



OPEN Analysis of social media language reveals the psychological interaction of three successive upheavals

Steven Mesquiti^{1,7}✉, Sarah Seraj^{2,7}, Andreas H. Weyland³, Ashwini Ashokkumar⁴, Ryan L. Boyd⁵, Rada Mihalcea⁶ & James W. Pennebaker²

Using social media data, the present study documents how three successive upheavals: the COVID pandemic, the Black Lives Matter (BLM) protests of 2020, and the US Supreme Court decision to overturn *Roe v. Wade* interacted to impact the cognitive, emotional, and social styles of people in the US. Text analyses were conducted on 45,225,895 Reddit comments from 2,451,289 users and 889,402 news headlines from four news sources. Results revealed significant shifts in language related to self-focus (e.g., first-person singular pronouns), collective-focus (e.g., first-person plural pronouns), negative emotion (anxiety and anger words), and engagement (e.g., discussion of upheaval-related topics) after each event. Language analyses captured how social justice-related upheavals (BLM, *Roe v. Wade*) may have affected people in different ways emotionally than those that affected them personally (COVID). The onset of COVID was related to people becoming increasingly anxious and people turned inward to focus on their personal situations. However, BLM and the overturning of *Roe v. Wade* aroused anger and action, as people may have looked beyond themselves to address these issues. Analysis of upheaval-related discussions captured the public's sustained interest in BLM and COVID, whereas interest in *Roe v. Wade* declined relatively quickly. Shifts in discussions also showed how events interacted as people focused on only one national event at a time, with interest in other events dampening when a new event occurred. The findings underscore the dynamic nature of culturally shared events that are apparent in everyday online language use.

Keywords Upheavals, LIWC, Social Media, Reddit, Language

Society-scale events can disrupt the lives of millions. In the last decade, the U.S. has witnessed wars, natural disasters, mass shootings, economic disasters, social injustice, and a pandemic. Large-scale upheavals can result in shifts in collective emotions and behaviors^{1,2}. Historically, the study of collective upheavals has relied almost exclusively on closed-ended self-reports to document psychological changes^{3–6}. Questionnaire-based studies have revealed the temporal effects of upheavals on emotional, cognitive, and social processes⁷ and produced models of coping that describe how individuals and communities communicate before, during, and after events⁸. These models operate under the assumption that upheavals unfold within a few days and that society can quickly recover to navigate a post-upheaval world in the weeks that follow⁹. Further, past work on upheavals has often failed to account for differences like events (e.g., an upheaval centered around social injustice vs. a natural disaster) and the co-occurrence of multiple events.

The limitations of this research raise important questions. First, how do people respond when they face a new upheaval in the context of an existing long-term upheaval? Second, how do successive upheavals compare and interact with each other regarding the duration of their psychological effects? With these questions in mind, we examine three social upheavals: the COVID pandemic starting in March 2020, the Black Lives Matter protests following the killing of George Floyd in May 2020, and the US Supreme Court's decision to overturn *Roe v. Wade* in June 2022. The onset of the COVID-19 pandemic, an event spanning several years that affected people living in most parts of the world, provides a unique opportunity to study the interaction of multiple, large-scale upheavals.

¹Princeton University, Princeton, USA. ²University of Texas at Austin, Austin, USA. ³Rice University, Houston, USA. ⁴New York University, New York, USA. ⁵University of Texas at Dallas, Richardson, USA. ⁶University of Michigan, Ann Arbor, USA. ⁷Steven Mesquiti and Sarah Seraj contributed equally to this work. ✉email: sm9518@princeton.edu

COVID-19 pandemic

The COVID-19 pandemic dramatically reshaped the lives of millions^{10,11}. While much of the initial research was cross-sectional, now, more than four years later, researchers have extensively documented the long-term ramifications of the pandemic (e.g.^{10–13}. Analyses of social media data captured the population-level effects of the COVID pandemic on loneliness¹⁴, anxiety¹⁵, and shifts in social cohesion¹⁶. As a result, past work has led to a better understanding of the society-level psychological effects of the pandemic.

Black Lives Matter (BLM)

The BLM movement began in 2013 after the acquittal of Trayvon Martin's killer, George Zimmerman, and gained followers every year. Organizers and leaders became more effective at mobilizing their supporters and distributing their message each year¹⁷. On May 25th, 2020, George Floyd was murdered by Minneapolis police officer Derek Chauvin, sparking nationwide protests and reinvigorating BLM as a national movement. Following Floyd's killing, the nation was gripped by protests. Floyd's death occurred two and half months after the shelter-in-place COVID lockdowns had started. Tens of millions of people watched television and social media coverage of the murder and its aftereffects for days. While some states had begun to relax social distancing restrictions, most continued to follow social distancing protocols. Beginning about a week after Floyd's death, millions of people began joining outdoor BLM protests in cities across the U.S. and much of the world.

Overturing of Roe v. Wade

In 1973, the Supreme Court of the United States ruled that the decision to continue or terminate a pregnancy belonged to the individual, not a governing body—maintained by the Fourteenth Amendment of the U.S. Constitution. The 1973 Roe v. Wade decision argued that abortion was an essential liberty. However, in May 2022, it was leaked that the U.S. Supreme Court would overturn the 1973 Roe v. Wade ruling¹⁸. A month later, the Court released the *Dodds* decision that reversed Roe v. Wade now claiming that there was no constitutional right to abortion. This ruling has since resulted in numerous states banning abortion and forcing women to travel long distances or carry unwanted pregnancies¹⁹.

The psychological effects of upheavals

The experience of overlapping, collective upheavals can produce multiple, large-scale psychological effects. However, these events can be difficult to study with traditional closed-ended, self-report measures. For example, large studies typically require very large samples with repeated measurements over long periods—methods that can be expensive and time-consuming⁷. Because most upheavals are unexpected, surveys typically require several days or weeks to be developed and distributed after the event has occurred. It is even rarer for investigators to have collected pre-upheaval measures for the same population.

With the parallel growth of social media platforms and increasingly sophisticated computational social science methods, it is now possible to track the natural language of very large groups of people over large time scales in ways that would have been unimaginable two decades ago. Social media data from public forums such as Reddit, Twitter, and blogs can grant insights into how people are thinking and feeling at almost a minute-by-minute scale from before, during, and after upheavals. A variety of text analytic and AI strategies, then, allow researchers to track psychological changes over time in theoretical ways to better understand people's natural responses to collective upheavals²⁰. Past work using this approach has demonstrated that analyzing the words people use before, during, and after upheavals can reveal shifts in people's attention^{21,22}, relationship health²³, and emotions²⁴. Therefore, the use of language analysis techniques on social media data offers a promising way to accurately track people's evolving social and psychological processes as they navigate multiple large social upheavals.

During upheavals, individuals often shift their focus as they grapple with an event's ramifications²⁵. The experience of a social upheaval has been linked to increased psychological distress and poorer adjustment¹⁵, which can be tracked through increased first-person singular pronoun use or self-focus language^{26–28}. On the other hand, tracking changes in first-person plural pronoun usage or collective-focus language (e.g. we, us, ours) allows researchers to measure changes in social cohesion in the periods around upheavals. Some research has found that collective upheavals are often followed by a long-term increase in collective language²⁴, which captures the sense of *togetherness* related to the experience of an upheaval^{15,29}. However, little is known about the psychological consequences of sequential upheavals that continually strain society^{7,30}. One potential way to understand these consequences is through examining changes in self and collective focus usage, which allows researchers to track shifts in social orientation and the psychological adjustments related to the experience of each event.

