Bystander Intervention on Social Media:

Examining Cyberbullying and Reactions to Systemic Racism

By Rashawn Ray, Melissa Brown, Ed Summers, Samantha Elizondo, and Connor Powelson

B | Governance Studies
at BROOKINGS

Table of Contents

Executive Summary

Over four billion people worldwide are estimated to use social media by 2025. Though a majority of people use social media to engage with family and friends, people also use platforms and apps to obtain news and engage with communities on a range of issues. The polarization and sharing of news content in an era of "alternative facts" and misinformation exacerbates potential conflicts online and can reinforce false rhetoric about specific social issues and racial groups. As a result, social media provides a forum for hate speech and cyberbullying to flourish with limited understanding about tools or tactics to counter these attacks. Consequently, about 70 percent of people report doing something abusive to someone online, a majority of whom report being cyberbullied themselves. Even more troubling, nearly 90 percent of teenagers report witnessing bullying online.

While false rhetoric, hate speech, and cyberbullying have many deleterious effects, there is a silver lining: over 80 percent of youth report seeing others stand up during cyberbullying incidents and engage in bystander intervention online. This high percentage shows the power of bystander intervention—a strategy that has proven effective in ensuring the dissemination of more fact-based public health information—and holds much promise for addressing and curbing online interactions that reinforce systemic racism. Even more promising, a majority of youth report wanting to identify effective strategies to intervene in cyberbullying situations.

While existing studies mostly focus on cyberbullying related to gender or LGBTQIA issues, research focused on hate speech and cyberbullying related to race and racism is less examined. Racism continues to be one of the most polarizing topics in America. Social media polarization has helped re-open the Pandora's box that allows white supremacy and racism to wreak havoc on people's lives. As previous research has shown, reactions to the #BlackLivesMatter movement have created echo chambers on social media that enhance hate speech and cyberbullying related to race and racism.[1] Social media offers the opportunity for people to mask their identity, similar to the KKK hoods of the past.

This study aims to identify effective strategies to combat hate speech and misinformation online. By examining how people respond to cyberbullying, our goal is to highlight bystander intervention strategies that are effective at constructing healthy communication, calming anger and frustration, and changing attitudes. This research has broader implications for leveraging strategies, tools, and tactics, many of which have already helped address the spread of public health misinformation, and

for the development and implementation of positive coping strategies for better mental and emotional health outcomes among marginalized communities.[2]

Accordingly, we conceptualize effective bystander strategies as those that:

• are perceived by other social media users as favorable;

• alter the discussion in more positive, objective, and less antagonistic ways; and

• change the online behavior of the bully or agitator.

Through this effort, the team aims to answer the following questions: How do people combat misinformation online, particularly related to systemic racism, and, more specifically, how do people engage in bystander intervention on social media? What strategies do they use and how effective are people at changing attitudes? How do people encourage healthy coping strategies for better mental and emotional health outcomes?

Analyzing over two million tweets and posts scraped from Twitter and Reddit from 2020, we examined the effectiveness of bystander strategies used online to combat racism.

Analyzing over two million tweets and posts scraped from Twitter and Reddit from 2020, we examined the effectiveness of bystander strategies used online to combat racism. These social media platforms were specifically chosen because they have inherent ranking systems that allow us to examine which strategies are viewed as most effective. On Twitter, people like and retweet messages. On Reddit, people rank comments that move them up or down in the hierarchy queue of importance to be seen more by others. Both platforms are also open allowing most people to comment on most tweets or posts.

Methodologically, we conducted a quantitative analysis of tweets and posts and a content analysis of comments. The analysis focused on four domains related to addressing racism (systemic racism, police brutality, education inequality, and employment and wealth) by using synonyms to each term to search hashtags on Twitter and search posts on Reddit that use these terms.

We found four primary types of racist discourses: stereotyping, scapegoating, accusations of reverse racism, and echo chambers. We also found four types of bystander intervention strategies, which include: call-outs, insults or mocking, attempts to educate or provide evidence, and content moderation. However, only one in six Twitter discussions and slightly less than 40 percent of Reddit discussions featured bystander action. Our findings contribute to research identifying and disseminating results about patterns regarding online communication and effective strategies to combat hate speech and misinformation about systemic racism.

In the sections below, we provide an overview of academic research on cyberbullying and social support, a detailed methods section regarding our analytical approach, and quantitative and qualitative findings from our investigation into how discussions about systemic racism manifest on social media.

Background

Cyberbullying

With 95 percent of American teens having access to a smartphone, adolescents are more connected than ever.[3] Many adolescents interact over social media on a daily basis, and this computer-mediated communication is pivotal to their self-esteem in positive and negative ways.[4] While social media is a way to engage and share exciting accolades with others, this technology has compressed social networks in ways that enable the weaponization of shame and embarrassment on a broad scale.[5] In fact, nearly 60 percent of adolescents have been victims of one form of online harassment, known as cyberbullying, within the past year.[6] Cyberbullying is any harassing behavior that occurs online by peers that is repeated multiple times.[7] While considerable research has focused on the differences between cyberbullying and traditional bullying, less attention has been given to how the topics of race and racism might impact cyberbullying experiences.[8] As the United States becomes more racially diverse,[9] there is a growing need to understand how race impacts responses to cyberbullying and bystander intervention.

Because of the dearth of research on the influence that race has on cyberbullying and responses to cyberbullying, there is still much that remains unknown. Racial minorities, for instance, may face different types of cyberbullying than their majority peer counterparts. When bystanders are less willing to help[10] or less empathetic toward victims of color,[11] it does not bode well for cases of race-based aggressions on social media. One clear consequence is the underreporting and under-intervention of cyberbullying for marginalized people.[12] While some research has found that bystanders are more ready and willing to jump into action to help Black and Latinx students compared to others,[13] other studies document the exact opposite effect of race, noting that Black and Hispanic students reported more missed opportunities for bystander intervention and a decreased likelihood for intervention.[14] Such differences in the perception and reporting of cyberbullying can contribute to disparate health outcomes. Being the victim of bullying is known to contribute to mental health problems including depression, anxiety, lower self-esteem, and suicidal thoughts.[15] These outcomes definitely manifested themselves in 9-year-old McKenzie Adams.

In 2018, 9-year-old McKenzie Adams hung herself at her grandmother's home in Alabama. What would drive a 4th grader who aspired to be a scientist to commit suicide?[16] Her family blames the repeated bullying she experienced at school with messages such as "kill yourself" and "you ugly… Black b-tch."[17] Not only do these statements call into question the role of gender and physical attractiveness but also race. Many may also wonder why no one intervened on McKenzie's behalf, either teachers or particu-larly other students. The tragic death of McKenzie not only has scholarly and policy implications for bullying and bystander intervention, but it also has implications for the role that race and gender may play in the experiences of victims and the likelihood of bystanders to intervene.

Silence and inaction do nothing but cause biased perpetrator behaviors to proliferate as they feel unquestioned. We know that bystander intervention can help prevent some of the most detrimental

As the United States becomes more racially diverse, there is a growing need to understand how race impacts responses to cyberbullying and bystander intervention.

effects of targeted aggressions on the victim's mental and physical health by intervening to make the aggressions visible, disarming the situation, educating the perpetrator, and seeking external reinforce-ment or support.[18] Furthermore, it seems likely that deliberate bystander stances against racism can establish social norms that will both constrain the racist behavior immediately addressed and even affect long-term social attitudes in the group.[19] Unfortunately, not everyone who witnesses a discrim-inatory incident actually moves to intervene. Especially when looking at cases of online harassment such as the racially-biased aggressions we examine in our report, it is important to look at if, how, and at what stage bystander intervention occurs to better understand it as a key tool against these attacks.[20]

In one investigation on cyberbystanders' willingness to intervene in situations of cyberbullying on Twitter, researchers found that bystanders were more likely to intervene in situations with multiple offenders and situations with original content. These two scenarios influenced bystanders' appraisal of whether a tweet or tweets were considered to be cyberbullying and thus increased their feelings of responsibility to intervene.[21] Another study—examining bystanders' willingness to argue against hate speech on Facebook—identifies the severity of hate speech content on Facebook, the number of other bystanders, and the reactions of other bystanders to the hateful content as determinants of a user's intention to counter hate speech.[22] People online may choose not to engage in bystander intervention if they feel group pressure that leans more in support of a racist comment or of nonaction. Additionally, not all comments are directly aggressive; when responding to racist humor or other forms of microaggressions, the implications for bystander intervention change.[23] While bystanders have been shown to act in cases of high severity aggressions,[24] waiting for such trauma to occur is deleterious to the health of the victim. And, in cases of racial aggressions online, it is an entire community of people that feels the effects of highly visible biased views.

Social media and online spaces set up a much different social interaction order for cyberbullies and victims than offline. What makes cyberbullies such powerful adversaries is their ability to remain anonymous and avoid retaliation.[25] Heavy social media use combined with anonymity facilitates disinhibition and fosters cyberbullying.[26] For people in school settings, when the identity of the perpetrator is known, they are often someone from the victim's school and part of their offline social networks.[27] Therefore, schools are an important setting for studying cyberbullying. Student peers are often the first line of defense to bullying. Peers play a role in identifying and reacting to cyberbullying and can report it to their parents or school authorities. These peer reports are important since targets of cyberbullying are often hesitant to share their victimization with others, particularly parents and teachers.[28] However, teachers and administrators can have the ability to identify bullies and victims, and to intervene, too. In exclusive online and social media spaces, intervention may be more difficult.[29]

The impact of persistent racial bias and violence is facilitated by witnesses to the discrimination who fail to act or intercede.[30] The bystander effect provides important insights on recent acts of racial violence such as the murder of Mr. George Floyd. Social media can be used to study communities impacted by traumatic events,[31] which is in some ways what we see with the uses of #BlackLivesMatter and Black Twitter. Sentiments such as sadness, anger, and anxiety reflect from conversations engaging with biased comments, comments around instances of real-world violence, and the system racism that frames them. Conversely, the positive social reactions most often indicate validation for survivors and calls for social change. Social media platforms offer unique opportunities for targets to share personal stories and receive emotional and social support they may not have access to in person.[32]

> The bystander effect provides important insights on recent acts of racial violence such as the murder of Mr. George Floyd.

Adolescents experience bullying at disproportionate rates.[33] When considering how Black communities already experience a higher risk for other negative health effects, the effects of racial victimization on school performance and social development pose an even more immediate concern.[34] Complicating this problem is that the bullying of yesterday has spread beyond the school domain through technology.[35] The effects of face-to-face bullying and cyberbullying are similar with one major point of divergence. Though both can result in depression, isolation, and suicidal ideation, cyberbullying has catapulted the adolescent suicide rate.[36] Its prevalence has become such an issue that suicide, the third leading cause of death, has a subcategory named after it called cyberbullicide.[37] Despite the high cost to victims, this widespread and fatal social problem lacks community-level approaches for procedural change.

The bulk of the research on cyberbullying continues to come from psychology. This research normally examines the individual components that underlie bullying.[38] While this is important, the ways that social interactions in online environments as well as the topic, in this case systemic racism, are understudied. Though some researchers criticize the individualization of responses of isolated instances of racism that may be reported on social media,[39] the usage of key terms for social media studies allows for a

larger, systemic narrative to be uncovered in the combined discourses. Accordingly, actual content of aggressive posts has not been explored in great detail, nor have the methods of response from bystanders. Research by Schweinberger, Haugh, and Hames (2021) points out that studies on social media that use text-mining approaches to analyze responses to major events often fail to consider the distinct sub-discourses that make up the larger online discourse being examined and can also confound structured behavioral data (i.e., tagging) with unstructured user-generated data (i.e., content of tweets).[40] With this analytic orientation, we combine a data-driven, text-mining approach and examine the corpus of tweets surrounding conversations that interact with racialized key terms to fill these important gaps in the literature of race-based online aggressions.

Bystander Intervention as Social Support

Social support theories provide a framework for understanding who gets bullied, who reports bullying, and who is more resilient to the effects of bullying.[41] In particular, being socially marginalized increases the risk of being bullied.[42] Yet, having a support system in place is associated with better health and wellbeing outcomes after difficult events.[43] This is known as the "buffering effect of social support."[44] The buffering effect highlights the power of the community to intervene. Bystander intervention can have a direct impact to change social interactions, mental health outcomes, and future life events.[45] To design these group-based interventions, however, group-level approaches are needed.[46] Peer prevention strategies to traditional bullying have demonstrated the highest rates of success, which makes the role of the bystander an advantageous area for the disruption of the aggression process.[47] Because bystanders typically choose to not intervene,[48] there is ample room to design strategies that increase intervention rates, which may act as central buffers to cyberbullying.[49]

In this regard, our work advances the current literature on peer aggression through cyberbullying in a few key ways. First, our work treats bullying as a social network process to which the audience is a participant, rather than strictly as a series of individual effects. Secondly, we pay special attention to ways that bullying events are race-based where prior studies have focused almost exclusively on gender and sexuality. Lastly, we investigate the role, recognition, and motivations of the bystander to intervene.

The knowledge gained from this research is useful to develop community-based intervention strategies for local implementation and policy adoption. Whereas much research has treated bystanders in social media environments as a passive group, we treat this population as a powerful group whose actions enforce an existing social order.[50] By focusing on reactions to social media posts and comments about systemic racism in determinants of bystander intervention, this study addresses peer aggression in digital spaces and the growing public health issue of cyberbullicide. This study breaks ground on analyses of interactions in that it offers a typology of social media conversations on race and informs practitioners of best practices for the use of social networking sites as a toolkit for anti-racist education.

Methodology

We conducted a quantitative and qualitative analysis of Twitter tweets and Reddit posts and a content analysis of comments. The methodological approach used is similar to previous analyses used by members of the project team as co-editors of a special issue in *Ethnic and Racial Studies* that examined the role of Twitter in the rise of the #BlackLivesMatter movement.[51] This project conducted an analysis of the following systemic racism focus areas: systemic racism, police brutality, education inequality, employment, and wealth. We examine these four domains using synonyms to each term to search hashtags on Twitter in 2020 and identify posts on Reddit in 2020 that use these terms.

On Twitter, a micro-blogging social networking site, users create profiles where they generate text-based digital content in 280 characters or less. Users can also use Twitter to: upload and exchanges audiovisual content such as URLS, images, videos, or GIFS; use hashtags, geotags, or emojis in text posts; reply to other users; interact with other users' posts through likes, quote tweets, or retweets; follow others and maintain a following; connect to other users privately through direct messages; and save posts as bookmarks. Excluding bookmarks, direct messages, or content made by private accounts, we accessed the content users generate on Twitter through its Application Programming Interface (API). We used web crawling and digital ethnographies to generate datasets on social networking sites (SNSs), which essentially include internet-based services that enable people to (1) create user profiles to participate within a virtual, bounded social system, (2) share a connection with other users, and (3) view the content generated by other users that participate in this virtual social system.[52] Mobile social networking (MSN) applications include the features of SNSs, and users access them through the interface of mobile internet-enabled smartphone technology.[53]

Through web crawling, we used the API to access archived user-generated content and its meta-data—"the data about the data which describe the main data content in some way— that includes information such as "the creator or author of the network data, time and date of creation, sources of data, means of creation, legal rights, etc."[54] Then, we generated datasets in multiple file formats (e.g., JSON or CSV) suitable for quantitative analysis with software like Microsoft Excel or Tableau. Qualitatively, we conducted a digital ethnography by drawing upon traditional qualitative techniques to examine Twitter tweets and Reddit posts. We observed the dynamics, digital practices, and connections of a set of users related to a particular digital artifact such as a hashtag or keyword.[55]

Selected Platforms

There are some important similarities and differences between Twitter and Reddit that make them optimal for this analysis. While Twitter permits only favorites of posts, users on Reddit can both "downvote" or "upvote" a post. Additionally, though Twitter enables any user to add audiovisual content to a post, Reddit permits the creator of the original post to insert images or video into the body of their post, but does not permit direct uploads of images in comments or replies. Lastly, while Twitter enables users to link to similar posts through the use of hashtags (hyperlinked text preceded by a '#' symbol) embedded in a post, Reddit does not have functional hashtags and users can only connect to other posts through direct links.

