication* 12: 450–472. Available at: https://ijoc.org/index.php/ijoc/arti cle/view/6407 (accessed 21 January 2021).

Ross L, Greene D and House P (1977) The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology* 13(3): 279–301.

Salminen J, Sengün S, Corporan J, et al. (2020) Topic-driven toxicity: Exploring the relationship between online toxicity and news topics. *PLoS One* 15(2): e0228723.

Sandman PM (1991) *Twelve Principal Outrage Components* (handout). Available at: https://psandman.com/handouts/sand58.pdf (accessed 21 January 2021).

Sandman PM (1993) *Responding to Community Outrage: Strategies for Effective Risk Communication*. Falls Church, VA: American Industrial Hygiene Association.

Sandman PM (2015) 10 things you need to know about outrage management and social media. Available at: www.psandman.com/col/social-media.htm (accessed 21 January 2021).

Sandman PM (2020) Pro-mask advocacy when even COVID-19 is polarized. Available at: www.psand man.com/articles/Corona30.htm (accessed 10 December 2020).

Schild L, Ling C, Blackburn J, et al. (2020) "go eat a bat, chang!": An early look on the emergence of Sinophobic behavior on web communities in the face of COVID-19. Available at: https://arxiv.org/abs/2004.04046 (accessed 10 December 2020).

Sousa-Pinto B, Anto A, Czarlewski W, et al. (2020) Assessment of the impact of media coverage on COVID-19–related google trends data: Infodemiology study. *Journal of Medical Internet Research* 22(8): e19611.

Stewart E (2020) Anti-maskers explain themselves: "If I'm going to get Covid and die from it, then so be it." *Vox*, 7 August. Available at: www.vox.com/the-goods/2020/8/7/21357400/anti-mask-protest-rallies-donald-trump-covid-19 (accessed 10 December 2020).

Twenge JM and Campbell WK (2003) "Isn't it fun to get the respect that we're going to deserve?" Narcissism, social

rejection, and aggression. *Personality & Social Psychology Bulletin* 29(2): 261–272.

World Health Organization (2020a) Coronavirus disease (COVID-19) advice for the public: When and how to use masks. Available at: www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks (accessed 10 December 2020).

World Health Organization (2020b) 1st WHO infodemiology conference. Available at: www.who.int/news-room/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference (accessed 10 December 2020).

Zimmerman DW (2004) A note on preliminary tests of equality of variances. *The British Journal of Mathematical and Statistical Psychology* 57: 173–181.