The active experience of a collective upheaval related to the spread of disease or social injustice has the potential to produce large-scale negative emotional effects as people cope with loss, inequity, and uncertainty. These shifts in emotion and uncertainty can be tracked through the increased use of anger (e.g., hate, mad, angry)^{4,31,32} and anxiety words (e.g., worry, fear, afraid)^{4,31,32}. The examination of anger word use is relevant to the three upheavals, as it captures the social frustrations of society with insufficient responses and negligence from governing authorities surrounding upheavals^{33,34}. On the other hand, tracking anxiety words is relevant as the start of a novel, life-threatening upheaval (e.g., global pandemic) can arouse anxiety and a myriad of fears^{7,24}. Feelings of frustration and uncertainty can be further exacerbated through inaction or unclear responses from community leaders (e.g., government responses to COVID-19, and state-level responses to abortion policies). While research has documented the effects of single upheavals on the anger and anxiety of communities^{24,32,35,36}, researchers have rarely attempted to understand how successive upheavals, with different social origins, impact the public's anger and anxiety levels⁷. Looking at changes in anger and anxiety word-use before, during, and

after sequential upheavals provides a means of tracking people's emotional and adjustment challenges as these unique events unfold.

When collective upheavals occur, individuals often flock to social media platforms and news sites to learn and share information about events³⁴. A strong increase in traffic and information sharing grants researchers a simple and reliable way to track engagement (and attention shifts) through mentions of key terms or phrases relevant to an upheaval, as word frequencies can reflect the relative amount of attention people pay to a given topic^{37,38}. Based on past collective upheaval research^{5,8}, discussions related to an upheaval should increase as each event unfolds as people attend to the event, with mentions decreasing in the weeks following each event.

Overview of the current research study

To explore the effects of three different national upheavals: the COVID-19 pandemic, George Floyd's killing, and the overturning of *Roe v. Wade*, we analyzed Reddit conversations (i.e., subreddit comments) occurring across 30 U.S. city-level subreddits. Reddit is a social media platform that hosts over one hundred thousand online communities (or subreddits) where users can interact about various topics³⁹. Within the city subreddits, users may discuss shared experiences and concerns with others in their shared geographical space. City subreddits are particularly valuable in that they reflect the everyday issues and concerns of tens of thousands of people living in a given area, providing rich insights regarding changes in social and emotional processes⁴⁰. In conjunction with U.S. city subreddit comments, we analyzed U.S. news headlines from four major outlets, allowing us to track what topics the media discussed and what information the public consumed⁴¹.

In the current study, over 45 million Reddit comments were analyzed from 2.45 million people across 30 U.S. cities (32 U.S. city subreddits), as well as over 800,000 news headlines from 4 major U.S. news outlets. The combination of these two data sources allows us to observe the evolution of five psychological processes: *self-focus*, *collective-focus*, *anger*, *anxiety*, and *engagement* through existing and validated text analysis methods³⁷. Our study builds on the examination of one-off collective upheaval events, like terrorist attacks^{5,24}, the spread of disease³², and natural disasters⁴². However, our approach is novel in that it attempts to explore the longer-term psychological impacts of *sequential* collective upheavals, as well as the interplay of events. In the following sections, we pose the following set of hypotheses as they pertain to each psychological process.

Self and collective-focus

Past work¹⁵ has extensively detailed the impact of social upheavals on people's orientation towards themselves and others. Upheavals that impact an individual on a personal level (e.g., a break-up) may cause one to turn inward to address these internal issues²³. Based on this, we expected self-focus language to increase across events related to people's anticipation, processing, and coping with the threats posed to their individual lives by the three upheavals. On the other hand, upheavals that disrupt the dynamics of a community (e.g., an earthquake or act of terror) may result in people turning towards others to move towards collective action^{5,24}. We expected collective-focus language to increase after each upheaval, consistent with events like natural disasters⁴³, acts of terror²⁴, and pandemics¹⁶, capturing how people come together to overcome these events.

Anger and anxiety

The onset of each upheaval will likely be related to increases in people's expressions of anger and anxiety as communities navigate each upheaval's repercussions. For anger, we expected the experience of these events to be related to different effects depending on their unique origins. We expected expressions of anger to increase following the murder of Floyd and the overturning of *Roe v. Wade*, capturing nationwide frustration with systemic injustice. In contrast, we did not expect increased expressions of anger at the onset of the pandemic due to the lack of a clear entity to direct frustrations at, as the virus is not an obvious target for anger. Building on research investigating anxiety dynamics during events such as COVID and the Black Lives Matter movement^{35,36}, we anticipated an increase in expressions of anxiety with each successive upheaval unfolding against the backdrop of preceding events.

Engagement

The extent to which an individual attends to an upheaval can be gauged by tracking the frequency of key terms or phrases related to the event, as greater immersion in the event often leads to increased discussion³⁸. In the current study, attention to an upheaval was operationalized as the mention of topics related to an event (e.g., Black Lives Matter, Floyd, abortion, COVID-19, virus, etc.). What we were most interested in was the length of time each upheaval remained relevant in people's lives. In the case of the BLM and *Roe v. Wade*, we were especially interested in how long it takes for the event's influence to dissipate on the backdrop of a preexisting upheaval, the pandemic. We anticipated that discussions about COVID-19 would be most intense at the onset of the pandemic in March 2020, would persist for the longest duration, and fluctuate with each subsequent surge of the virus. With regards to BLM and *Roe v. Wade*, we expected discussions to be the strongest at the onset of the event, replacing discussions of other concurrent upheavals (e.g., a COVID surge). Discussions should then dissipate in the following weeks as society returns to normalcy.

Methods

Datasets

The current study analyzed two types of datasets: Reddit comments and U.S. news headlines. The Reddit sample was an archival dataset of comments extracted from the social media website Reddit⁴⁴. The headlines sample consisted of headlines of news articles from four major US news sources extracted using the Global Database of Events, Language, and Tone (GDELT 2.0)⁴⁵ application programming interface.

Reddit

To understand how civic engagement on race, reproductive rights, and COVID issues changed over time, we relied on the social news and discussion website Reddit. Over 12 million unique users visit Reddit daily to participate in forums (subreddits) related to their interests and hobbies³⁹. Relevant to our study were city-related subreddits, where people can talk about issues about that specific city. The current study extracted comments on US city subreddits between January 2019 to March 2023 using the Pushshift application programming interface⁴⁴ to compare shifts in the language. We chose subreddits that had at least 30,000 subscribers and were geographically dispersed across the US. The extracted dataset consists of 45,225,895 comments from 2,451,289 unique users and 30 US cities, which encompasses 32 subreddits; New York City and Minneapolis each have two active city subreddits, which were combined for analysis (See Fig. 1 for the geographic distribution of subreddits).