While we used computational approaches for collecting and selecting data, our data analysis pipeline was designed for sampling and analyzing the conversations using qualitative methods. The Twitter and Reddit platforms were not chosen because they provide a representative sample of what bystander interventions look like more generally in online conversations. Social media platforms have their own particular affordances that encourage specific modes of online interaction, and are used by particular types of users, that do not necessarily reflect the larger population of people who converse online. However, these interactions do provide topical, deeply contextual snapshots of how bystander interventions occur online in social media, as well as what their outcomes look like. Reddit and Twitter provide crucial examples of how bystander interventions can manifest online.

Another factor for the selection of Reddit and Twitter is that these platforms allow for programmatic access to conversation data using Application Programming Interfaces (APIs). APIs allow searches to be performed without bespoke, and often error prone, web scraping techniques. The data collection and selection were done using Jupyter Notebooks, and qualitative coding was performed in Excel. All the code used, and identifier datasets for collected social media posts have been made available as a public dataset.[56]

Using a software service called Twarc as well as Twitter's new native API data collection for academics, we used the metadata for quantitative analyses of posts that contain the keywords. We quantified, per day: the frequency of tweets that contain the keywords; the frequency of likes for each post; the hashtags associated with these keywords; the frequency of retweets for each post; and the users most frequently retweeted whose posts contain the keywords. Using Nvivo 12 software, we conducted a content analysis to identify and describe the patterns observed through digital ethnography, maintaining field notes of the virtual field site to ascertain the patterns of interaction that strengthen or weaken user responses to the topics related to the keywords. Nvivo 12 will also allow the team to generate visualizations such as wordclouds to assess which terms occur most frequently alongside the keywords.

While researchers have examined how users of Twitter engage topics related to structural racism and police violence,[57] fewer studies examine the dynamics of conversations on these topics on Reddit. Reddit functions as a repository of countless web forums (subreddits) that users can access on a wide range of topics. While anyone can observe a subreddit through a web browser, users must sign up for a profile to comment on text-based posts and interact with other users. Unlike Twitter, Reddit

does not permit users to upload profile pictures. Instead, users create avatars based on Reddit's "Sloo" logo, which resembles an extraterrestrial creature. This creates an environment where fewer people reveal their true identities on Reddit compared to Twitter, where people often signal who they are in the biography section of their profile pages or through profile pictures. Once users have an account, they can start a conversation in a subreddit as a title post and expand on it in the body of the post. Users, known as moderators, create these subreddits and maintain a set of norms on them that are typically referred to as rules. Other users can interact with the original post through comments, replies to other users' comments, or voting on the post or comments and replies.

Accordingly, we analyzed keywords in the following subreddits that each contain over 150,000 subscribed users: R/2020policebrutality, a forum created in the wake of the death of George Floyd; r/BlackLivesMatter, a forum centered on news related to the Black Lives Matter movement; and r/Conservative, a space for political conservatives on Reddit. In addition to using the same digital ethnographic strategies proposed for Twitter, the team will also use the Reddit API to collect data comparable to the data collected on Twitter. Since these are two different software services, the type of data collected will be similar, though not identical. The team will use the software service PRAW or Pushshift to access the Reddit API and collect metadata on posts that contain the keywords. This data, measured per day, includes: the number of users per discussion thread; number of posts in a subreddit; number and percentage of upvotes for a post; and the number of comments and percentage of upvotes for comments.
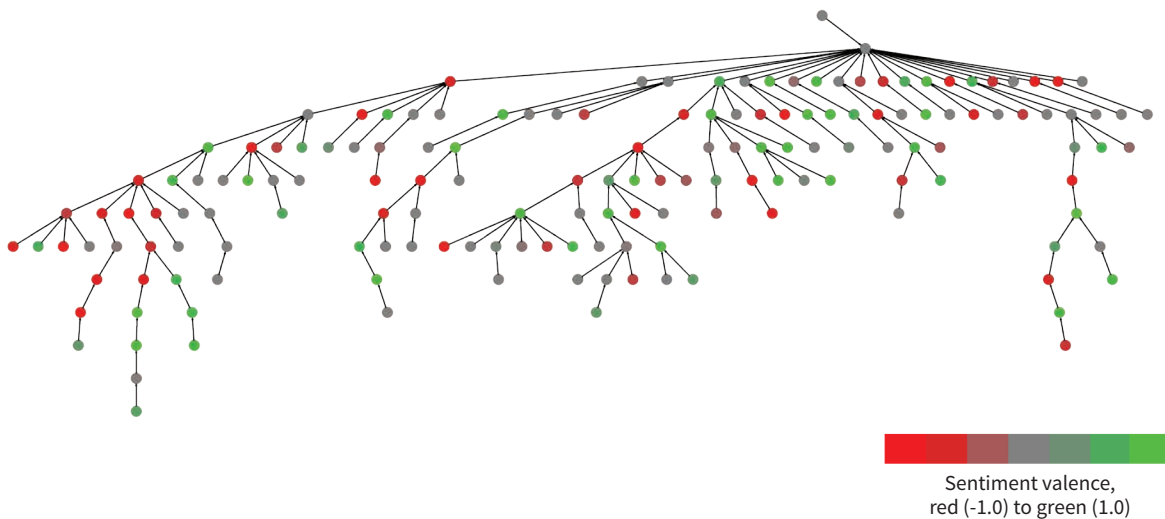
As shown in Table 1, we selected three search queries: "racism", "black people" and "racial wealth gap" to run on each platform. Search results were limited to the year following the murder of George Floyd, starting May 25, 2020, and ending May 25, 2021. These three queries were chosen because these key terms constitute topics of interest to policymakers invested in proposing solutions toward racial equity. Additionally, these terms reflect topics that emerge in mediated discourse on race and racism. For the Twitter platform we took the extra step of limiting our search to original posts, which filtered out retweets (amplifications of specific posts by other users) and would potentially cause oversampling of shared conversations. The twarc tool [58] was used inside a Jupyter notebook to collect the tweets from Twitter's API using Twitter's Academic Research Track API access, which allows for searching for historical tweets, but is limited to 5 million tweets per month.

In order to obtain the Reddit data we used PushShift, which is a third party service that has collected Reddit posts in real-time since 2015.[59] PushShift makes Reddit posts available as a downloadable dataset, but also provides an API service which lets researchers run queries and collect results as structured data (JSON). Once the searches were completed, we analyzed the collected Twitter and Reddit data to identify conversations that occurred. The search results themselves did not provide a complete picture of conversation threads, since not all posts in the conversation may have included the query. Fortunately, both Twitter and Reddit employ a unique identifier for each conversation thread (*conversation_id* and *parent_id* respectively). We identified the unique conversation ids for each dataset and then randomly sampled 30 conversations for each query/platform combination to analyze, resulting in a total of 180 sampled conversations. Some of the conversations were small (n=2) whereas others large (n=43,186).

Table 1: Topics of Analysis by Platform

Query	Platform	Posts	Conversations	Selected Posts
"racism"	Twitter	1,000,377	214,756	4,918
	Reddit	103,993	48,013	1,669
"black people"	Twitter	1,000,066	231,400	1,420
	Reddit	33,571	17,638	4,019
"racial wealth gap"	Twitter	36,246	8,725	8,012
	Reddit	342	97	5,473

The 30 randomly selected conversations[60] for each query were then analyzed to extract the directed network graph of posts in order to visualize the structure of conversations. Each post was assigned a sentiment score using Vader, since it was designed to be used with micro-blog size social media posts similar to Twitter and Reddit.[61] A minimal web application was developed to allow for rapid viewing of each conversation.[62] Each node in the network is a post which is made in response to another post (unless it is the first post, or conversation head). This visualization was color coded on a gradient scale from red to green to indicate the sentiment valence, red (-1.0) to green (1.0) (see Figure 1 for example). Twitter conversations often appeared fragmented when interstitial posts had been deleted, which broke the reply chain, while the conversation identifier still allowed the disconnected posts to be recognized as part of the same conversation.

Figure 1: A Reddit Conversation Thread

Sentiment valence,
red (-1.0) to green (1.0)

View our Twitter and Reddit conversation maps with this interactive query tool.

Analysis

Quantitative Text Analysis

We analyzed this data using both quantitative and qualitative techniques. First, we utilized natural language processing (NLP) techniques to engage in quantitative textual analysis to get insight into the words and phrases that recurred throughout the corpus of material for each social media platform concerning the racial wealth gap, racism, and Black people. Then, while we gathered the data and stored it in JSON files, we utilized Python to transform it into plain text files for each query (CSVs). Following that, we imported each CSV file encoded in UTF-8 straight into R Studio as data frames, where we began analysis using the R programming language and software environment.

Our quantitative text analysis procedure consisted of many stages, and the first was to clean the text in each data frame. To begin, we cleaned the data twice. We started by eliminating hyphens, retweets, and trailing whitespace on the first pass. Additionally, we replaced "&" with "and," substituted a blank space for newline characters, and converted all text to lowercase, storing this content as a new column in the current data frame. After removing all URLs, punctuation, numerals, and remaining special characters in the second pass, we replaced the original cleaned data with the stripped-down text. Second, we utilized this condensed material to build a corpus of text. Following creating the text corpus, we decided to eliminate stop words using the standard English dictionary provided in the "tm" package. Third, we determined the 25 most commonly occurring phrases in each data frame to evaluate additional stop words to remove. Then, we generated a list of additional stop words that we deleted from each data frame before storing our final text corpus.

We utilized the resultant text corpus to generate data frames or text corpora for summary statistics following data cleansing. In contrast, many of the subsequent processes are the same for the Twitter and Reddit datasets, the differences in how communication occurs on each platform resulted in somewhat different data. While we used our keywords to identify individual tweets on Twitter, we also used them to identify Reddit posts' titles using Pushshift.io. Furthermore, Reddit headlines do not include hyperlinked hashtags or mentions seen in tweets. We created a smaller data frame for both sets of data that included the analysis' main metrics. The Reddit data included: id, author, created, num_comments,

score, title, title_clean. For the Twitter data, this included: id, author.username, author.public_metrics. followers_count, created_at, public_metrics.quote_count, public_metrics.reply_count, public_metrics. retweet_count, public_metrics.like_count, lang, text, text_clean.

Following that, we created data visualizations for each piece of data. First, we displayed the titles of the top five Reddit posts according to the number of comments per user using the Reddit data. The top 25 most often used words in each title were then identified. Additionally, we examined the frequency over time of posts containing our keywords in the title. Then, using the top 100 terms for each keyword, we generated word clouds. Following that, we determined sentiments (positive, negative) and emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) using the NRC Word-Emotion Association Lexicon, a list of English language words regularly used in studies of text data.[63]

Additionally, we displayed the top 10 titles that included five-word sentences, including each key term. We also visualized the top 10 titles that mentioned each keyword in five-word phrases. Finally, we generated a document word matrix from the titles of each Reddit post for topic modeling, which involves "statistical methods that analyze the words of the original texts to discover the themes that run through them, how those themes connect, and how they change over time".[64] Our topic models use latent Dirichlet allocation (LDA), a method intended to produce distributions of words throughout a collection of documents inside a set of discrete data, especially text corpora, expressed as multiple topics.[65] In addition to the models and visualizations described earlier, we performed a few extra steps with our Twitter data. First, we visualized the top 15 users' tweets, retweets, and likes. Additionally, we identified the top 50 hashtags and mentions associated with each key phrase in tweets. Following that, we showed which hashtags and mentions co-occurred.

Qualitative Text Analysis

Our qualitative study of this social media data concentrated solely on the 180 discussions produced across the two platforms—90 on each platform and 30 per crucial phrase. We analyzed these data using the grounded theory technique, an inductive approach to data gathering and analysis that entails a two-step coding procedure, memo writing, and theoretical sampling.[66] While we utilized network visualizations to see how discussions diverged from the original post, we also downloaded CSV files of the talks organized by timestamp. Then, we examined them directly in their individual platforms' browsers. In the CSV files, we kept track of the term that was the focus of the conversation, the total number of tweets or comments in that conversation, the total number of users in the conversation, the average sentiment score for the conversation, a direct quote from each conversation to serve as an example of the dialogue, whether or not bystander intervention occurred, whether more than one conversation occurred in each thread, and a URL for the initiating post or tweet. For the Reddit data, we also kept track of deleted users and text, mainly by moderators, and the name, description, and

URL of the subreddits that each conversation took place on. In total, these conversations took place across 66 different subreddits.

We assessed whether or not an interaction between users included bystander action during the first coding phase. We defined bystander action as occurring anytime a user challenged another user for alleged acts of racism or responded to an absent third party's observed racism (e.g., the subject of a news story). First, we documented the scene of the bystander intervention via memo writing, noting which user started the discussion and what they wrote. Additionally, we documented which users responded to this first user, and what they wrote. We next investigated what led people to disagree over the subject of the discussion or to challenge the views of a third person who was not there.

Additionally, we assessed if and how disagreements transpired. We then utilized these notes to create categories from which we could code the various forms of racism and bystander intervention that we saw. We then used these categories to analyze and interpret the identified scenes of bystander intervention, allowing each conversation to fall within multiple categories, which we elaborate on further in the findings section.

Research Findings

Quantitative Findings

Our quantitative methods relied on textual analysis techniques that identify commonly occurring words or phrases in vast text corpora. We used these methods for each of the three keywords relevant to our study and analyzed text data from each platform using similar strategies wherever feasible. Applying quantitative techniques to the patterns of recurring text that occur in conjunction with the keywords in our dataset offers insight into the overarching narratives that define social media discussions about these subjects.

For the purposes of this report, we focus on the top 25 most frequently used terms in tweets and the titles of Reddit posts for each key term. For tweets, we also look at the top 25 most frequently used hashtags and the top 25 most mentioned usernames. In general, our results indicate that themes or speech promoted by mainstream media members and, especially journalists, celebrities, and politicians, spark social media discussions. However, social media allows ordinary people to engage with members of the mainstream media and initiate independent conversations about race and racism.

Tweets and Reddit Posts that Reference "Black People"

We collected 1,000,066 tweets and 33,571 titles of Reddit posts that contained the phrase "Black people." Figures 2A and 2B show the most frequently used terms associated with the use of the phrase on Reddit and Twitter. Aside from the two terms that make up the primary phrase for this keyword search, the top four most used terms for both tweets and Reddit titles are the same: "white," "police," "racist," and "racism."