News outlet headlines

To understand how media outlets' discussions of upheavals changed over time and how that might differ from the general public's discussions, we used GDELT 2.0⁴⁵ GDELT collects news every 15 min from various news sources across the world. This presents the unique opportunity to capture the description of societal events through the eye of the media, making it an ideal data source for measuring what the general public attends to. The current study sampled approximately 250 news headlines per day from January 2019 to April 2023, totaling 889,402 news article headlines across four major US news outlets: *Fox News*, *New York Times*, *Washington Post*, and *CNN*.

Text analysis procedures

We used two approaches to measure changes in people's engagement, emotional expressions, and social orientation in the periods surrounding the three social upheavals. To gauge engagement shifts, we developed custom dictionaries to identify upheaval-related topics in both Reddit comments and news headlines. These dictionaries tagged mentions of COVID, race, and abortion-related terms over time. This approach allowed us to quantify the percentage of comments and headlines referencing these topics across both datasets. For instance, the COVID dictionary included terms such as 'COVID', 'Coronavirus', 'Pandemic', and 'Lockdown'. The race dictionary contained terms like Black Lives Matter, racism, and the names of Black victims of police brutality. The abortion dictionary contained terms such as abortion, pro-choice, pro-life, *Roe vs. Wade*, and women's rights (See Table S1 for the full list of terms).

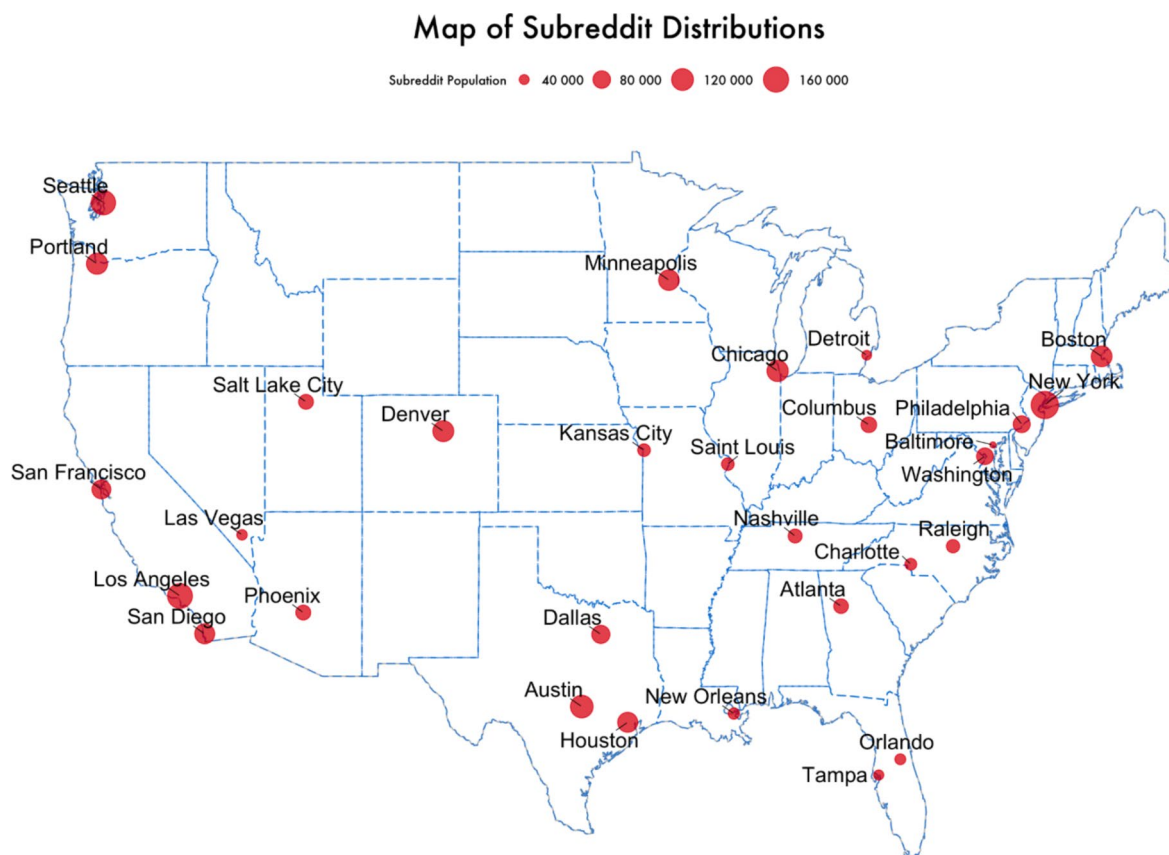


Fig. 1. Geographic distribution and populations of subreddits included in the analyzed dataset (01/01/2019–04/03/2023). Subreddit populations are denoted by the size of the red circles. See Table S4 for a breakdown of statistics on unique authors and posts.

Our second approach involved analyzing Reddit comments to measure emotional expressions and social orientation using the Linguistic Inquiry and Word Count program (LIWC-22)⁴⁶. LIWC is a dictionary-based text analysis program that connects important psychosocial constructs and theories with words, phrases, and other linguistic constructions providing a principled approach to measuring the psychological impact of social upheavals. We focused on the following LIWC dimensions: anger (e.g., hate, angry), anxiety (e.g., worry, fear), collective-focus (e.g., we, our), and self-focus (e.g., I, me). Analyses were performed on comments with at least 25 words, a common cutoff for bag-of-words text analysis²⁰, resulting in a dataset of 14,678,577 comments. Scores for each dimension were calculated as a percentage of total words. LIWC analyses were not applied to News Outlet Headlines due to their brevity, which fell below the 25-word threshold, as this would make such analysis excessively noisy. For more details on the dataset and additional dimensions, see Figures S3-S6 and visit <https://smesquiti.shinyapps.io/blm-shiny-app-test/>.

Results

For all analyses presented, we established a baseline period using data from 2019 to provide a comparison time point. User-level data from the year 2019 (all the users 2019 data) was used as a reference because during this period, there were no significant changes between time blocks for any of the language categories that could be directly attributed to the upheavals in question. For statistical analyses, each user was weighted equally at each time point, regardless of their posting frequency. This approach ensured that each user had only one datapoint per week; past research has shown that aggregating data in this manner can enhance the clarity and interpretability of trends and patterns⁴⁷.

Changes in self-focus, collective-focus, anger, and anxiety words

The social and emotional language patterns were examined for 10 weeks before and after each upheaval (see Fig. 2). Figure 2 depicts graphical trends for the language categories corresponding to self-focus, collective-focus, anger, and anxiety. Each data point represents a weekly mean score of Reddit comments across all the users in the dataset, with each user weighted equally at each time point regardless of the frequency of posts.