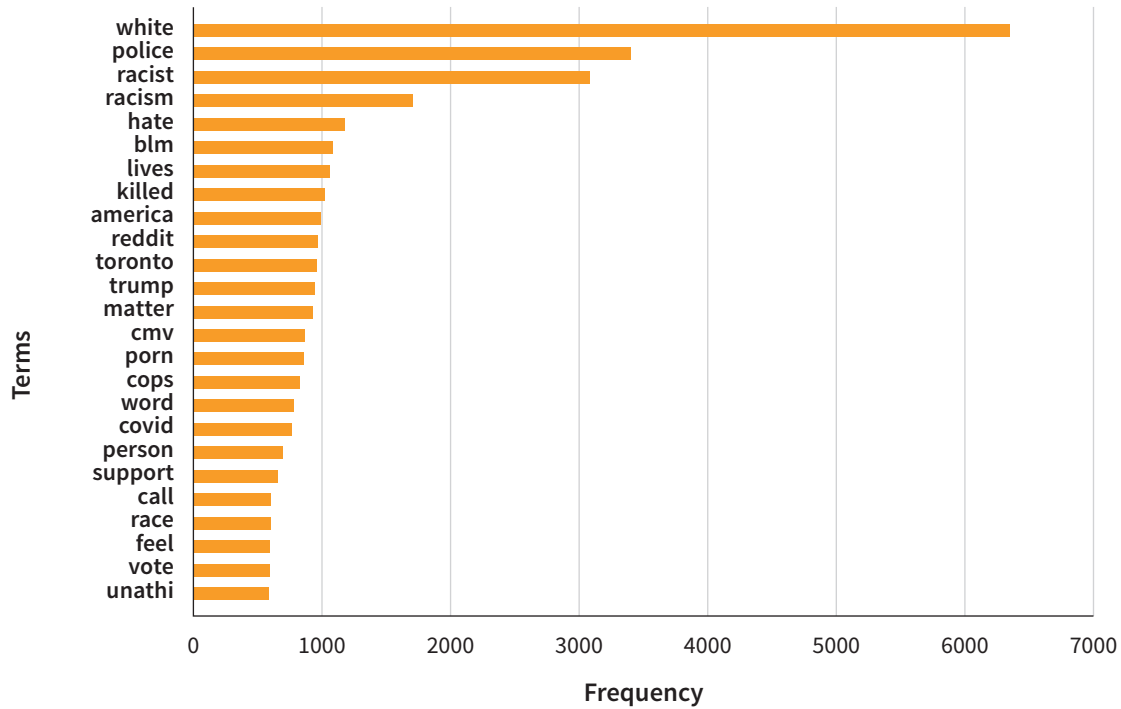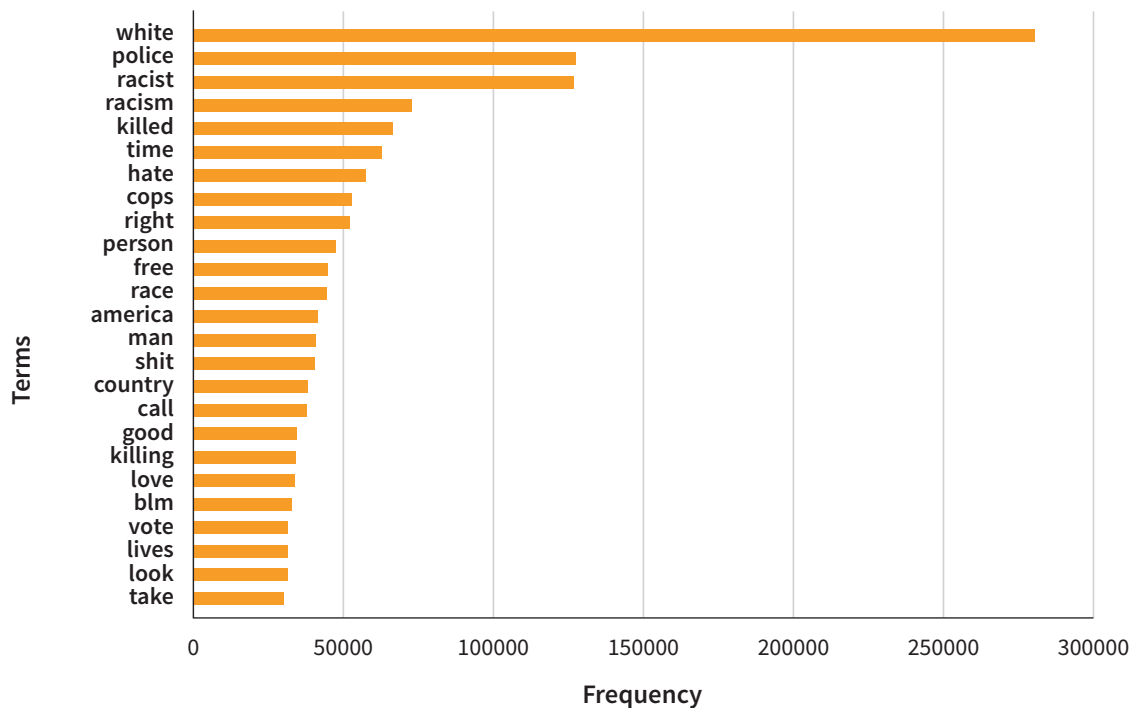## Figure 2A: Most Frequently Used Terms in Titles of Reddit Posts about Black People

Figure 2B: Most Frequently Used Terms in Tweets about Black People

Aside from these terms, several other words re-occurred within the text posted on these platforms, albeit at differing frequencies (see Tables A.1-1 through A.1-3 in Appendix). These terms include "killed," "hate," "cops," "person," "race," "America," "BLM," "call," and "lives." Words that appeared most frequently in tweets that reference Black people, but not in Reddit posts, include "time," "right," "free," "man," "shit," "country," "good," "love," and "vote." Words that appeared in the title of Reddit posts about Black people, but not in tweets, include "Toronto," "Trump," "matter," "cmv," "porn," "word," "covid," "support," and "feel."

Given the time period for which we chose to collect this data, these results suggest that the most frequent discourse on the text we collected appears to associate Black people most routinely with topics of police brutality, particularly from white police officers, and also associates Black people with race and racism both on Reddit and Twitter. Further, to varying degrees, the dialogue around Black people on these platforms also brings up a reference to anti-racist social movements (e.g., "BLM," "lives"). However, while Twitter users reference Black people for topics related to politics or govern- ment more broadly (e.g., "vote," "country," "free"), the titles of posts we collected from Reddit touch on a broader range of topics, likely due to the ways the platform silos topics into different categories through subreddits.

Figure 3 shows the matrix of the top fifty hashtags that tended to appear together in tweets that referenced Black people (see Tables A.2-1 through A.2-3 in Appendix for a full list of hashtags).

Figure 3: Matrix of Hashtags that Co-Occur in Tweets Containing the Phrase "Black People"

While one set of hashtags co-occurred together frequently in tweets associating Black people with hashtags that addressed issues related to the Middle East and North Africa (e.g., "#gazaunderattack," "#jerusalem," etc.), Figure 3 indicates that "#freepalestine" appeared to be the primary link between this set of tweets and another set of tweets about Black people that contained phrases about Black people. This dense network of hashtags included several hashtags typically associated with topics that affect Black people including the names of victims of police brutality (e.g., "#makhiabryant," "#dauntewright," #justicefordauntewright), references to anti-police brutality campaigns or movements (e.g., "#defundthepolice," "#blm," "#policereformnow"). Sparse networks of hashtags addressed topics such as politics (e.g., "#votersuppression") or entertainment (e.g., "#verzuz," "#oscars"). Overall, the hashtags that most frequently appeared in the tweets we collected that referenced Black people appeared to tap into a diversity of hyperlinked discourses broadly related to structural racism.

Figure 4: Matrix of Username Mentions that Co-Occur in Tweets Containing the Phrase "Black People"

Figure 4 displays the matrix of usernames mentioned in tweets about Black people (see the full list in Table A.3 in the Appendix). By and large, the network of mentions in tweets about Black people was

sparse, indicating that the individuals who produced the text we examined seldom referred to more than one person when tweeting about Black people. When users sent tweets about Black individuals, they often referenced the accounts of mainstream media personalities or governmental organizations. President Joe Biden, the Republican Party, White House Domestic Policy Advisor Susan Rice, Attorney Ben Crump, and Senator Tim Scott are among the political personalities and organizations named. *The Hill*, the *New York Times*, ABC, the *New York Post*, and *The Daily Caller* are among the media outlets cited. Finally, Sean Hannity, Joy-Ann Reid, Ben Shapiro, Andy Ngo, and Matt Walsh are all named as media personalities. Finally, these tweets on Black people included references to individuals with media exposure or platforms, such as Charlie Kirk, Tomi Lahren, Candace Owens, Bree Newsome, and LeBron James. By and large, cited usernames suggest that both left-leaning and right-leaning media and political personalities are often addressed in the tweets we gathered regarding Black people.

Tweets and Reddit Posts that Contain the Phrase "Racism"

While the text we collected from Twitter that referenced racism included 1,000,377 tweets, our collection of Reddit posts included 103,993 titles. Figures 5A and 5B display the top 25 most frequently used terms on tweets and Reddit titles about racism.

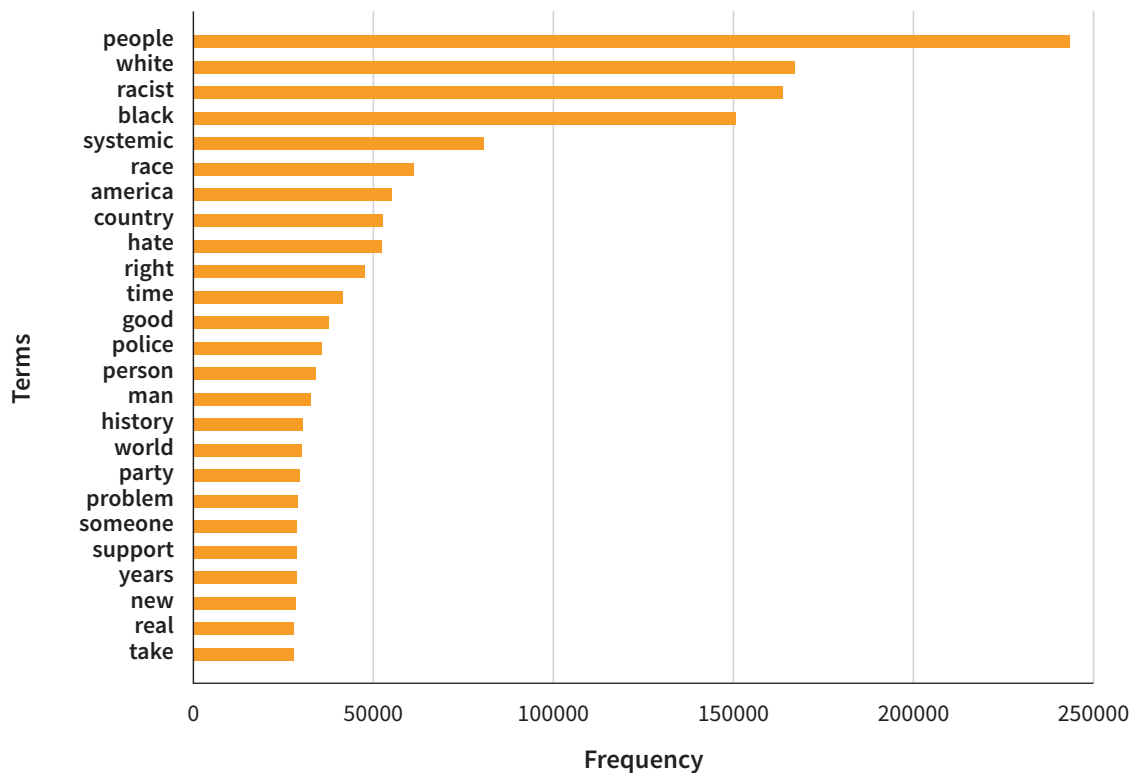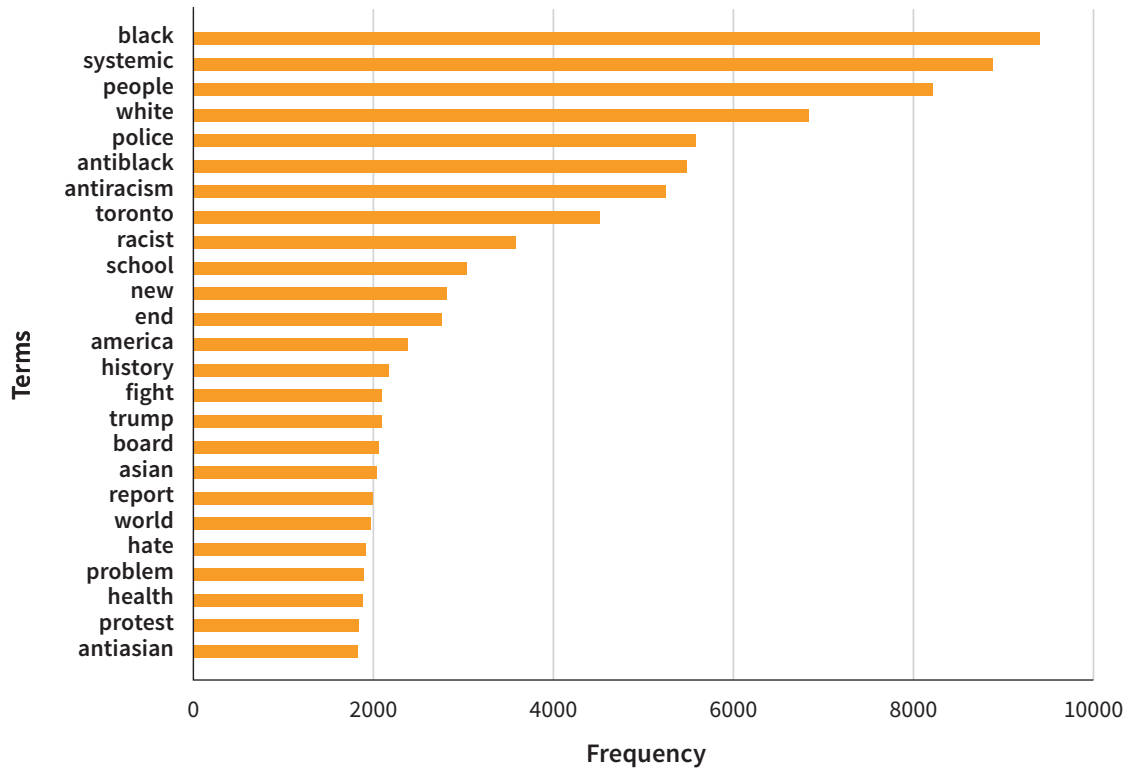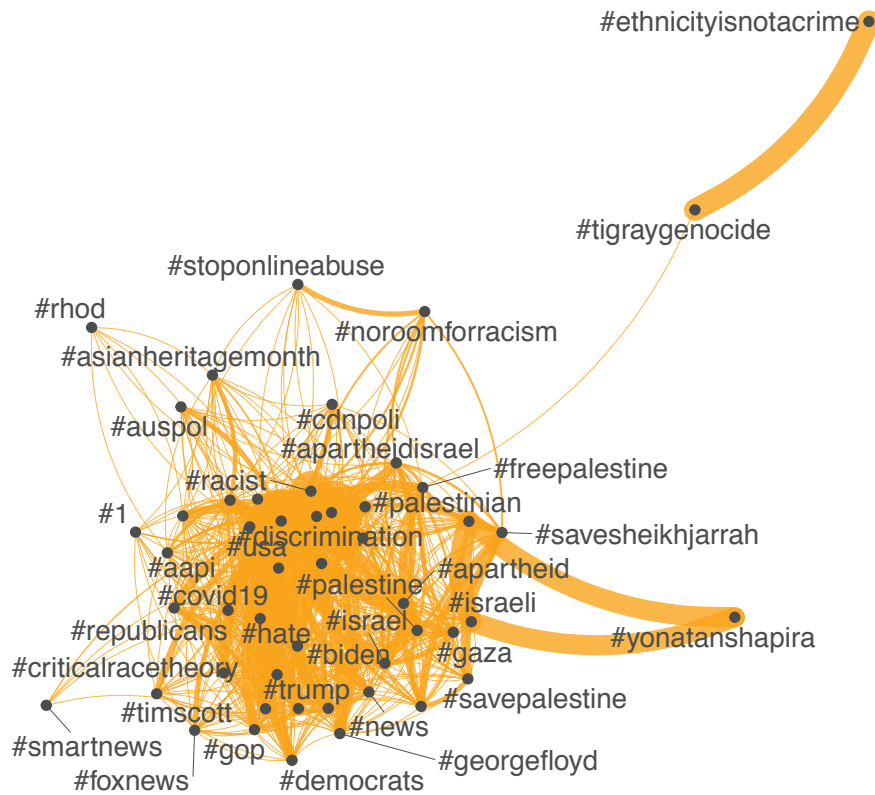### Figure 5A: Most Frequently Used Terms in Tweets about Racism

Figure 5B: Most Frequently Used Terms in Titles of Reddit Posts about Racism

To varying degrees, users on both platforms, aside from the terms "racism" itself most frequently used the phrases "people," "white," "racist," "black," "systemic," "America," "hate," "history," "police," "world," and "new." Twitter users also frequently mentioned the words "race," "country," "right," "time," "good," "person," "man," "party," "problem," "someone," "support," "years," and "real" in their tweets about racism. Reddit users, on the other hand, frequently mentioned "anti-black," "anti-racism," "Toronto," "school," "end," "fight," "Trump," "board," "Asian," "report," "problem," "health," and "protest" in the titles of their posts. On both sites, users seem to connect racism with systemic problems but also use the term in discussions about white individuals and Black people. While the words associated with racism on Twitter imply that people debate the phenomena in terms of broad societal or political issues, allusions to racism in Reddit post names often evoke particular, local concerns.

Figure 6 displays a matrix of the hashtags that co-occurred most frequently in tweets about racism. Hashtags that overlapped most frequently revolved around broad topics related to racism (e.g., "#systemicracism," "#hate," "#discrimination," etc.) and those more specific to racial minority groups in the United States (e.g., "#stopasianhate," "#aapi," "#antisemitism," etc.).
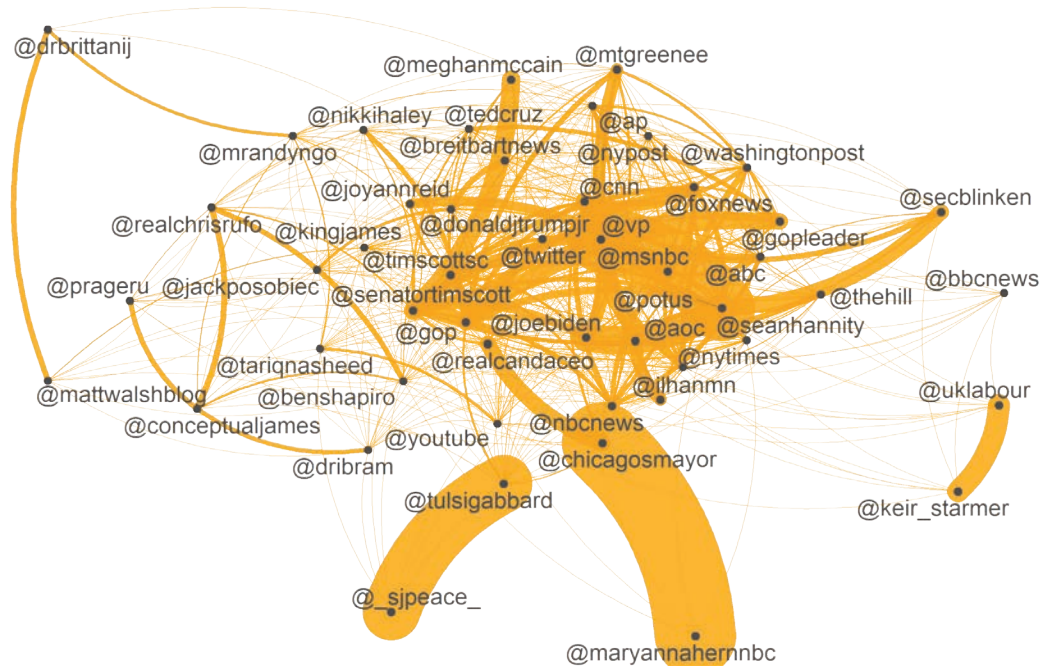
Figure 6: Matrix of Hashtags that Co-Occur in Tweets Containing the Phrase "Racism"

While tweets that included references to racism and hashtags related to Middle Eastern or North African issues tended to occur outside the network of tweets on systemic racism and other racial minorities, hashtags like "#palestinian," "#gaza," and "#israel," were more closely associated with hashtags included alongside others in tweets about racism. Another somewhat sparser cluster of hashtags referenced either conservative news outlets or politicians, while hashtags that referenced Democrats and Joe Biden appeared more closely to tweets that mentioned systemic or ethnic-racial issues.

Figure 7 displays the matrix of usernames in the tweets we collected referencing racism. Like the tweets collected that referenced Black people, the majority of mentioned usernames in the tweets on racism referred to media or political figures. Users more frequently mentioned the usernames of accounts for political entities or figures like Senator Tim Scott, Alexandria Ocasio-Cortez, and the GOP alongside the usernames of media outlets such as CNN, the *New York Times*, MSNBC, and the *New York Post*.

Figure 7: Matrix of Username Mentions that Co-Occur in Tweets Containing the Phrase "Racism"

Less prominent U.S. political figures and media outlets were less frequently mentioned together in tweets about racism. To a lesser extent, tweets about racism would include tweets that mentioned political and media entities in the United Kingdom including the UK Labour Party and BBC News. Overall, the way users invoked political and media figures in tweets about racism suggest that these entities drive a large share of social media discourse on racism, likely due to the ways media outlets amplify politicians, political parties, and government agencies who in turn use the features of Twitter, such as retweets, quote tweets, and mentions, to engage with media outlets online.

> The way users invoked political and media figures in tweets about racism suggest that these entities drive a large share of social media discourse on racism.

Tweets and Reddit Posts that Contain the Phrase "Racial Wealth Gap"

For this analysis, we collected 36,246 tweets that included the phrase "racial wealth gap," and 342 Reddit posts that included the phrase in the title between May 2020 and May 2021. Figure 8A displays the top 25 most frequently used terms on tweets about the racial wealth gap. Figure 8B displays the top 27 most frequently used terms in Reddit posts that included the phrase in the title. Since fewer words occurred more than 25 times overall in this sample, we chose to display words that occurred at least 12 times.

Figure 8A: Most Frequently Used Terms in Tweets About the Racial Wealth Gap
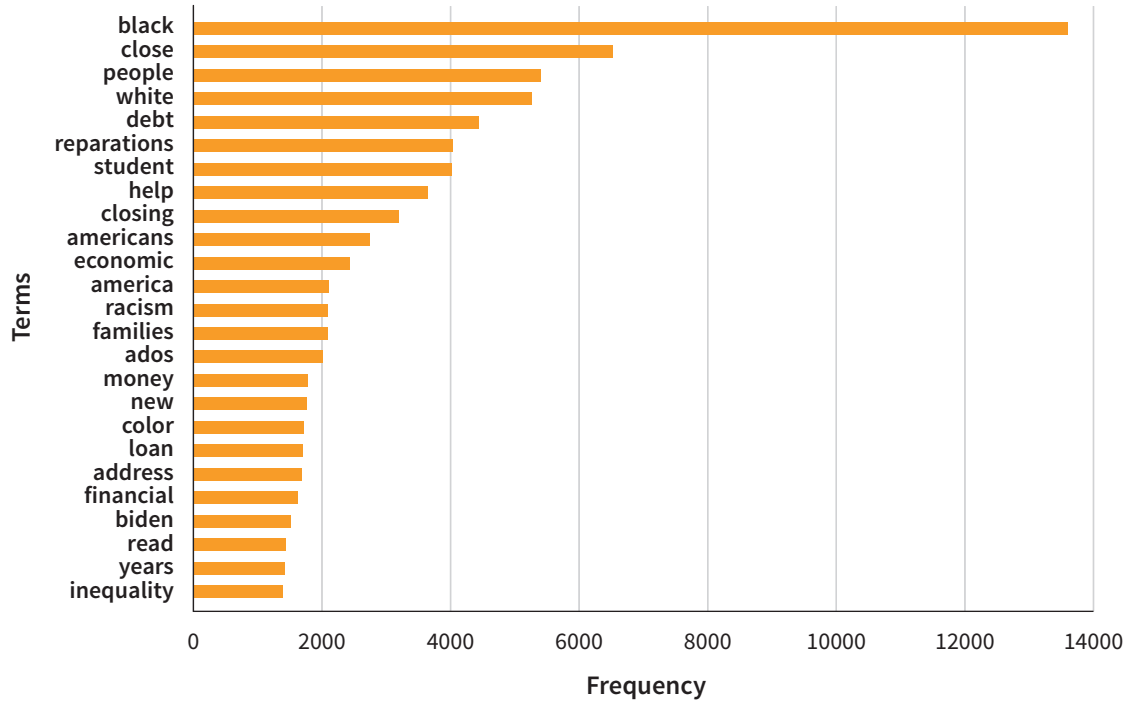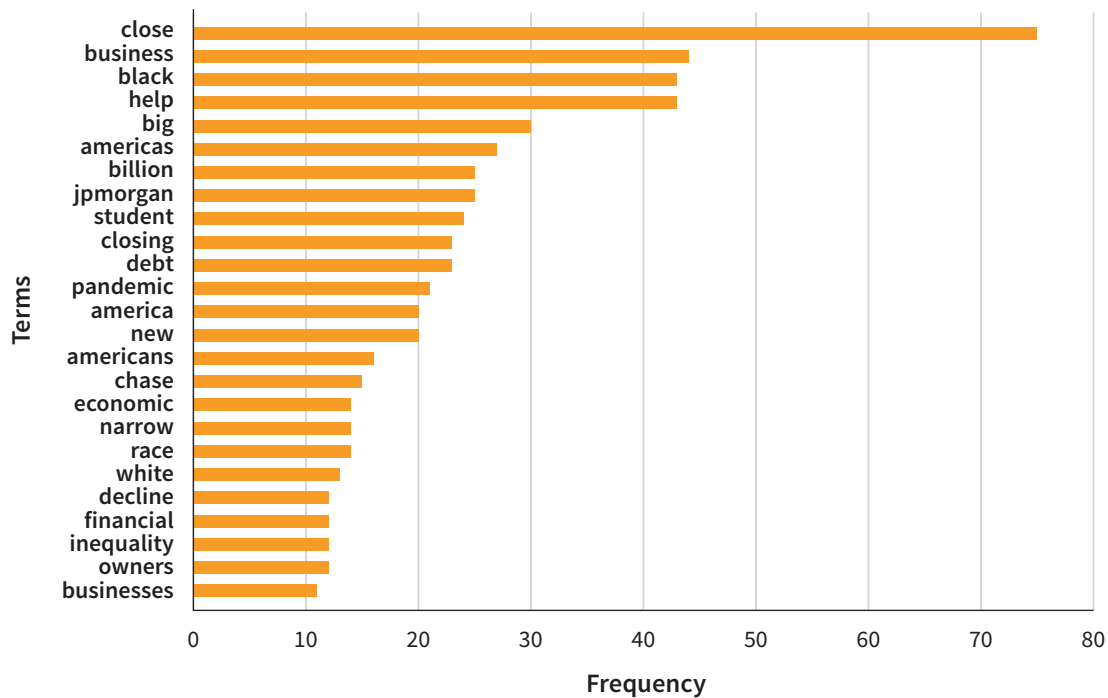

Figure 8B: Most Frequently Used Terms in Titles of Reddit Posts About the Racial Wealth Gap

Note: Due to the smaller size of the dataset, this figure displays only terms that appear at least 12 times.

Other than the key terms, the words that occurred most frequently on both platforms, though to varying degrees include "Black," "close," "closing," "white," "debt," "student," "help," "Americans," "America," "economic," "financial." Terms that showed up most frequently in tweets on the racial wealth gap, but not titles of Reddit posts, include "people," "reparation," "racism," "families," "ados," "money," "color," "loan," "address," and "Biden." Terms appearing most frequently in Reddit post titles, but not in tweets, include "business," "big," "Americas," "JPMorgan," "pandemic," "new," "Chase," "narrow," "race," and "decline."

Our results indicate that users of both platforms perpetuated dialogue around the racial wealth gap that addressed it as a characteristic of racial inequality between Black people and white people that required assistance to ameloriate, possibly through programs that targeted student debt. However, Twitter users were interested in more systemic or broad topics related to the racial wealth gap particularly in the financial sector (e.g., "money," "reparation," "loan"). Reddit users, however, appeared to respond to the notion of the racial wealth gap in reference to financial institutions such as JP Morgan Chase Bank.

Figure 9 depicts the hashtag matrix for tweets about the racial wealth gap. One sparse cluster of hashtags linked together issues related to finance (e.g., "#banking," "#fintech," "#business," etc.) to issues of race and racism (e.g., "#systemicracism," "#racialequity," "#diversity," etc.).

Figure 9: Matrix of Hashtags in Tweets with the Phrase "Racial Wealth Gap"

These hashtags appear to occur together, though they link to a broader set of issues related to the racial wealth group through #reparations. This hashtag connects to other issues of economic inequality (e.g., "#cancelstudentdebt"), but appears primarily with hashtags that address reparations as a topic of systemic racism (e.g., "#supporthr40," "#1619project," "#americandos," etc.). This matrix of hashtags indicates users on Twitter who discuss the racial wealth gap using hashtags see the topic as one that bridges issues of systemic economic and racial inequality, particularly through the vehicle of reparations as a solution.
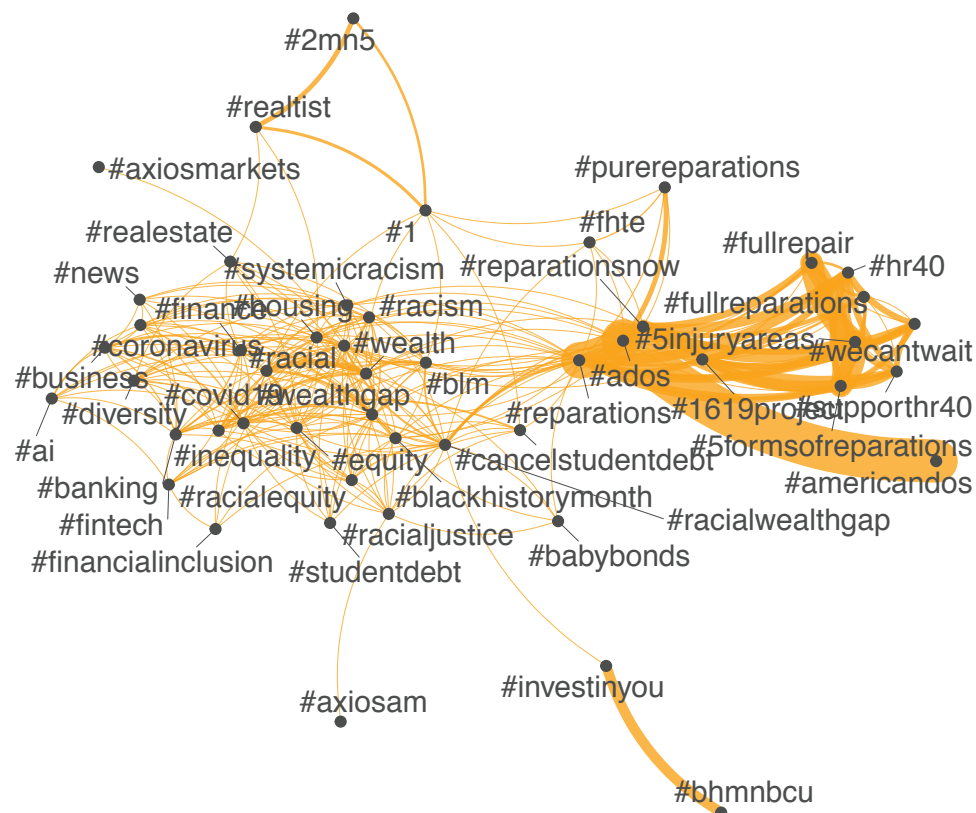
Figure 10 displays the matrix of usernames mentioned in tweets that reference the racial wealth gap. The most frequently mentioned usernames in tweets about the racial wealth gap follow a similar pattern to the matrix of usernames for the other key terms, in that the most frequently mentioned usernames refer to political or media figures or entities. Named politically-associated usernames included Vice President Kamala Harris, former President Donald Trump, former President Barack Obama, Senator Chuck Schumer, and Senator Elizabeth Warren. Usernames for political entities also include the Congressional Black Caucus and the NAACP.