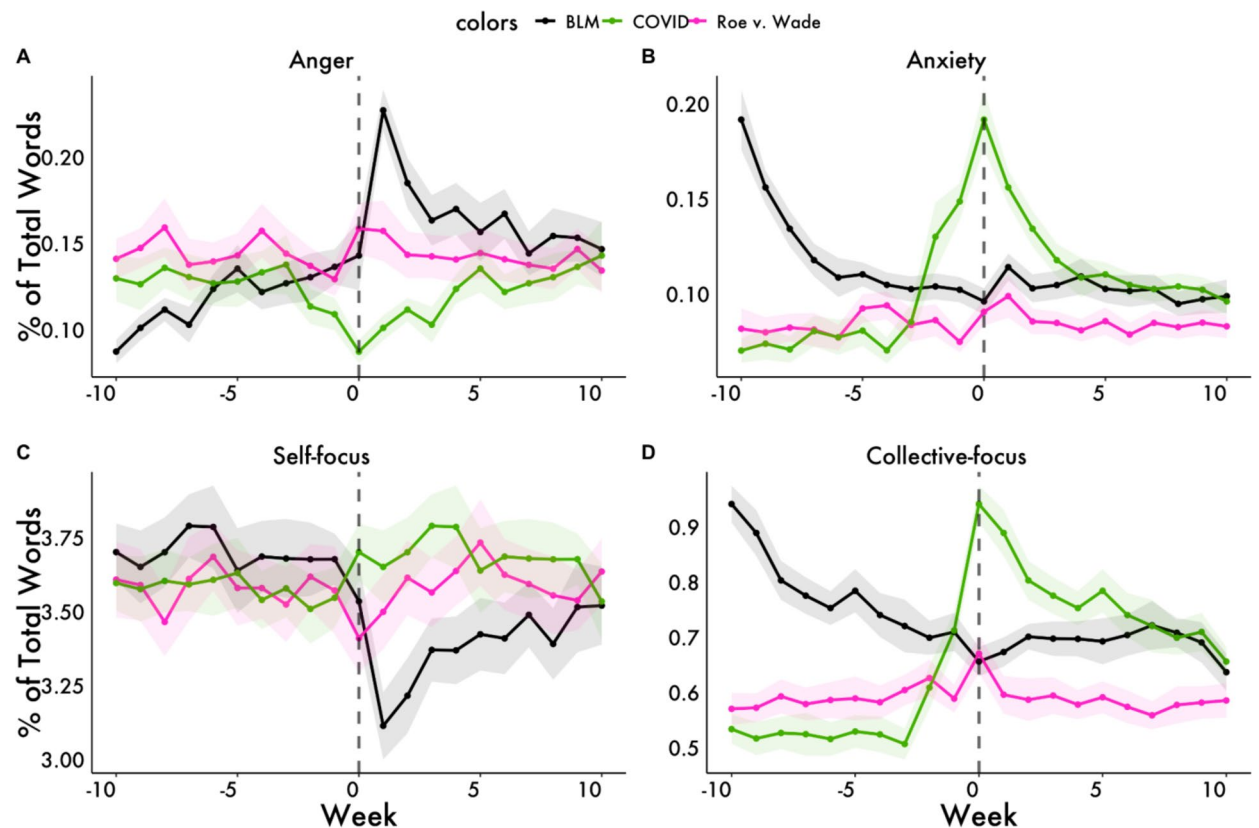


Fig. 2. Social and emotional language shifts in US city subreddits during the Pandemic, BLM protests and the Overturning of Roe v Wade. Time zero (vertical dashed line) represents the onset of each event: BLM (05/25/2020), Roe v. Wade (06/24/2022), COVID (03/15/2020). Each datapoint represents a weekly mean score of Reddit comments across all the users in the dataset, with each user weighted equally at each time point regardless of their frequency of posting. The linguistic categories are dictionaries from the text analysis tool LIWC-22. Self-focus is measured using I-words and Collective-focus using we-words. The y-axis refers to the percentage of words in the posts pertaining to each category. The shaded areas are 95% CIs for weekly datapoint.

Time zero (vertical dashed line) represents the onset of each event: George Floyd is killed, leading to BLM protests (05/25/2020), *Roe v. Wade* was overturned (06/24/2022), and the start of COVID lockdowns in the US (03/15/2020).

Self- and collective-focus language

To quantify changes in self-focus language for each upheaval, paired-sample t-tests were computed to compare each weekly average against the baseline period, using a False Discovery Rate Correction to account for Type-I error inflation⁴⁸ (all FDR-corrected effects in the text have $p < 0.001$ unless noted otherwise). Examining self-focus language shifts, relative to baseline, the public became much less self-focused after Floyd's killing, reaching its lowest point one week after the event (Cohen's $d = 0.14$). The decrease in self-focus following Floyd's killing persisted into August 2020 until finally returning to baseline. There was a similar decrease after the overturning of *Roe v. Wade* (Cohen's $d = 0.04$), yet decreases in self-focus language following the Supreme Court's ruling lasted for only a fraction of the time. Interestingly, there was an opposite trend with the start of the COVID lockdown (Cohen's $d = 0.06$) as self-focus language increased and remained elevated for several months following the COVID lockdown.

Results showed that collective-focus language increased immediately after the onset of the pandemic (see Fig. 3; Cohen's $d = 0.12$) and peaked in the week following lockdown (Cohen's $d = 0.31$). Significant weekly shifts in collective-focus for the pandemic persisted into April of 2020, while for BLM, they subsided after mid-June. Similarly, there was a spike in collective-focus language after the overturning of *Roe v. Wade* in June of 2022 (Cohen's $d = 0.13$). A similar trend was not observed following Floyd's killing, possibly due to already elevated levels during lockdown.

To summarize, the pandemic was related to the most sustained increase in collective-focus language, reflecting widespread effects on social unity, while simultaneously increasing self-focus as people turned inward to address personal issues like job loss. In contrast, social justice events like the BLM protests and the overturning of *Roe v. Wade* were related to decreased self-focus and increased collective-focus, highlighting how people directed their attention outwards to address systemic injustices.

Anger and anxiety

The anger language analyses, relative to the baseline, revealed a slight drop at the start of the pandemic (Cohen's $d = 0.09$). This drop was then followed by a large increase after Floyd's killing (Cohen's $d = 0.16$) and a more muted rise following the overturning of *Roe v. Wade* (Cohen's $d = 0.03$) in June 2022. The largest shift in anger words for BLM occurred one week after Floyd's death (Cohen's $d = 0.15$), while for COVID (Cohen's $d = 0.09$) and *Roe v. Wade* (Cohen's $d = 0.03$), the peak shifts appeared the week the events occurred.

Anxiety language shifts showed increased public anxiety due to pandemic-related uncertainty (Cohen's $d = 0.25$) following the lockdown, which remained elevated until late May 2020. In contrast, Floyd's killing (Cohen's $d = 0.08$) and *Roe v. Wade*'s overturning (Cohen's $d = 0.03$) were associated with only small increases in anxiety, though these events occurred against the backdrop of elevated anxiety from the pandemic.

Taken together, our examination of anger and anxiety words reveals the public's distinct emotional responses related to the different societal upheavals. The COVID-19 pandemic elicited prolonged anxiety, while social justice events like Floyd's killing and *Roe v. Wade*'s overturning triggered more immediate and intense anger. The changes in anger and anxiety language highlight the complex interplay between personal impact and societal issues in shaping public emotional reactions.

Shifts in upheaval-related discussions: news headlines and reddit engagement

To capture shifts in the public's engagement with upheaval-related discussions, we computed the percentage of total city subreddit comments and news headlines that referenced pandemic, racial, and abortion-related issues were analyzed by week. Figure 3a and b depict rates of weekly engagement on race (black lines), COVID (green line), and reproductive rights topics (pink line) from January 2019 to April 2023. Across the two corpora, COVID discussion sparked immediately following the start of the pandemic and remained elevated for the entire three-year period of our dataset until April 2023. Discussion of racial issues peaked the week after Floyd's killing and remained elevated for 6 months until the end of 2020. The increase in race discussions for a prolonged period is juxtaposed with the fact that many Americans were still very much experiencing the challenges presented by the pandemic. While mentions of reproductive rights issues peaked the week of the Supreme Court's Decision, it is noteworthy that the spikes associated with this event were short-lived, returning to near baseline levels in the following weeks (see Table S3.1 for full statistical analyses).

Importantly, to quantify the inter-relationships between topic engagement across the three events, we calculated cross-correlations between the weekly discussion rates for each upheaval of interest across both datasets. Our analyses revealed that discussions of race and abortion ($r_{\text{Reddit}} = -0.074$, $p = 0.014$; $r_{\text{headlines}} = -0.186$, $p < 0.001$), as well as COVID and abortion ($r_{\text{Reddit}} = -0.217$, $p < 0.001$; $r_{\text{headlines}} = -0.305$, $p < 0.001$) were negatively correlated at lag-0. In other words, discussions of abortion-related topics replaced discussions of race and COVID-related topics (and vice-versa; See Fig. 3a and b). Interestingly, the cross-correlations at lag-0 between race and COVID-related discussions ($r_{\text{reddit}} = 0.275$, $p < 0.001$; $r_{\text{headlines}} = 0.061$, $p = 0.042$) decreased relative to lag - 10 ($r_{\text{reddit}} = 0.478$, $p < 0.001$; $r_{\text{headlines}} = 0.76$, $p < 0.001$) and - 5 ($r_{\text{reddit}} = 0.478$, $p < 0.001$; $r_{\text{headlines}} = 0.406$, $p < 0.001$). This profound shift captures how discussions of COVID-issues dropped in tandem with an increase in race-related discussions; which was in spite of the fact that COVID-issues were as relevant as ever, with minorities being disproportionately affected during the pandemic. Additional cross-correlation analyses, which examine the interactions between corpora are available in the Supplementary Information (see Figures S7-8 and Tables S5-6).

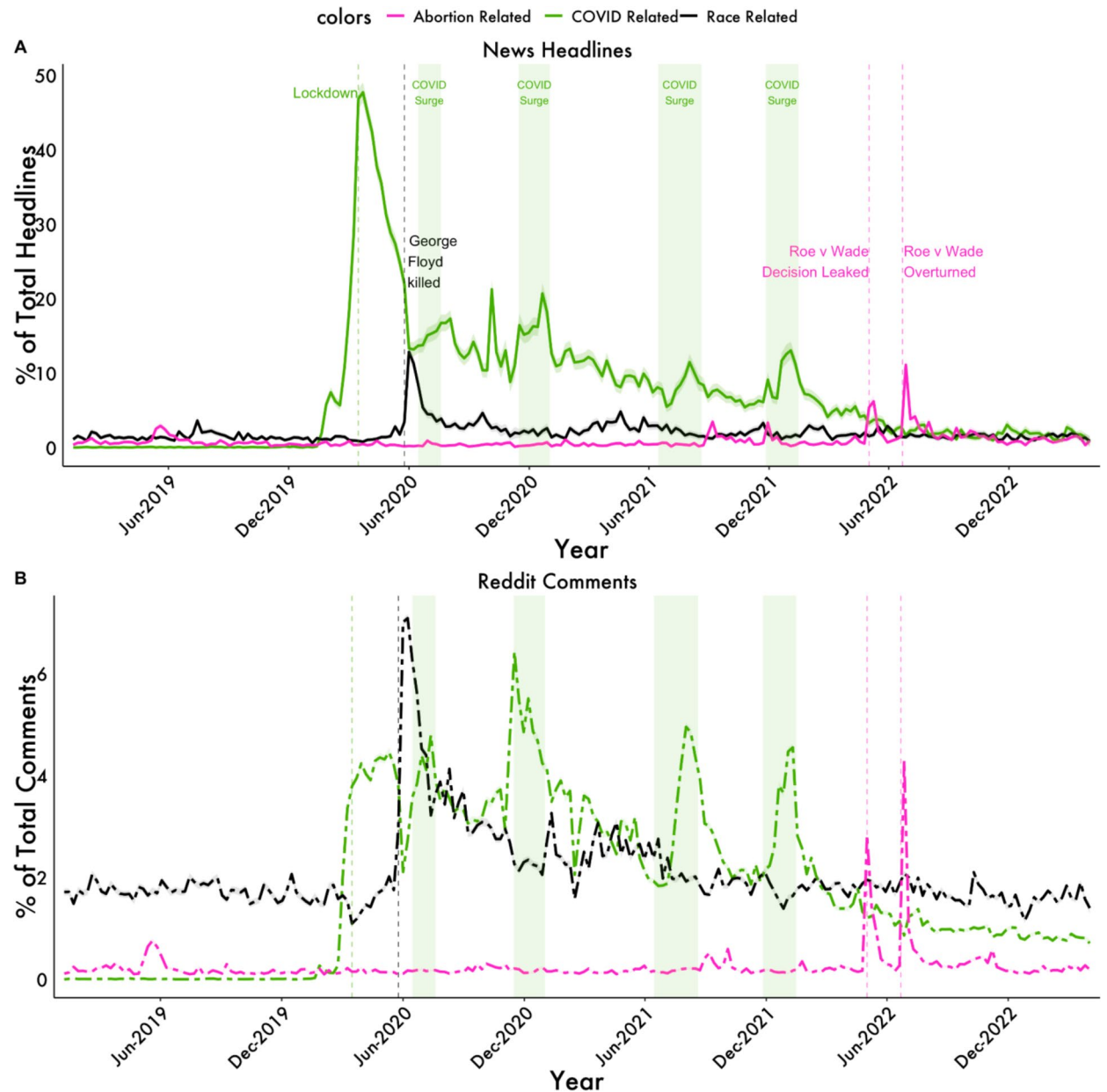


Fig. 3. Rate of engagement on Race (black lines), COVID (green line), and Reproductive Rights topics (pink line) within Reddit city subreddits (dashed lines) and News Headlines (solid lines) from January 2019 to April 2023. The y-axis denotes the percentage of posts that pertain to either Race, COVID, or Reproductive rights. Key events are annotated with vertical dashed lines in the event's respective color. COVID surges refer to large increases in infections, hospitalizations, and deaths and are denoted by green vertical bars. The shaded areas are 95% CIs for each datapoint. The confidence intervals are narrow, which makes the error bands hard to see, especially in (b).