Figure 10: Matrix of Username Mentions in Tweets with the Phrase "Racial Wealth Gap"

Mentioned usernames of media outlets include those of *Forbes Magazine*, the *New York Times*, *Business Insider*, *Politico*, and the *Washington Post*. Given the topic at hand, mentioned usernames also included accounts associated with JP Morgan and Citi Group. Unsurprisingly, given the scope of the work done here, the username for the primary Brookings Institution account (@brookingsinst) also appeared as a frequently mentioned username in this sample of tweets.

Figure 10 shows that usernames associated with political figures or entities co-occurred together whereas references to media outlets or financial institutions were sparser. These findings suggest that at least within the sample we collected from Twitter, discourse on the racial wealth gap revolves primarily in reference to users affiliated with politics, though the hashtag matrix in Figure 9 suggests the conversations bridge topics of racism and economic inequality.

We found that just 16 percent (15 of 90) of Twitter discussions featured bystander action.

Qualitative Findings

We found that just 16 percent (15 of 90) of Twitter discussions featured bystander action. Comparatively, 39 percent (35 of 90) of Reddit conversations included bystander action. As shown in Table 2, racist discourses were coded as one of four types: stereotyping, scapegoating, accusations of reverse racism, and echo chambers. We also found four types of bystander intervention strategies, which include: call-outs, insults or mocking, attempts to educate or provide evidence, and content moderation. Two of these bystander intervention strategies (education or evidence and content moderation) are clearly better than others. We label these intervention strategies as prosocial. Conversely, call-outs and insults and mocking are labeled as antisocial. Below, we provide additional details on the types of racist discourses and intervention strategies. The section below describes how the dynamics of disagreements on topics related to race and racism transpired on Reddit and Twitter.

Table 2: Percentage of Conversation Threads that Involved Bystander Intervention in Racist Discourse by Category

	Category	Reddit	Twitter
Racist Discourse	Racial stereotyping	57%	20%
	Accusations of reverse racism	40%	47%
	Racial scapegoating	37%	73%
	Racism echo chambers	11%	27%
Bystander Interventions	Prosocial education or evidence	77%	73%
	Content moderation	43%	13%
	Antisocial call-outs	40%	73%
	Insults and mocking	9%	60%
	Total discussions featuring bystander intervention	15	35

Note: Due to the multivocal nature of social media discourse, several conversation threads contained multiple categories.

Types of Discourses

Seventy-three percent of the 15 tweets online bystanders intervened in involved racial scapegoating, 47 percent included charges of reverse racism, 27 percent involved echo chambers of racism, and 20 percent of the tweets that bystanders intervened in included racial stereotyping. Comparatively, 57 percent of the 35 conversations we analyzed on Reddit included stereotyping. Forty percent included charges of reverse racism. Thirty-seven percent included racial scapegoating, and 11 percent included echo chambers.

Racial stereotyping occurred when users made claims based on tropes steeped in racist charac-terizations of racially and ethnically marginalized groups' behaviors, beliefs, and identities. *Racial scapegoating* involved shifting the blame for racist actions, behaviors, policies, and institutions onto the marginalized racial group. Racial scapegoating also included blaming the activism or social justice organizing of marginalized racial groups on their outcomes. Lastly, racial scapegoating also involved attributing evidence of racial inequality to a source other than racism, or claiming that the evidence did not exist. In general, users that engage in racial scapegoating do so to dismiss the grievances of ethnically and racially marginalized people, often reinforcing perspectives steeped in the status quo or more egregious white supremacist perspectives.

Reverse racism charges generally involved alleging that members of a racial group typically considered victims of racism acted as perpetrators of racism, despite no evidence to support these claims. These accusations tended to take the form of a) "If a white person…" statements—implying that had a white person engaged in an action or behavior that a person of color had done, there would be more social consequences, or b) "POC-on-POC violence," which typically implied that Black people were dispro-portionately perpetrators of anti-Asian violence. In this sense, this framing was used to reverse the dynamics of white-on-Black acts of violence seen in incidents of vigilantism or police brutality. Black people were the perpetrators of racist hate crimes rather than victims.

Racism echo chambers involved dialogue where users rejected information of race or racism that disconfirmed their perspectives. Likewise, these echo chambers also emerged when users converged around information that appeared to support their perspectives on a marginalized group.

Bystander Intervention

Seventy-three percent of the interventions from online bystanders involved call-outs, directly naming the offending user racist or ignorant. Seventy-three percent of interventions drew on educational material or other forms of evidence, including news stories, statistical evidence, or historical evidence. Sixty percent of users that responded to these tweets engaged in insults or mocking. Finally, 13 percent of these interventions involved moderation in blocking the user or insisting on norms of behavior as the initiator of the thread.

On Reddit, 77 percent of these responses drew on education or evidence as a strategy of address-ing these forms of racist discourse. Forty-three percent involved some form of content moderation, whether by the subreddit's moderators, Reddit's spam filters, or the downvote mechanism used by other Redditors. Forty percent of interventions included call-outs. Finally, roughly nine percent of interventions involved insults or mocking.

Insults and Mocking

Users who resorted to **insults or mocking** used vulgar or impolite language when confronting per-spectives they deemed racist. These users criticized the intelligence of racist users or called them nonracial slurs. These users also often used crass humor or sarcasm to deride the perspectives of disagreeable users.

One way that insults and mocking was displayed as an intervention is exemplified in a thread of 187 tweets contributed to by 156 users.[67] The thread begins with a tweet from the official Twitter account for the far-right outlet *Breitbart News Network* in which they link to a May 2021 story on their website by David Ng, entitled "Disney Feels 'Hard Impact' of Disastrous Oscars Ratings as ABC Faces Advertiser Uncertainty."

This tweet and the interactions among users represent an echo chamber of opposition towards Disney's attempts to address diversity, likely kicked off by the *Breitbart News* account's initial tweet of "Go Woke, Go Broke" that accompanied the link to their story. Overwhelmingly, users in this thread emulate this initial tweet in terms of derision of Disney and Hollywood, associating these icons of the U.S. entertainment industry with everything from "Black Lives Matter" to critical race theory to communism. For example, one user quips that "Hollywood should have kept its leftist bullshit to itself," and "stuck to entertaining" rather than address political issues. Another user mentioned the Twitter account for Disney in a tweet where they asserted the company needed to "stop catering to the minority of these people" who were invested in "social justice woke BS" and instead focus on entertainment. Most of the users in the thread appear to conceal their identities through profile pictures and usernames other than those associated with their personal identity.

Despite the fact that the majority of the over 150 users in the thread appeared to agree with the *Breitbart News* account's initial sentiment, one person attempted to challenge the perspectives shared in the thread through mocking. Their initial response to this tweet called out the irony that a conservative news outlet would root for the downfall of a U.S. business before sarcastically calling them "such patriots." Several celebrants of the news of Disney's lower-than-expected Oscar ratings respond in various ways. Some users complained Disney itself was anti-American, so it only made sense not to support them. The user who called out *Breitbart News* appeared to adopt a strategy of displaying a one-line perspective in response to these, essentially pointing out how the desire to "own the libs" by wishing a U.S. corporation failed seemed contrary to the typically pro-capitalist conservative stance. However, they varied how they stated it.

Despite this user's commitment to call-outs and mocking as an intervention strategy, the users in the thread dismiss this user's argument and maintain their disfavorable attitudes toward Disney's attempts to address diversity issues. Potentially, this may in part reflect how insults and mocking as an intervention rarely address the issue of racism itself, but rather the perception that those who project prejudiced or biased attitudes towards diversity in the virtual public sphere lack wisdom or intellect.

Call-outs

Second, **call-outs** occurred when intervening users responded to the offending users by explicitly calling them racist, white supremacist, or accused them of supporting racist views or actions on behalf of others. These call-outs also included accusing the user of using or endorsing dog whistles—subtle remarks or framings of ideas that reinforce marginalization without overt reference to racial or ethnic identity. These users also accused offending users of lacking education or having ignorance about the topic of discussion.

One example of how a call-out intervention transpired occurred in response to a case of racial stereo-typing and reverse racism accusations on a post made on the subreddit r/RacismAgainstAsians. The description of this subreddit describes it as an outlet for "documenting racist outbursts and incidents

specifically targeted towards people of the Asian race." The title of the thread addressed a claim made by a Reddit user that accused the StopAsianHate movement of originating from the Chinese Communist Party.[68] The title also stated the Reddit user claimed that "Asians are beating and assaulting thousands of Black people every week without evidence." The body of the post features a screenshot of a post made on another subreddit, r/TrueUnpopularOpinion, displaying the text where the Reddit user addressed in the title makes these accusations.

Assuming that the person leveraging these allegations and misinformation against the StopAsianHate movement identified as Black, a commenter on the post responds by calling the user "ebola virus." The user then asserts that "we," referencing other Asians, would "just go around attacking you guys for a change," thus implying that Black people were perpetrators of violence against Asians. Their comment received four upvotes.

In response, a second Reddit user on the thread comments that "racism is still racism, friend," to call out the first user's comment as racist, regardless of if another person of color makes a racist comment about Asians. Their comment received two upvotes from other Reddit users. However, this attempted intervention falters as the first commenter ultimately defends their racism. They reply that they both know and do not care. Their statement invoked racism as they are "saying the same thing to them as they say to us" because "two can play the same game." This final comment receives six upvotes.

Education and Evidence

While social media users may explicitly call out racism in another user's post, the anonymity of a site like Reddit indicates that people feel comfortable perpetuating preconceptions about racial minorities, particularly when their views are reinforced via mechanisms such as upvotes. Users that drew on *education and evidence* would often draw on historical, sociological, or statistical sources to challenge the perspectives of disagreeable users. While most users refuted points directly in their posts, others would link to an external website that contained supportive information, including news stories, academic journal articles, and reports from government, policy, and research or academic institutions. Lastly, educational interventions included breaking down or analyzing the thought process or arguments presented by the user they disagree with, pointing out flaws in the logic of that user's argument.

A discussion between two users on Twitter throughout nine tweets demonstrates how providing evidence or education may function as an effective intervention at times, in this case against racial scapegoating.[69] While all Reddit users are required to use their proprietary "Snoo" avatar as their profile picture, Twitter users have a range of profile image standards. While some users show their own image, others maintain their anonymity by utilizing profile photos of inanimate things, animals, celebrities, or other social media users. In this conversation, one user's profile photo seems to be a selfie, but it is unclear if it is really the person. The other user maintains anonymity via the usage of a black-and-white picture of a person wearing what seems to be a cowboy hat.

The Twitter thread begins when the first user links to an article published in *Forbes* titled "Here's What The Racial Wealth Gap In America Looks Like Today" by Sarah Hansen.[70] In addition to posting this link, the user critiques corporations for showing solidarity without material support. Posted less than two weeks after the death of George Floyd, the tweet references the flurry of corporate statements released by major corporations in the days after his death, some of which included promises of financial support for the #BlackLivesMatter organization and other advocacy groups and movements that address issues of racial justice.

A second Twitter user replies to this post and invokes racial scapegoating as they contemplate that the racial wealth gap "seems more like the result of individual decisions" rather than systemic causes as argued in Hansen's article, which cites a Brooking Institute blog post titled "Examining the Black-white wealth gap" as evidence. In response, the first user dismisses this user's argument by stating that they don't "decide to racially discriminate my way out of" receiving a loan from the bank or equal interest rates to white families of equal economic status.

This initial response seems not to deter the second Twitter user, who responds that "there are a lot of factors that go into" the racial wealth gap. Their reply then prompts the user who initially posted the *Forbes* article that prompted this dialogue to respond that much of the racial wealth gap results from structural causes. They then urge their respondent to understand how banking institutions historically redlined Black neighborhoods from home ownership. To end their tweet, they ask their conversant if they knew about the finding from a study that found that a college-educated Black man had a similar likelihood of getting a job as a white man with a high school diploma.

Moderation

Fourth, ***moderation***, which occurred almost exclusively on Reddit, involved users using the platform's features to minimize or exclude discourse that contained ignorant or stereotypical perspectives on race and racism. This approach included when users downvoted a particular sentiment or moderators of a subreddit removing content at their discretion. At times, on Reddit, moderation would include a bot designed to monitor the content posted to a particular subreddit. On the rare occasion that moderation did occur on Twitter, it involved either blocking users or directly regulating the tone of the conversation as the original poster.

Content moderation as a form of bystander intervention figures more prominently on Reddit. The features of the multi-forum website enable users to create sub-forums where they can then assign other users to the role of moderator. Additionally, some users include a moderator bot on their subreddits. These moderator bots rely on algorithms to enact specific actions on posts. For example, some moderator bots automatically reply to new posts that present users with a list of subreddit rules. Though sometimes a reply to the initial Reddit post indicated the specific reasons why a moderator had taken a particular action, for the most part, when users violated subreddit norms, their comment gets deleted, indicated typically by the statement "comment removed by a moderator." While content moderation

typically involves the decision-making of subreddit moderators or bots, the downvote feature of Reddit enables other users in a subreddit to rank posts with certain perspectives lower in a thread. For these reasons, content moderation on Reddit generally also occurred alongside other forms of interventions.

One way that content moderation played out on Reddit occurred on the subreddit r/WayoftheBern, which, despite the reference to Bernie Sanders, offers a space to discuss "politics along a top/bottom divide" according to the subreddit's description. The title of the post begins "Blue lives don't exist…" and links a tweet by a journalist-activist that also serves as the remainder of the title that explains how the Blue Lives Matter movement differs from Black Lives Matter in that people are not born cops.

A thread of 62 comments and 25 users, most users in this subreddit agree with this initial claim, though several users argue against it, often invoking the controlling image of Black people as criminals or the notion of policing as a dangerous job to justify their perspective.

For example, one user quotes part of the post title that refers to "trigger-happy officers…abusing their power and killing people, especially Black people." In response, the user links to a graphic on Statista and a data table from the Uniform Crime Reporting (UCR) program at the Federal Bureau of Investigation (FBI). The Statista graphic is titled "Number of people shot to death by the police in the United States from 2017 to 2021, by race" and the UCR link leads to the "Expanded Homicide Data Table 6," on "Murder" by "Race, Sex, and Ethnicity of Victim by Race, Sex, and Ethnicity of Offender, 2018." Using this information, the user asserts that fewer Black people were killed by law enforcement in 2019 than were killed by other Black people in 2018.

Several users respond to this manipulation of information by educating the user on their logical fallacy. For example, a user responds that the proportion of Black people far exceeds the proportion of law enforcement officers. Additionally, they point out how interracial killings often lead to incarceration, unlike in the case of police killings. This user, whose comment receives seven upvotes, also calls out the commenter's intentions by writing they were certain the user knew these facts and intended only to respond to the thread on the specific subreddit "in bad faith to 'own the libs' with out-of-context statistics."

The user, however, dismisses this rebuttal by stating that "a cop is much more likely to be involved in a death than a normal person," particularly if they are "patrolling in a heavy crime area," invoking an association of Black people and predominantly Black neighborhoods with criminality. This user then asserts that statistics do not reflect "that cops are going around killing black people." Their response receives one downvote. In addition to this downvote, the user receives yet another response from a third user that functions as an educational intervention. This user, echoing the strategies of the second user to intervene during this exchange, claims that data from the Occupational Safety and Health Administration (OSHA) appears to indicate that being a police officer is a less risky job than a pizza delivery driver when it comes to the possibility of experiencing a fatal shooting while at work.

When another user makes a similar argument on a separate thread within the same post, asking "what group of people commit more violent crimes per capita," they receive six downvotes. Several users respond with educational information that explains the links among poverty, crime, and race in the United States. However, this user seems unswayed, responding with yet another question of "are we just gonna pretend that affirmative action never happened," a comment that received two downvotes.

In this same post, a moderator of the r/WayoftheBern subreddit has removed a comment in a separate thread. However, responses from other users, some drawing on insults, indicate the user had engaged in racist discourse based on how other users replied, presumably before a moderator had intervened. These users name the now-removed commenter a "blatant troll account" and "disgusting racist creep." Overall, the exchange on this subreddit exemplifies how Reddit users draw on the features of the social media platform, namely the ranking system and moderation allowance, to penalize those who invoke various forms of racial scapegoating or stereotyping, provided anti-racist perspectives characterize the norms of the subreddit.

Broader Impacts

Our analysis identified four types of racist discourses that occur during social media discussions: stereotyping, scapegoating, accusations of reverse racism, and echo chambers. We also found four types of bystander intervention strategies, which include: attempts to educate or provide evidence, content moderation, call-outs, insults or mocking. However, there was much variation in the deployment of these intervention strategies. On Twitter, where people can be more anonymous and may not see all the social network branches, people were less likely to intervene and more likely to use antisocial intervention strategies. On Reddit, where the social network is more closed and interactional (e.g., repeated interactions, ratings up and down, and people can see most of the information), people were more likely to engage in prosocial bystander intervention.

Some social media sites have aimed to address how intergroup dynamics matter during online discussion by creating small groups and online communities. Our research suggests the importance of repeated interaction to create accountability rather than the continuous ability to be anonymous. Anonymity seems to help drive contagion and reduces the employment of content moderation as a bystander intervention strategy. On Reddit, like in groups and communities on other platforms, someone or even a group of people are often responsible for content moderation and aim to actively intervene.

Accordingly, our report provides theoretical, methodological, and policy-oriented contributions. First, our research contributes in a meaningful way to the growing academic literature on this topic, bolstering literature on bystander intervention and responses to systemic racism and racist language. Second, our research provides an analytical blueprint for continuing to investigate these patterns. Our thematic patterns establish future possibilities for scholars to examine. Finally, our work may have immediate impacts on youth, who are often victimized online, and the many experts and parents aiming to develop effective strategies to engage in bystander intervention for healthy mental and emotional health development.

About the Authors

Rashawn Ray is a Senior Fellow in Governance Studies at The Brookings Institution. He is also a professor of sociology and executive director of the Lab for Applied Social Science Research (LASSR) at the University of Maryland, College Park. Ray is one of the co-editors of *Contexts Magazine: Sociology for the Public*. His research addresses the mechanisms that manufacture and maintain racial and social inequality with a particular focus on police-civilian relations, racial-health disparities, and voting. His work also speaks to ways that inequality may be attenuated through racial uplift activism and social policy.

Melissa Brown is a scholar of Black feminist thought, digital sociology, social movements, and sexual politics. She has a chapter titled "For a Black feminist digital sociology" in the recently published edited volume published by Routledge and for which she serves as the *Black Feminist Sociology* digital editor of the companion website. Prior to arriving at Santa Clara University, Melissa was a postdoctoral research fellow at the Clayman Institute for Gender Research at Stanford University. Previously, she has published her research on anti-racist and Black feminist social media activism in the journals *Ethnic and Racial Studies* and *Sociology Compass*. She currently maintains Blackfeminisms.com as a public scholarly project that disseminates academic research and theory by and about Black women across the African Diaspora.

Ed Summers is research faculty at the Maryland Institute for Technology in the Humanities at the University of Maryland. His research focuses on the role of the web and social media platforms in social justice, community building, and memory work. He holds a PhD in Information Studies from the University of Maryland.

Samantha Elizondo is a research assistant in the Race, Prosperity, and Inclusion Initiative at The Brookings Institution. She works on a series of projects related to inclusive economic opportunities, racial inequality, voting, health, and criminal justice reform. Elizondo holds a bachelor's degree in Public Policy from the University of Chicago.

Connor Powelson, M.A., is a doctoral candidate in the Department of Sociology and Lead Research Assistant for the Lab for Applied Social Science Research (LASSR) at the University of Maryland, College Park. His research and teaching interests include sociological social psychology, emotions, race/gender/class, social inequality, and mixed methodology. Connor's current research addresses the mechanisms that influence biased misconduct, accountability, and wellbeing of organization members, particularly among police. His work applies affect control theory to evaluate subcultural sentiment expression among groups of varying attitude, race, and gender composition. He is currently involved in an ongoing evaluation of police implicit bias using virtual reality and procedurally generated language technology at LASSR. Connor is also affiliated with the Group Processes Laboratory at the University of Maryland.

Endnotes

1   Brown, Melissa, Rashawn Ray, Edward Summers, and Neil Fraistat. 2017. "#SayHerName: A Case Study of Intersectional Social Media Activism." *Ethnic and Racial Studies* 40(11): 1831–46. Retrieved from https://doi.org/10.1080/01419870.2017.1334934.

Ray, Rashawn, Melissa Brown, Neil Fraistat, and Edward Summers. 2017. "Ferguson and the Death of Michael Brown on Twitter: #BlackLivesMatter, #TCOT, and the Evolution of Collective Identities." *Ethnic and Racial Studies* 40(11): 1797–1813. Retrieved from https://doi.org/10.1080/01419870.2017.1335422.

2   Low, Sabina, and Dorothy Espelage. 2013. "Differentiating Cyber Bullying Perpetration From Non-Physical Bullying: Commonalities Across Race, Individual, and Family Predictors." *Psychology of Violence* 3: 39–52. Retrieved from https://www.researchgate.net/publication/235220407_Differentiating_Cyber_Bullying_Perpetration_From_Non-Physical_Bullying_Commonalities_Across_Race_Individual_and_Family_Predictors.

Seals, Dorothy, and Jerry Young. 2003. "Bullying and Victimization: Prevalence and Relationship to Gender, Grade Level, Ethnicity, Self-Esteem, and Depression." *Adolescence* 38(152): 735–47.

3   Anderson, Monica, and Jingjing Jiang. 2018. "Teens, Social Media & Technology 2018." *Pew Research Center: Internet, Science & Tech*. Retrieved from http://publicservicesalliance.org/wp-content/uploads/2018/06/Teens-Social-Media-Technology-2018-PEW.pdf.

4   Davis, Jenny L. 2014. "Triangulating the Self: Identity Processes in a Connected Era" *Symbolic Interaction* 37(4): 500–523. Retrieved from https://doi.org/10.1002/symb.123.

Gonzales, Amy L. and Jeffrey T. Hancock. 2011. "Mirror, Mirror on My Facebook Wall: Effects of Exposure to Facebook on Self-Esteem." *Cyberpsychology, Behavior, and Social Networking* 14(1): 79–83. Retrieved from https://doi.org/10.1089/cyber.2009.0411.

Gonzales, Amy L. 2014. "Text-Based Communication Influences Self-Esteem More than Face-to-Face or Cellphone Communication." *Computers in Human Behavior* 39:197–203. Retrieved from https://doi.org/10.1016/j.chb.2014.07.026.

McDonald, Daniel G., and Hyeok Kim. 2001. "When I Die, I Feel Small: Electronic Game Characters and the Social Self." *Journal of Broadcasting & Electronic Media* 45(2): 241–58.

5   Marwick, Alice E., and danah boyd. 2011. "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience." *New Media & Society* 13(1): 114–33. Retrieved from http://www.tiara.org/wp-content/uploads/2018/05/Marwick_boyd_TweetHonestly.pdf.

Phillips, Whitney. 2015. *This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture*. MIT Press.

Ronson, J. 2016. *So You've Been Publicly Shamed.* Riverhead Books.

6   Anderson, Monica. 2018. "A Majority of Teens Have Experienced Some Form of Cyberbullying." *Pew Research Center: Internet, Science & Tech*. Retrieved from https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2018/09/PI_2018.09.27_teens-and-cyberbullying_FINAL.pdf.

7   Gladden, Matthew R., Alana M. Vivolo-Kantor, Merle E. Hamburger and Corey D. Lumpkin. 2014. "Bullying Surveillance Among Youths: Uniform Definitions for Public Health and Recommended Data Elements," Version 1.0. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention and US Department of Education. Retrieved from https://www.cdc.gov/violenceprevention/pdf/bullying-definitions-final-a.pdf.

8   Mueller, Anna S., Wesley James, Seth Abrutyn, and Martin L. Levin. 2015. "Suicide Ideation and Bullying among US Adolescents: Examining the Intersections of Sexual Orientation, Gender, and Race/Ethnicity." *American Journal of Public Health* 105(5): 980–85.

Patchin, Justin W., and Sameer Hinduja. 2010. "Cyberbullying and Self-Esteem." *Journal of School Health* 80(12): 614–21. Retrieved from https://onlinelibrary.wiley.com/doi/10.1111/j.1746-1561.2010.00548.x.

9   United States Census Bureau. 2018. "The Nation's Older Population Is Still Growing, Census Bureau Reports." Census Bureau QuickFacts, United States Census Bureau. Retrieved from https://www.census.gov/newsroom/press-releases/2017/cb17-100.html.

10   Katz, Jennifer, Christine Merrilees, Jill C. Hoxmeier and Marisa Motisi. 2017. "White female bystanders' responses to a black women at risk for incapacitated sexual assault." *Psychology of Women Quarterly* 41(2): 273–285. Retrieved from https://doi.org/10.1177/0361684316689367

11   Franklin, Cortney A., and Alondra D. Garza. 2021. "Sexual Assault Disclosure: The Effect of Victim Race and Perpetrator Type on Empathy, Culpability, and Service Referral for Survivors in a Hypothetical Scenario." *Journal of Interpersonal Violence* 36(5–6): 2327–52. https://doi.org/10.1177/0886260518759656 (October 8, 2021).

12   Mueller, Anna S., Wesley James, Seth Abrutyn, and Martin L. Levin. 2015. "Suicide Ideation and Bullying among US Adolescents: Examining the Intersections of Sexual Orientation, Gender, and Race/Ethnicity." *American Journal of Public Health* 105(5): 980–85.

Popp, Ann Marie, Anthony A. Peguero, Kristin R. Day, and Lindsay L. Kahle. 2014. "Gender, Bullying Victimization, and Education." Violence and Victims 29(5): 843–56. Retrieved from http://www.scopus.com/inward/record.url?scp=84929030895&partnerID=8YFLogxK.

13   Christensen, Candance M. and Richard J. Harris. 2019. "Correlates of Bystander Readiness to Help Among a Diverse College Student Population: An Intersectional Perspective." *Research in Higher Education* 60(2): 1195–1226. Retrieved from https://link.springer.com/article/10.1007/s11162-018-09544-6.

14   Bayraktar, Faith et al. 2015. "Cyberbullying: The Discriminant Factors Among Cyberbullies, Cybervictims, and Cyberbully-Victims in a Czech Adolescent Sample." *Journal of Interpersonal Violence* 30(18), 3192–3216. Retrieved from https://doi.org/10.1177/0886260514555006.

Brewster, Mary, and Jane M. Tucker. 2016. "Understanding Bystander Behavior: The Influence of and Interaction between Bystander Characteristics and Situational Factors." *Victims & Offenders* 11(3): 455–81.

Hoxmeier, Jill C., Alan C. Acock and Brian R. Flay. 2017. "Students as prosocial bystanders to sexual assault: Demographic correlates of intervention norms, intentions, and missed opportunities." *Journal of Interpersonal Violence* 35(3–4): 731–754. Retrieved from https://www.researchgate.net/publication/313126709_Students_as_Prosocial_Bystanders_to_Sexual_Assault_Demographic_Correlates_of_Intervention_Norms_Intentions_and_Missed_Opportunities.

15   Olenik-Shemesh, Dorit, Tali Heiman, and Sigal Eden. 2012. "Cyberbullying Victimisation in Adolescence: Relationships with Loneliness and Depressive Mood." *Emotional & Behavioural Difficulties* 17(3–4): 361–74.

Wang, Jing, Tonja R. Nansel, and Ronald J. Iannotti. 2011. "Cyber and Traditional Bullying: Differential Association with Depression." *Journal of Adolescent Health* 48: 415–417. Retrieved from https://www.sciencedirect.com/science/article/pii/S1054139X10003435?casa_token=N-km8gHnOwcAAAAA:frQ37a3v5vbJ3_q4OAlq4YcpkhpF-7Bqit22_AEGeXUMjveAXizptzR45z_VLgOEgcFBfbCCH.

Perren, Sonja, Julian Dooley, Thérèse Shaw, and Donna Cross. 2010. "Bullying in School and Cyberspace: Associations with Depressive Symptoms in Swiss and Australian Adolescents." *Child and Adolescent Psychiatry and Mental Health* 4(1): 28. Retrieved from https://doi.org/10.1186/1753-2000-4-28.

Schneider, Shari Kessel, Lydia O'Donnell, Ann Stueve, and Robert W. S. Coulter. 2012. "Cyberbullying, School Bullying, and Psychological Distress: A Regional Census of High School Students." *American Journal of Public Health* 102(1): 171–77. Retrieved from http://www.promoteprevent.org/sites/www.promoteprevent.org/files/resources/Cyberbullying%20school%20bullying%20and%20psychological%20distress.pdf.

16   Messias, Erick, Kristi Kindrick, and Juan Castro. 2014. "School Bullying, Cyberbullying, or Both: Correlates of Teen Suicidality in the 2011 CDC Youth Risk Behavior Survey." Comprehensive psychiatry 55(5): 1063–1068.