Discussion

This study aimed to identify the psychological effects of overlapping upheavals and track societal discourse surrounding these events by analyzing changes in language use within a large-scale social media dataset. Overall, the analyses revealed the psychological impacts upheavals have on people's lives, through language, and the rate at which people engage with various upheaval-related topics. For example, COVID had the greatest long-term effects on people's lives^{11,13}, which was related to the largest level of conversation and psychological effects across the entire study. Rates of COVID-related discussions peaked multiple times corresponding with the spread of new variants of the virus. On the other hand, reactions to the BLM protests were related to a larger but much briefer surge in topic-related discussions, with major psychological disruptions peaking one week after the killing of Floyd. Responses to the Supreme Court's Roe v. Wade decision were related to a surge in discussions that lasted for only a few weeks. The differences in effects across events may have been a function of the other two

upheavals occurring against the backdrop of the pandemic. Taken together, our work contributes to the growing body of research that explores the psychological effects of social upheavals on society, while also shedding light on how these effects may vary depending on the nature and context of each upheaval.

The different ways people communicate about social upheavals suggest each one has its own psychological fingerprint. Through analyzing the words people use around each upheaval, with validated Natural Language Processing techniques grounded in psychological theory⁴⁹, we can infer the probable actions (e.g., sequestering or protesting) and emotional states (e.g., anger or anxiety) of the affected populations. The COVID-19 pandemic, for instance, was related to a notable increase in both collective-focused and self-focused language following the lockdown. While the collective language increase underscored the shared nature of the experience, the rise in self-focused language reflects individual-level distress in response to possible challenges such as job loss, childcare difficulties, and social isolation. In contrast, the decrease in self-focused language following Black Lives Matter protests and the reversal of *Roe v. Wade*, capture a shift in attention towards collective action addressing systemic injustices. This difference in linguistic patterns across events highlights a crucial distinction: global health crises like the pandemic directly affect daily routines (e.g., school attendance, shopping, social gatherings), making them virtually impossible to ignore on a personal level. Conversely, social justice issues related to race and reproductive rights may have a less pronounced personal impact if they do not directly disrupt an individual's daily patterns. These linguistic variations across different social upheavals provide valuable insights into the psychological processes underlying people's responses to societal challenges and have the potential to offer researchers a deeper understanding of the complex interplay of these events.

Our analysis of anxiety and anger language revealed different emotional responses to disease outbreaks and social justice-related upheavals⁷. The COVID-19 pandemic's onset was followed by a significant and persistent increase in anxiety, which suggests that the social sharing of emotions may have amplified the collective response to the pandemic⁵⁰. These language patterns, when combined with the increase in self-focus language, suggest that while people experienced considerable worry during the pandemic, they struggled to direct their frustrations outwardly as they grappled with internal issues. When comparing events, the pandemic and social justice events activated different emotional processes, among those that can be captured through language use. For example, the BLM protests and *Roe v. Wade* reversal elicited a response characterized by increased collective-focus and anger language and reduced self-focus and anxiety-related terms. This shift in language reflects how individuals may have transcended personal concerns to engage in collective action addressing systemic social inequities, such as protesting. The contrasting emotional responses to health crises and social justice events again highlight the complex interplay between collective and individual emotional processing during times of societal upheaval, revealing how different types of crises can shape public sentiment and behavior in unique ways³².

When an upheaval unfolds, changes in people's attentional processes become evident as discussions related to the event increase on social media platforms and news sites provide more information about it³⁴. The observed patterns of attention shifts align with established research on collective upheavals, which suggests that event-related discussions typically intensify during the occurrence and gradually decline thereafter^{5,8}. For example, while the public maintained a persistent focus on the pandemic due to recurrent surges and novel variants, George Floyd's killing precipitated a significant shift in discourse, with race-related conversations temporarily eclipsing COVID-19 discussions. A similar pattern emerged with the overturning of *Roe v. Wade* in June 2022: discussions about COVID decreased as attention shifted toward this ruling. Despite its continued presence in daily life, each sequential upheaval seemed to draw focus away from COVID-related topics⁵¹. Furthermore, our findings extend the field's understanding of how discussions of upheavals interact. Notably, following the *Roe v. Wade* decision, discussions on reproductive rights increased alongside race topics rather than replacing them. This simultaneous rise may reflect concerns about the ruling's disproportionate impact on marginalized communities, leading to concurrent discussions relevant to both issues⁵². This phenomenon highlights the public's capacity to process and connect social issues simultaneously, rather than treating them as mutually exclusive concerns.

Despite the current work's strengths, including the use of valid language analysis techniques coupled with large-scale social media data, this study is not without limitations. First, as is common with social media studies our design was correlational and therefore we are unable to infer causality unlike experiments conducted in a controlled laboratory setting²⁰. Furthermore, the exclusion of users not active in both periods introduces potential bias, as individuals may have started or ceased activity due to the upheavals being studied. This limitation may affect the representativeness of the sample. When considering our findings, it's also important to acknowledge the limitations of using social media text data as an indirect measure of psychological processes. For example, expressions of emotion words may not always distinguish between genuine feelings and mere discussions about emotions. However, when studying collective upheavals in large populations, discussions about others' emotions and social expectations are integral to the phenomenon under investigation. These expressions may even provide insights about individuals who are not active on the platform^{53,54}. While social media analyses should not be seen as a replacement for surveys, they offer a valuable complementary approach to investigating collective emotions and when combined with other research methods, contribute to a more comprehensive understanding of psychological dynamics of social upheavals at scale.

While LIWC allows for in-depth exploration of psychological states like other bag-of-words methods (e.g.^{20,55}), it does not consider the context in which words are used. Understanding contextual qualities of language use, at times, can be informative for the study of topic discussion. Methods utilizing Large language models (LLM) like Google's Bidirectional Encoder Representations from Transformers⁵⁶ or Generative Pre-trained Transformers (e.g., The LLM that powers the chatbot ChatGPT)⁵⁷ can provide accurate insights for context-specific use of natural language⁵⁸. However, these methods are not without their own limitations. LLMs can be computationally and financially expensive^{20,59}. They can also produce ambiguous results relative to bag-of-words methods and

(in the case of ChatGPT) are based on models so complex as to be uninterpretable and, in many cases, not open source⁶⁰.

While our findings provide valuable, novel insights regarding the psychological effects of three distinct upheavals, several future avenues may be useful. One direction is to corroborate insights from social media data with self-report data, which could allow for additional resolution regarding the study of afflicted psychological processes. Additionally, there are distinct individual differences in language use⁶¹. A rich body of literature suggests that intrapersonal qualities like psychological states⁶², values⁶³, demographics⁶⁴, and motivations⁶⁵ may provide additional insights into exploring individual-level differences in language use during upheavals and ‘for whom?’ are these changes most pronounced. This is especially interesting given that the overturning of *Roe v. Wade* and BLM revolved around social injustices of marginalized groups and may not elicit strong observed effects in individuals who are not directly affected.

Using language analysis tools, social scientists can identify how people naturally talk about collective upheavals in the days and weeks surrounding the event, as well as track the psychological effects of overlapping upheavals. Our study contributes to a growing body of literature that has begun to explore the longer-term coping patterns around sequential and overlapping events. Overall, our research underscores the importance of using natural language data to study the impact of large-scale sequential upheavals on society, as well as how different types of upheavals may touch the lives of groups and cultures in very different ways.

Data availability

Data and Code Availability: All the Reddit datasets used in this study were previously publicly available. (Accessed: 2023-05-01). The processed datasets used in this study and the associated code can be found at the following link: https://osf.io/u785x/?view_only=829a6dcd6d3e439686dba59e68573dcc.

Received: 27 September 2024; Accepted: 3 February 2025

Published online: 17 February 2025

References

1. Goldenberg, A., Garcia, D., Halperin, E. & Gross, J. J. Collective emotions. *Curr. Dir. Psychol. Sci.* **29**, 154–160 (2020).
2. Goldenberg, A., Sweeny, T. D., Shpigel, E. & Gross, J. J. Is this my group or not? The role of ensemble coding of emotional expressions in group categorization. *J. Exp. Psychol. Gen.* **149**, 445–460 (2020).
3. Holman, E. A., Garfin, D. R. & Silver, R. C. Media’s role in broadcasting acute stress following the Boston Marathon bombings. *Proc. Natl. Acad. Sci.* **111**, 93–98 (2014).
4. Jalloh, M. F. et al. Impact of Ebola experiences and risk perceptions on mental health in Sierra Leone, July 2015. *BMJ Glob Health.* **3**, e000471 (2018).
5. Silver, R. C., Holman, E. A., McIntosh, D. N., Poulin, M. & Gil-Rivas, V. Nationwide longitudinal study of psychological responses to September 11. *Jama* **288**, 1235–1244 (2002).
6. Wortman, C. B. & Silver, R. C. The myths of coping with loss. *J. Consult. Clin. Psychol.* **57**, 349 (1989).
7. Silver, R. C., Holman, E. A. & Garfin, D. R. Coping with cascading collective traumas in the United States. *Nat. Hum. Behav.* **5**, 4–6 (2021).
8. Pennebaker, J. W. & Harber, K. D. A social stage model of collective coping: the Loma Prieta Earthquake and the Persian Gulf War. *J. Soc. Issues.* **49**, 125–145 (1993).
9. Gortner, E. M. & Pennebaker, J. W. The archival anatomy of a disaster: media coverage and community-wide health effects of the Texas A&M Bonfire Tragedy. *J. Soc. Clin. Psychol.* **22**, 580–603 (2003).
10. Al-Aly, Z., Agarwal, A., Alwan, N. & Luyckx, V. A. Long COVID: long-term health outcomes and implications for policy and research. *Nat. Rev. Nephrol.* **19**, 1–2 (2023).
11. Benke, C., Autenrieth, L. K., Asselmann, E. & Pané-Farré, C. A. One year after the COVID-19 outbreak in Germany: long-term changes in depression, anxiety, loneliness, distress and life satisfaction. *Eur. Arch. Psychiatry Clin. Neurosci.* **273**, 289–299 (2023).
12. Carey, R. L., Bailey, M. J. & Polanco, C. I. How the COVID-19 pandemic shaped adolescents’ future orientations: insights from a global scoping review. *Curr. Opin. Psychol.* **53**, 101655 (2023).
13. Paglino, E. et al. Excess natural-cause mortality in US counties and its association with reported COVID-19 deaths. *Proc. Natl. Acad. Sci.* **121**, e2313661121 (2024).
14. Badal, V. D. et al. Do words matter? Detecting social isolation and loneliness in older adults using natural language processing. *Front. Psychiatry.* **12**, 728732 (2021).
15. Guntuku, S. C. et al. Tracking mental health and symptom mentions on Twitter during COVID-19. *J. Gen. Intern. Med.* **35**, 2798–2800 (2020).
16. Ashokkumar, A. & Pennebaker, J. W. Social media conversations reveal large psychological shifts caused by COVID-19’s onset across US cities. *Sci. Adv.* **7**, eabg7843 (2021).
17. The Second Act of Social-Media Activism | The New Yorker. <https://www.newyorker.com/culture/cultural-comment/the-second-act-of-social-media-activism>.
18. Politico Supreme Court has voted to overturn abortion rights, draft opinion shows - POLITICO. <https://www.politico.com/news/2022/05/02/supreme-court-abortion-draft-opinion-00029473> (2022).
19. Dench, D. L., Lifchez, K., Lindo, J. M. & Liu, J. L. Are People Fleeing States with Abortion Bans? Working Paper at <https://doi.org/10.3386/w33328> (2025).
20. Kern, M. L. et al. Gaining insights from social media language: methodologies and challenges. *Psychol. Methods.* **21**, 507–525 (2016).
21. Dzogang, F., Lightman, S. & Cristianini, N. Circadian mood variations in Twitter content. *Brain Neurosci. Adv.* **1**, 2398212817744501 (2017).
22. Varol, O., Ferrara, E., Ogan, C. L., Menczer, F. & Flammini, A. Evolution of online user behavior during a social upheaval. In *Proceedings of the ACM conference on Web science* 81–90 <https://doi.org/10.1145/2615569.2615699> (ACM, 2014).
23. Seraj, S., Blackburn, K. G. & Pennebaker, J. W. Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proc. Natl. Acad. Sci.* **118**, e2017154118 (2021).
24. Garcia, D. & Rimé, B. Collective emotions and social resilience in the digital traces after a terrorist attack. *Psychol. Sci.* **30**, 617–628 (2019).
25. Cohn, M. A., Mehl, M. R. & Pennebaker, J. W. Linguistic markers of psychological change surrounding September 11, 2001. *Psychol. Sci.* **15**, 687–693 (2004).