17   Stanley-Becker, Isaac. 2018. A 9-year-old's family says she committed suicide after racist bullying, including taunts of 'kill yourself.' *Washington Post.* Retrieved from http://www.washingtonpost.com/nation/2018/12/12/year-olds-family-says-she-hung-herself-after-racist-bullying-including-taunts-kill-yourself/.

18   Sue, Derald Wing et al. 2019. "Disarming Racial Microaggressions: Microintervention Strategies for Targets, White Allies, and Bystanders." *American Psychologist* 74(1): 128–42.

19   Nelson, Jacqueline K., Kevin M. Dunn, and Yin Paradies. 2011. "Bystander Anti-Racism: A Review of the Literature: Bystander Anti-Racism." *Analyses of Social Issues and Public Policy* 11(1): 263–84. Retrieved from https://onlinelibrary.wiley.com/doi/10.1111/j.1530-2415.2011.01274.x.

20   McMahon, Sarah, and Victoria L. Banyard. 2012. "When Can I Help? A Conceptual Framework for the Prevention of Sexual Violence through Bystander Intervention." *Trauma, Violence & Abuse* 13(1): 3–14.

McMahon, Sarah, Jessica Burnham, and Victoria L. Banyard. 2020. "Bystander Intervention as a Prevention Strategy for Campus Sexual Violence: Perceptions of Historically Minoritized College Students." *Prevention Science: The Official Journal of the Society for Prevention Research* 21(6): 795–806.

21   Kazerooni, Franccesca, Samuel Hardman Taylor, Natalya N. Bazarova, and Janis Whitlock. 2018. "Cyberbullying Bystander Intervention: The Number of Offenders and Retweeting Predict Likelihood of Helping a Cyberbullying Victim." *Journal of Computer-Mediated Communication* 23(3): 146–162. Retrieved from https://doi.org/10.1093/jcmc/zmy005.

22   Leonhard, Larissa, Christina Rueß, Magdalena Obermaier, and Carsten Reinemann. 2018. "Perceiving Threat and Feeling Responsible. How Severity of Hate Speech, Number of Bystanders, and Prior Reactions of Others Affect Bystanders' Intention to Counterargue against Hate Speech on Facebook." *Studies in Communication and Media* 7: 555–79.

23   Compton, D'Lane, Tony P. Love, and Jane Sell. 2012. "Developing and Assessing Intercoder Reliability in Studies of Group Interaction." *Sociological Methodology* 42(1): 348–364. Retrieved from https://doi.org/10.1177/0081175012444860.

Mulvey, K. L., Palmer, S. B., and Abrams, D. 2016. "Race-Based Humor and Peer Group Dynamics in Adolescence: Bystander Intervention and Social Exclusion." *Child Development* 87(5): 1379–91.

24   Duffy, Amanda L. and Drew Nesdale. 2009. "Peer Groups, Social Identity, and Children's Bullying Behavior." *Social Development* 18(1): 121–139. Retrieved from https://doi.org/10.1111/j.1467-9507.2008.00484.x.

Pöyhönen, Virpi, Jaana Juvonen, and Christina Salmivalli. 2012. "Standing Up for the Victim, Siding with the Bully or Standing by? Bystander Responses in Bullying Situations." *Social Development* 21(4): 722–41. Retrieved from https://www.academia.edu/11857279/Standing_up_for_the_victim_siding_with_the_bully_or_standing_by_Bystander_responses_in_bullying_situations.

Wang, Sai. 2021. "Standing up or standing by: Bystander intervention in cyberbullying on social media." *New Media & Society*, 23(6): 1379–1397. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/1461444820902541.

25   Bonanno, Rina A., and Shelley Hymel. 2013. "Cyber Bullying and Internalizing Difficulties: Above and Beyond the Impact of Traditional Forms of Bullying." *Journal of Youth and Adolescence* 42(5): 685–697. Retrieved from https://link.springer.com/content/pdf/10.1007/s10964-013-9937-1.pdf.

Raskauskas, Juliana. 2010. "Text-Bullying: Associations with Traditional Bullying and Depression among New Zealand Adolescents." *Journal of School Violence* 9(1): 74–97.

Rigby, Ken, and Peter K. Smith. 2011. "Is School Bullying Really on the Rise?" *Social Psychology of Education* 14(4): 441–55. Retrieved from https://doi.org/10.1007/s11218-011-9158-y.

Slonje, Robert, and Peter Smith. 2008. "Cyberbullying: Another Main Type of Bullying?" *Scandinavian Journal of Psychology* 49: 147–54. Retrieved from https://www.academia.edu/2384042/Cyberbullying_Another_main_type_of_bullying.

26   Lowry, Paul Benjamin, Jun Zhang, Chuang Wang, and Mikko Siponen. 2016. "Why Do Adults Engage in Cyberbullying on Social Media? An Integration of Online Disinhibition and Deindividuation Effects with the Social Structure and Social Learning Model." *Information Systems Research* 27(4): 962–86. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2830405.

27   Calvete, Esther, Izaskun Orue, Ana Estevez, Lourdes Villardon and Patricia Padilla. 2010. "Cyberbullying in Adolescents: Modalities and Aggressors' Profile." *Computers in Human Behavior* 26(5): 1128–1135. Retrieved from https://doi.org/10.1016/j.chb.2010.03.017.

Slonje, Robert, and Peter Smith. 2008. "Cyberbullying: Another Main Type of Bullying?" *Scandinavian Journal of Psychology* 49: 147–54. Retrieved from https://www.academia.edu/2384042/Cyberbullying_Another_main_type_of_bullying.

28   Cassidy, Wanda, Chantal Faucher, and Margaret Jackson. 2013. "Cyberbullying among Youth: A Comprehensive Review of Current International Research and Its Implications and Application to Policy and Practice." *School Psychology International* 34(6): 575–612. Retrieved from https://doi.org/10.1177/0143034313479697.

Gini, Gianluca, Paolo Albeiro, Beatrice Benelli and Gianmarco Altoe. 2008. "Determinants of Adolescents' Active Defending and Passive Bystanding Behavior in Bullying." *Journal of Adolescence* 31(1): 93–105. https://doi.org/10.1016/j.adolescence.2007.05.002.

29   Hinduja, Sameer and Justin W. Patchin. 2007. "Offline Consequences of Online Victimization." *Journal of School Violence* 6(3): 89–112. Retrieved from https://doi.org/10.1300/J202v06n03_06.

Hoffman, Jan. "Online Bullies Pull Schools Into the Fray." *New York Times*, 28 June 2010. https://www.nytimes.com/2010/06/28/style/28bully.html.

Horner, Stacy, Yvonne Asher and Gary D. Fireman. 2015. "The Impact and Response to Electronic Bullying and Traditional Bullying Among Adolescents." *Computers in Human Behavior* 49: 288–295. Retrieved from https://doi.org/10.1016/j.chb.2015.03.007.

30   Murrell, A. 2020. "Why Someone Did Not Stop Them? Aversive Racism and the Responsibility of Bystanders."

31   Doré, Bruce, Leonard Ort, Ofir Braverman, and Kevin N. Ochsner. 2015. "Sadness Shifts to Anxiety over Time and Distance from the National Tragedy in Newtown, Connecticut." *Psychological Science* 26(4): 363–73.

Jones, Nickolas, Sean Wojcik, Josiah Sweeting, and Roxane Silver. 2016. "Tweeting Negative Emotion: An Investigation of Twitter Data in the Aftermath of Violence on College Campuses." *Psychological Methods* 21: 526–41.

32   Hill, Marc. 2018. "Thank You, Black Twitter': State Violence, Digital Counterpublics, and Pedagogies of Resistance." *Urban Education* 52(2): 286–302. Retrieved from https://doi.org/10.1177/0042085917747124.

Ince, Jelani, Fabio Rojas and Clayton A. Davis. 2017. "The social media response to Black Lives Matter: how Twitter users interact with Black Lives Matter through hashtag use." *Ethnic and Racial Studies* 40(11): 1814–1830. Retrieved from https://doi.org/10.1080/01419870.2017.1334931.

33   Reed, Karen P., R. Lyle Cooper, William R. Nugent, and Kathryn Russell. 2016. "Cyberbullying: A Literature Review of Its Relationship to Adolescent Depression and Current Intervention Strategies." Journal of Human Behavior in the Social Environment 26(1): 37–45. Retrieved from https://doi.org/10.1080/10911359.2015.1059165.

34   Beran, Tanya N., and Qing Li. 2007. "The Relationship Between Cyberbullying and School Bullying." *Journal of Student Wellbeing* 1(2): 15–33. Retrieved from https://doi.org/10.21913/JSW.v1i2.172.

Cassidy, Wanda, Margaret Jackson, and Karen N. Brown. 2009. "Sticks and Stones Can Break My Bones, But How Can Pixels Hurt Me?: Students' Experiences with Cyber-Bullying." *School Psychology International* 30(4): 383–402. Retrieved from https://www.academia.edu/21601182/Sticks_and_Stones_Can_Break_My_Bones_But_How_Can_Pixels_Hurt_Me_Students_Experiences_with_Cyber_Bullying.

Juvonen, Jaana and Elisheva F. Gross. 2008. "Extending the School Grounds—Bullying Experiences in Cyberspace." *Journal of School Health* 78(9): 496–505. Retrieved from 10.1111/j.1746-1561.2008.00335.x

35   Slonje, Robert, Peter K. Smith, and Ann Frisén. 2013. "The Nature of Cyberbullying, and Strategies for Prevention." *Computers in Human Behavior* 29(1): 26–32.

36   Cash, Scottye. J., and Jeffrey A. Bridge. 2009. "Epidemiology of Youth Suicide and Suicidal Behavior." *Current Opinion in Pediatrics* 21(5): 613. Retrieved from 10.1097/MOP.0b013e32833063e1.

Hinduja, Sameer and Justin W. Patchin. 2010. "Bullying, Cyberbullying, and Suicide." *Archives of Suicide Research* 14(3): 206–221. Retrieved from https://doi.org/10.1080/13811118.2010.494133.

Klomek, Anat B., Frank Marrocco, Marjorie Kleinman, Irvin Schonfeld, Madelyn Gould. 2007. "Bullying, Depression, and Suicidality in Adolescents." *Journal of the American Academy of Child and Adolescent Psychiatry* 46(1): 40–49.

37   Cash, Scottye. J., and Jeffrey A. Bridge. 2009. "Epidemiology of Youth Suicide and Suicidal Behavior." *Current Opinion in Pediatrics* 21(5): 613. Retrieved from 10.1097/MOP.0b013e32833063e1.

38   Olweus, Dan. 2012. "Cyberbullying: An Overrated Phenomenon?" *European Journal of Developmental Psychology* 9(5): 520–538. Retrieved from https://doi.org/10.1080/17405629.2012.682358.

Pieschl, Stephanie, Torsten Porsch, Tobias Kahl, and Rahel Klockenbusch. 2013. "Relevant Dimensions of Cyberbullying—Results from Two Experimental Studies." *Journal of Applied Developmental Psychology* 34(5): 241–52.

Strauss, A., and Corbin, J. M. 1997. *Grounded Theory in Practice.* Sage.

39   Lentin, Alana. 2020. *Why Race Still Matters*. Polity Press.

40   Schweinberger, Martin, Michael Haugh, and Sam Hames. 2021. "Analysing Discourse around COVID-19 in the Australian Twittersphere: A Real-Time Corpus-Based Analysis." *Big Data & Society* 8(1): 20539517211021436. Retrieved from https://journals.sagepub.com/doi/full/10.1177/20539517211021437.

41   Vaux, Alan. 1988. *Social Support: Theory, Research, and Intervention.* Praeger publishers.

42   Cullen, Francis T, James D. Unnever, Jennifer L. Hartman, Michael G. Turner and Robert Agnew. 2008. "Gender, Bullying Victimization, and Juvenile Delinquency: A Test of General Strain Theory." *Victims and Offenders* 3(4): 346–364. Retrieved from https://doi.org/10.1080/15564880802338468.

43   Uchino, Bert N., John T. Cacioppo, and Janice Kiecolt-Glaser. 1996. "The Relationship Between Social Support and Physiological Processes: A Review with Emphasis on Underlying Mechanisms and Implications for Health." *Psychological Bulletin* 119(3): 488. Retrieved from https://www.researchgate.net/publication/14532031_The_Relationship_Between_Social_Support_and_Physiological_Processes_A_Review_with_Emphasis_on_Underlying_Mechanisms_and_Implications_for_Health.

44   Cohen, Sheldon and Garth McKay. 1984. "Social Support, Stress and the Buffering Hypothesis: A Theoretical Analysis." *Handbook of Psychology and Health* 4: 253–267.

45   Padgett, Sharon, and Charles E. Notar. 2013. "Bystanders Are the Key to Stopping Bullying." *Universal Journal of Educational Research* 1(2): 33–41. Retrieved from https://files.eric.ed.gov/fulltext/EJ1053992.pdf.

Menesini, Ersilia, Annalaura Nocentini, and Benedetta Palladino. 2012. "Empowering Students Against Bullying and Cyberbullying: Evaluation of an Italian Peer–Led Model." *International Journal of Conflict and Violence* 6: 313–20.

46   Bell, Christopher D., Katherine A. Raczynski, and Arthur M. Horne. 2010. "Bully Busters Abbreviated: Evaluation of a Group–Based Bully Intervention and Prevention Program." *Group Dynamics: Theory, Research, and Practice* 14(3): 257–267. Retrieved from https://www.researchgate.net/publication/232561349_Bully_Busters_Abbreviated_Evaluation_of_a_Group-Based_Bully_Intervention_and_Prevention_Program.

47   Menesini, Ersilia, Annalaura Nocentini, and Benedetta Palladino. 2012. "Empowering Students Against Bullying and Cyberbullying: Evaluation of an Italian Peer-Led Model." *International Journal of Conflict and Violence* 6: 313–20.

48   Peter, Levy et al. 1972. "Bystander Effect in a Demand-without-Threat Situation." *Journal of Personality and Social Psychology* 24(2): 166–71.

49   Padgett, Sharon, and Charles E. Notar. 2013. "Bystanders Are the Key to Stopping Bullying." *Universal Journal of Educational Research* 1(2): 33–41. Retrieved from https://files.eric.ed.gov/fulltext/EJ1053992.pdf.

50 Pascoe, C. J. 2011. *Dude, You're a Fag: Masculinity and Sexuality in High School*. University of California Press.

Pascoe, C. J. 2013. "Notes on a Sociology of Bullying: Young Men's Homophobia as Gender Socialization." *QED: A Journal in GLBTQ Worldmaking*, 87–104. Retrieved from https://www.jstor.org/stable/10.14321/qed.0087.

Pascoe, C. J. 2014. "Bullying as Social Inequality." *THE ENEMY/Issue One*, 17.

51 Brown, Melissa, Rashawn Ray, Edward Summers, and Neil Fraistat. 2017. "#SayHerName: A Case Study of Intersectional Social Media Activism." Ethnic and Racial Studies 40(11): 1831–46. Retrieved from https://doi.org/10.1080/01419870.2017.1334934.

Ray, Rashawn, Melissa Brown, Neil Fraistat, and Edward Summers. 2017. "Ferguson and the Death of Michael Brown on Twitter: #BlackLivesMatter, #TCOT, and the Evolution of Collective Identities." Ethnic and Racial Studies 40(11): 1797–1813. Retrieved from https://doi.org/10.1080/01419870.2017.1335422.

52 boyd, danah m., and Nicole B. Ellison. 2007. "Social Network Sites: Definition, History, and Scholarship." *Journal of Computer-Mediated Communication* 13(1): 210–30. Retrieved from https://doi.org/10.1111/j.1083-6101.2007.00393.x.