26. Edwards, T. & Holtzman, N. S. A meta-analysis of correlations between depression and first person singular pronoun use. *J. Res. Personal.* **68**, 63–68 (2017).
27. Rai, S. et al. Key language markers of depression on social media depend on race. *Proc. Natl. Acad. Sci.* **121**, e2319837121 (2024).
28. Stade, E. C., Ungar, L., Eichstaedt, J. C., Sherman, G. & Ruscio, A. M. Depression and anxiety have distinct and overlapping language patterns: results from a clinical interview. *J. Psychopathol. Clin. Sci.* **132**, 972–983 (2023).
29. Drury, J., Cocking, C. & Reicher, S. Everyone for themselves? A comparative study of crowd solidarity among emergency survivors. *Br. J. Soc. Psychol.* **48**, 487–506 (2009).
30. Vives, M. L., De Bruin, D., Van Baar, J. M. & FeldmanHall, O. Tweeting under uncertainty: the relationship between uncertain language and negative emotions in the wild. *Emotion* <https://doi.org/10.1037/emo0001376> (2024).
31. Forbes, M. K. & Krueger, R. F. The great recession and mental health in the United States. *Clin. Psychol. Sci.* **7**, 900–913 (2019).
32. Metzler, H. et al. Collective emotions during the COVID-19 outbreak. *Emotion* **23**, 844–858 (2023).
33. Coombs, W. T. Protecting organization reputations during a crisis: the development and application of situational crisis communication theory. *Corp. Reput. Rev.* **10**, 163–176 (2007).
34. Kim, H. K. & Niederdeppe, J. Exploring optimistic bias and the integrative model of behavioral prediction in the context of a campus influenza outbreak. *J. Health Commun.* **18**, 206–222 (2013).
35. Eichstaedt, J. C. et al. The emotional and mental health impact of the murder of George Floyd on the US population. *Proc. Natl. Acad. Sci.* **118**, e2109139118 (2021).
36. Smith, B. A. et al. Impact of COVID-19 on mental health: effects on screening, care delivery, and people with cystic fibrosis. *J. Cyst. Fibros. Off J. Eur. Cyst. Fibros. Soc.* **20**(Suppl 3), 31–38 (2021).
37. Boyd, R. L. & Schwartz, H. A. Natural language analysis and the psychology of verbal behavior: the past, present, and future states of the field. *J. Lang. Soc. Psychol.* **40**, 21–41 (2021).
38. Stone, L. D. & Pennebaker, J. W. Trauma in real time: talking and avoiding online conversations about the death of Princess Diana. *Basic. Appl. Soc. Psychol.* **24**, 173–183 (2002).
39. Bianchi, T. Worldwide visits to Reddit.com from July to December 2023. *Statista* <https://www.statista.com/statistics/443332/reddit-monthly-visitors/> (2024).
40. Aiello, L. M. et al. How epidemic psychology works on Twitter: evolution of responses to the COVID-19 pandemic in the U.S. *Humanit. Soc. Sci. Commun.* **8**, 1–15 (2021).
41. Ramakrishnan, N. et al. 'Beating the news' with EMBERS: forecasting civil unrest using open source indicators. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* 1799–1808 <https://doi.org/10.1145/2623330.2623373> (Association for Computing Machinery, 2014).
42. Gruebner, O. et al. Cities and mental health. *Dtsch. Arzteblatt Int.* **114**, 121–127 (2017).
43. Nolen-Hoeksema, S. & Morrow, J. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta Earthquake. *J. Pers. Soc. Psychol.* **61**, 115 (1991).
44. Baumgartner, J., Zannettou, S., Keegan, B., Squire, M. & Blackburn, J. The Pushshift Reddit Dataset. *Proc. Int. AAAI Conf. Web Soc. Media.* **14**, 830–839 (2020).
45. Leetaru, K. & Schrodt, P. A. GDELT: Global Data on Events, Location and Tone. (2013).
46. Pennebaker, J. W., Boyd, R. L., Booth, R. J., Ashokkumar, A. & Francis, M. *Linguistic Inquiry and Word Count: LIWC-22* (Pennebaker Conglomerates, Inc., 2022).
47. Giorgi, S. et al. The remarkable benefit of user-level aggregation for lexical-based population-level predictions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* 1167–1172 <https://doi.org/10.18653/v1/D18-1148> (Association for Computational Linguistics, 2018).
48. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
49. Boyd, R. L., Ashokkumar, A., Seraj, S. & Pennebaker, J. W. The Development and Psychometric properties of LIWC-22. <https://doi.org/10.13140/RG.2.2.23890.43205> (2022).
50. Rimé, B., Páez, D., Basabe, N. & Martínez, F. Social sharing of emotion, post-traumatic growth, and emotional climate: follow-up of Spanish citizen's response to the collective trauma of March 11th terrorist attacks in Madrid. *Eur. J. Soc. Psychol.* **40**, 1029–1045 (2010).
51. Smirnov, O. & Hsieh, P. H. COVID-19, climate change, and the finite pool of worry in 2019 to 2021 Twitter discussions. *Proc. Natl. Acad. Sci.* **119**, e2210988119 (2022).
52. Caraher, R. & Reproductive Injustice?, A. County-Level Analysis of the Impact of Abortion Restrictions on Abortion Rates. SSRN Scholarly Paper at <https://doi.org/10.2139/ssrn.4381662> (2023).
53. Galesic, M. et al. Human social sensing is an untapped resource for computational social science. *Nature* **595**, 214–222 (2021).
54. Garcia, D., Pellert, M., Lasser, J. & Metzler, H. Social media emotion macroscopes reflect emotional experiences in society at large. *Preprint at* <https://doi.org/10.48550/arXiv.2107.13236> (2021).
55. Sun, J., Schwartz, H. A., Son, Y., Kern, M. L. & Vazire, S. The language of well-being: tracking fluctuations in emotion experience through everyday speech. *J. Pers. Soc. Psychol.* **118**, 364–387 (2020).
56. Devlin, J., Chang, M. W., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Preprint at <https://doi.org/10.48550/arXiv.1810.04805> (2019).
57. Brown, T. B. et al. Language Models are Few-Shot Learners. Preprint at <https://doi.org/10.48550/arXiv.2005.14165> (2020).
58. Rathje, S. et al. GPT is an effective tool for multilingual psychological text analysis. *Preprint at* <https://doi.org/10.31234/osf.io/sekf5> (2023).
59. Narayanan, D. et al. Efficient large-scale language model training on GPU clusters using megatron-LM. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* 1–15 <https://doi.org/10.1145/3458817.3476209> (Association for Computing Machinery, 2021).
60. Chen, L., Zaharia, M. & Zou, J. How is ChatGPT's behavior changing over time? Preprint at <https://doi.org/10.48550/arXiv.2307.09009> (2023).
61. Newman, M. L., Groom, C. J., Handelman, L. D. & Pennebaker, J. W. Gender differences in language use: an analysis of 14,000 text samples. *Discourse Process.* **45**, 211–236 (2008).
62. Schwartz, H. A. et al. Personality, gender, and age in the language of social media: the open-vocabulary approach. *PLoS One.* **8**, e73791 (2013).
63. Kennedy, B. et al. The (moral) language of hate. *PNAS Nexus.* **2**, pga210 (2023).
64. Pennebaker, J. W. & Stone, L. D. Words of wisdom: language use over the life span. *J. Pers. Soc. Psychol.* **85**, 291 (2003).
65. Schultheiss, O. C. Are implicit motives revealed in mere words? Testing the marker-word hypothesis with computer-based text analysis. *Front. Psychol.* **4**, (2013).

Acknowledgements

This project was funded, in part, by the National Science Foundation (SES1758835, PI: Gosling), the John Templeton Foundation (#61156, PI: Mihalcea), National Institutes for Health (R01 GM112697, PI: De Choudhury), and the Federal Bureau Investigation (15F06718R00006603, PI: Pennebaker). We are indebted to the following people for support, advice, data gathering, and/or feedback on earlier drafts of the paper: Isabel Webb Carey,

the baristas at the Ultimo coffee shop in Graduate Hospital, Dr. Julie Swets, as well as members of the Falk and Pennebaker laboratories.

Author contributions

SM, SS, and JWP designed the study. SM, SS, and RLB processed and extracted the data. SM and SS analyzed the data. SM, SS, AA, AW, RM, and JWP collected and analyzed supplementary data. SM, SS, AW, AA, RLB, RM, and JWP wrote the paper.

Declarations

Competing interests

James W. Pennebaker, Ryan L. Boyd, Sarah Seraj, and Ashwini Ashokkumar are co-developers of LIWC-22, which is a commercial product. Steven Mesquiti, Andreas Weyland, and Rada Mihalcea declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-89165-z>.

Correspondence and requests for materials should be addressed to S.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025