Caliandro, Alessandro. 2018. "Digital Methods for Ethnography: Analytical Concepts for Ethnographers Exploring Social Media Environments." *Journal of Contemporary Ethnography* 47(5): 551–78. Retrieved from https://researchportal.bath.ac.uk/en/publications/digital-methods-for-ethnography-analytical-concepts-for-ethnograp.

53 Jabeur, Nafaa, Sherali Zeadally and Biju Sayed. 2013. "Mobile social networking applications." *Communications of the ACM* 56(3): 71–79. Retrieved from 10.1145/2428556.2428573.

54 Alhajj, Reda, and Jon Rokne (Eds.). 2014. *Encyclopedia of Social Network Analysis and Mining*. New York: Springer. Retrieved from https://www.springer.com/gp/book/9781461461692.

55 Murthy, D. 2008. "Digital Ethnography." *Sociology* 42(5):837–55.

Varis, Piia. 2014. "Digital Ethnography." *Tilburg Papers in Culture Studies*. Retrieved from https://pure.uvt.nl/ws/portalfiles/portal/30479758/TPCS_104_Varis.pdf.

56 Summers, Edward. et al. 2021. *DocNow/Twarc.* Zenodo. Retrieved from https://zenodo.org/record/5547755 (October 4, 2021).

57 Brown, Melissa, Rashawn Ray, Ed Summers, and Neil Fraistat. 2017. "#SayHerName: A Case Study of Intersectional Social Media Activism." *Ethnic and Racial Studies* 40(11): 1831–46. Retrieved from https://doi.org/10.1080/01419870.2017.1334934.

Ray, Rashawn, Melissa Brown, Neil Fraistat, and Edward Summers. 2017. "Ferguson and the Death of Michael Brown on Twitter: #BlackLivesMatter, #TCOT, and the Evolution of Collective Identities." *Ethnic and Racial Studies* 40(11): 1797–1813. Retrieved from https://doi.org/10.1080/01419870.2017.1335422.

58 Summers, Edward, Melissa Brown, Samantha Elizondo, Rashawn Ray. 2021. *edsu/bystander-action: v1.0.0.* Jupytr Notebook. Retrieved from https://zenodo.org/record/5520036#.YWBjt5rMI2w.

Summers, Edward et al. 2018. *Twarc*. Documenting the Now. Python. Retrieved from https://github.com/DocNow/twarc.

59 Baumgartner, Jason, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The PushShift Reddit Dataset. *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media* 14: 830–39. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/7347.

60 Summers, Edward. 2021. "Bystander Intervention and Race Social Media Conversations." Retrieved from https://edsu.github.io/bystander-action/convs/site/#twitter::racism::tweets_racism_convs_30_1394664103256469510.json.

61 Hutto, C. J. and Eric Gilbert. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the International AAAI Conference on Web and Social Media* 8(1): 216–225. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/14550.

62 Summers, Edward. et al. 2021. *DocNow/Twarc.* Zenodo. Retrieved from https://zenodo.org/record/5547755.

63 Jockers, Matthew. "Introduction to the Syuzhet Package." Microsoft R Application Network. Retrieved September 22, 2021 https://mran.microsoft.com/web/packages/syuzhet/vignettes/syuzhet-vignette.html.

Mohammad, Saif, and Peter Turney. 2010. "Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon." In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, Los Angeles, CA: Association for Computational Linguistics, 26–34. Retrieved from https://aclanthology.org/W10-0204.

Mohammad, Saif M., and Peter D. Turney. 2013. "Crowdsourcing a Word–Emotion Association Lexicon." *Computational Intelligence* 29(3): 436–65. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8640.2012.00460.x.

64 Blei, David M. 2012. "Probabilistic Topic Models." *Communications of the ACM* 55(4): 77–84. Retrieved from https://dl.acm.org/doi/10.1145/2133806.2133826

65 Blei, David M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.

Blei, David M. 2012. "Probabilistic Topic Models." *Communications of the ACM* 55(4): 77–84. Retrieved from https://dl.acm.org/doi/10.1145/2133806.2133826.

66 Charmaz, Kathy. 1983. "The Grounded Theory Method: An Explication and Interpretation." Pp. 109–26.

Kelle, Udo. 2007. "'Emergence' vs. 'Forcing' of Empirical Data? A Crucial Problem of 'Grounded Theory' Reconsidered." *Historical Social Research/Historische Sozialforschung*. Supplement, 19: 133–156. Retrieved from https://www.jstor.org/stable/40981074.

67 Breitbart News, tweet, May 18, 2021. Accessed October 6, 2021. Retrieved from. https://twitter.com/BreitbartNews/status/1395126949849280512?s=20.

68 Reddit, "Redditor thinks the StopAsianHate movement was created by The CCP and claims Asians are beating and assaulting thousands of Black people every week without evidence," post, r/RacismAgainst/Asians, created May 5, 2021. Retrieved from https://www.reddit.com/r/RacismAgainstAsians/comments/n5c6rl/redditor_thinks_the_stopasianhate_movement_was/gx1bxek/.

69 Twitter, tweet, June 7, 2020. Accessed October 4, 2021. Retrieved from https://twitter.com/Roger2dot0/status/1269731880183742470?s=20.

70 Hansen, Sarah. 2020. "Here's What The Racial Wealth Gap In America Looks Like Today." *Forbes*. Retrieved from https://www.forbes.com/sites/sarahhansen/2020/06/05/heres-what-the-racial-wealth-gap-in-america-looks-like-today/.

Appendix

Table A.1-1: Top 25 Terms for the Key Phrase "Black People" on Reddit and Twitter

Twitter		Reddit	
TERMS	N	TERMS	N
white	280416	white	6350
police	127533	police	3402
racist	126715	racist	3085
racism	72724	racism	1706
killed	66391	hate	1175
time	62839	blm	1086
hate	57346	lives	1061
cops	52938	killed	1021
right	52276	america	989
person	47547	reddit	965
free	44832	toronto	957
race	44636	trump	942
america	41510	matter	933
man	40737	cmv	869
shit	40657	porn	863
country	38035	cops	825
call	37719	word	785
good	34511	covid	769
killing	34263	person	695
love	33975	support	654
blm	32869	call	605
vote	31588	race	601
lives	31497	feel	596
look	31416	vote	596
take	30218	unathi	590

N = The frequency of occurrence for each term during the data collection period.

Table A.1-2: Top 25 Terms for the Key Phrase "Racial Wealth Gap" on Reddit and Twitter

Twitter		Reddit	
TERMS	N	TERMS	N
black	13598	close	75
close	6527	business	44
people	5406	black	43
white	5263	help	43
debt	4435	big	30
reparations	4037	americas	27
student	4022	billion	25
help	3641	jpmorgan	25
closing	3200	student	24
americans	2738	closing	23
economic	2441	debt	23
america	2114	pandemic	21
racism	2098	america	20
families	2086	new	20
ados	2022	americans	16
money	1783	chase	15
new	1773	economic	14
color	1717	narrow	14
loan	1709	race	14
address	1682	white	13
financial	1629	decline	12
biden	1524	financial	12
read	1432	inequality	12
years	1420	owners	12
inequality	1387	businesses	11

N = The frequency of occurrence for each term during the data collection period.

Table A.1-3: Top 25 Terms for the Key Phrase "Racism" on Reddit and Twitter

Twitter		Reddit	
TERMS	N	TERMS	N
people	243477	black	9402
white	167199	systemic	8878
racist	163813	people	8217
black	150673	white	6836
systemic	80808	police	5588
race	61274	antiblack	5486
america	55001	antiracism	5246
country	52672	toronto	4512
hate	52314	racist	3579
right	47573	school	3038
time	41537	new	2812
good	37563	end	2766
police	35761	america	2384
person	34086	history	2168
man	32498	fight	2096
history	30401	trump	2090
world	30229	board	2058
party	29716	asian	2035
problem	28981	report	1996
someone	28708	world	1968
support	28648	hate	1920
years	28621	problem	1894
new	28388	health	1881
real	27930	protest	1836
take	27813	antiasian	1828

N = The frequency of occurrence for each term during the data collection period.

Table A.2-1: Top 50 Hashtags in Tweets of the Key Term "Black People"

"Black People"			
1	#gazaunderattack	26	#votersuppression
2	#blm	27	#jerusalem
3	#dauntewright	28	#end1033
4	#them	29	#derekchauvintrial
5	#derekchauvin	30	#defundthepolice
6	#smartnews	31	#racist
7	#reparations	32	#justiceforgeorgefloyd
8	#foxnews	33	#police
9	#abolishthepolice	34	#policebrutality
10	#whitesupremacy	35	#freepalestine
11	#backtheblue	36	#alllivesmatter
12	#theview	37	#biden
13	#whiteprivilege	38	#verzuz
14	#blacklivesmatter	39	#sheikhjarrah
15	#georgefloyd	40	#stopasianhate
16	#black	41	#racism
17	#1	42	#blacktwitter
18	#makhiabryant	43	#chauvintrial
19	#covid19	44	#ados
20	#oscars	45	#duantewright
21	#democrats	46	#georgefloydtrial
22	#policereformnow	47	#makiyahbryant
23	#justicefordauntewright	48	#america
24	#adamtoledo	49	#acab
25	#georgia	50	#asianlivesmatter

Note: Hastags organized in order of frequency.

Table A.2-2: Top 50 Hashtags in Tweets of the Key Term "Racial Wealth Gap"

	"Racial Wealth Gap"		
1	#ados	26	#babybonds
2	#americandos	27	#business
3	#fullrepair	28	#financialinclusion
4	#hr40	29	#finance
5	#racialequity	30	#realestate
6	#bhmnbcu	31	#reparationsnow
7	#fintech	32	#5formsofreparations
8	#diversity	33	#investinyou
9	#housing	34	#1
10	#banking	35	#wealthgapwednesdays
11	#reparations	36	#axiosam
12	#wealthgap	37	#1619project
13	#covid19	38	#fullreparations
14	#blackhistorymonth	39	#wecantwait
15	#equity	40	#coronavirus
16	#racism	41	#racialwealthgap
17	#systemicracism	42	#blacklivesmatter
18	#news	43	#racial
19	#supporthr40	44	#purereparations
20	#fhte	45	#inequality
21	#cancelstudentdebt	46	#racialjustice
22	#5injuryareas	47	#studentdebt
23	#wealth	48	#axiosmarkets
24	#blm	49	#2mn5
25	#ai	50	#realtist

Note: Hastags organized in order of frequency.

Table A.2-3: Top 50 Hashtags in Tweets of the Key Term "Racism"

	"Racism"		
1	#racism	26	#aapi
2	#racist	27	#blm
3	#israel	28	#palestine
4	#whitesupremacy	29	#palestinian
5	#gop	30	#gazaunderattack
6	#israeli	31	#hate
7	#biden	32	#apartheid
8	#news	33	#gaza
9	#cdnpoli	34	#stoponlineabuse
10	#timscott	35	#rhod
11	#savepalestine	36	#truth
12	#noroomforracism	37	#yonatanshapira
13	#discrimination	38	#democrats
14	#blacklivesmatter	39	#stopasianhate
15	#freepalestine	40	#covid19
16	#savesheikhjarrah	41	#criticalracetheory
17	#antisemitism	42	#trump
18	#america	43	#antiracism
19	#foxnews	44	#auspol
20	#systemicracism	45	#georgefloyd
21	#usa	46	#tigraygenocide
22	#1	47	#ethnicityisnotacrime
23	#smartnews	48	#humanrights
24	#joebiden	49	#apartheidisrael
25	#asianheritagemonth	50	#republicans

Note: Hastags organized in order of frequency.

Table A.3-1: Top 50 Mentioned Usernames in Tweets of the Key Term "Black People"

	"Black People"		
1	@kingjames	26	@repmaxinewaters
2	@thehill	27	@cnn
3	@foxnews	28	@benshapiro
4	@msnbc	29	@youtube
5	@stillgray	30	@washingtonpost
6	@cbsnews	31	@joyannreid
7	@ambrice46	32	@latimes
8	@breitbartnews	33	@timcast
9	@dailycaller	34	@toure
10	@tariqnasheed	35	@realcandaceo
11	@mlb	36	@nypost
12	@vp	37	@joebiden
13	@ap	38	@breenewsome
14	@tomilahren	39	@prageru
15	@attorneycrump	40	@whcos
16	@talbertswan	41	@aoc
17	@speakerpelosi	42	@charliekirk11
18	@tmz	43	@mrandyngo
19	@potus	44	@jackposobiec
20	@mattwalshblog	45	@breaking911
21	@nytimes	46	@abc
22	@nbcnews	47	@cefaankim
23	@senatortimscott	48	@secdef
24	@gop	49	@seanhannity
25	@mtgreenee	50	@nypdhatecrimes

Note: Username mentions organized in order of frequency.

Table A.3-2: Top 50 Mentioned Usernames in Tweets of the Key Term "Racial Wealth Gap"

	"Racial Wealth Gap"		
1	@joebiden	26	@neilshenvi
2	@reppressley	27	@potus
3	@mattbruenig	28	@ayannapressley
4	@ninaturner	29	@tonetalks
5	@acorns	30	@barackobama
6	@jemelehill	31	@brookingsinst
7	@briebriejoy	32	@aoc
8	@mattyglesias	33	@netflix
9	@politico	34	@washingtonpost
10	@youtube	35	@senschumer
11	@senwarren	36	@icecube
12	@realdonaldtrump	37	@kamalaharris
13	@nhannahjones	38	@cnbc
14	@axios	39	@theblackcaucus
15	@annielowrey	40	@nytimes
16	@cnn	41	@rolandsmartin
17	@citi	42	@businessinsider
18	@officialcbc	43	@mehrsabaradaran
19	@sandydarity	44	@forbes
20	@darrickhamilton	45	@ewarren
21	@breakingbrown	46	@transition46
22	@naacp	47	@corybooker
23	@mckinsey	48	@tariqnasheed
24	@andreperryedu	49	@senbooker
25	@vp	50	@jpmorgan

Note: Username mentions organized in order of frequency.

Table A.3-2: Top 50 Mentioned Usernames in Tweets of the Key Term "Racism"

	"Racism"		
1	@senatortimscott	26	@cnn
2	@foxnews	27	@gop
3	@conceptualjames	28	@benshapiro
4	@mrandyngo	29	@realchrisrufo
5	@jackposobiec	30	@aoc
6	@dribram	31	@seanhannity
7	@ap	32	@joyannreid
8	@drbrittanij	33	@bbcnews
9	@secblinken	34	@tariqnasheed
10	@tulsigabbard	35	@chicagosmayor
11	@youtube	36	@msnbc
12	@joebiden	37	@nytimes
13	@nypost	38	@keir_starmer
14	@mtgreenee	39	@gopleader
15	@prageru	40	@_sjpeace_
16	@kingjames	41	@maryannahernnbc
17	@gatissjudy	42	@nbcnews
18	@potus	43	@timscottsc
19	@thehill	44	@washingtonpost
20	@breitbartnews	45	@realcandaceo
21	@vp	46	@tedcruz
22	@mattwalshblog	47	@nikkihaley
23	@donaldjtrumpjr	48	@ilhanmn
24	@twitter	49	@meghanmccain
25	@uklabour	50	@jpmorgan

Note: Username mentions organized in order of frequency.

B | Governance Studies
at BROOKINGS

The Brookings Institution
1775 Massachusetts Ave., NW
Washington, D.C. 20036
Brookings.edu