



Global evidence of expressed sentiment alterations during the COVID-19 pandemic

Jianghao Wang^{1,2,5}, Yichun Fan^{2,5}, Juan Palacios², Yuchen Chai², Nicolas Guetta-Jeanrenaud^{2,3}, Nick Obradovich⁴, Chenghu Zhou¹✉ and Siqi Zheng²✉

The COVID-19 pandemic has created unprecedented burdens on people's physical health and subjective well-being. While countries worldwide have developed platforms to track the evolution of COVID-19 infections and deaths, frequent global measurements of affective states to gauge the emotional impacts of pandemic and related policy interventions remain scarce. Using 654 million geotagged social media posts in over 100 countries, covering 74% of world population, coupled with state-of-the-art natural language processing techniques, we develop a global dataset of expressed sentiment indices to track national- and subnational-level affective states on a daily basis. We present two motivating applications using data from the first wave of COVID-19 (from 1 January to 31 May 2020). First, using regression discontinuity design, we provide consistent evidence that COVID-19 outbreaks caused steep declines in expressed sentiment globally, followed by asymmetric, slower recoveries. Second, applying synthetic control methods, we find moderate to no effects of lockdown policies on expressed sentiment, with large heterogeneity across countries. This study shows how social media data, when coupled with machine learning techniques, can provide real-time measurements of affective states.

The December 2019 coronavirus disease (COVID-19) outbreak has threatened the stability of health-care systems and generated unparalleled social and economic disruptions in nations across the globe¹. The health crisis has brought emotional distress to citizens beyond those contracting the disease². Over the course of 2020, individuals faced novel health risks associated with daily activities, shortages of resources and increased uncertainty in their financial futures and social lives³. In addition, numerous governments have imposed strict controls on movement, infringing on personal freedoms and increasing loneliness, depression, anxiety and other negative emotions^{4,5}.

Many governments worldwide are incorporating measures of citizens' subjective well-being into policy decision-making to complement economic indicators such as gross domestic product^{6–8}. Affective state (positive and negative emotions) is one of the central components of subjective well-being^{9,10} and is likely to be greatly impacted by the COVID-19 pandemic¹¹. As the COVID-19 crisis continues and the world faces the expectation of recursive virus outbreaks, governments worldwide are increasingly concerned about the emotional impacts of COVID-19 outbreaks and anti-contagion policies used to manage the pandemic¹². Nonetheless, while there have been numerous efforts to track COVID-19 infection and policy responses globally^{13,14}, there are no standardized high-frequency measures of the affective aspect of subjective well-being.

Tracking the affective states of citizens during disruptive events such as natural disasters and epidemics is especially challenging due to the unpredictability and volatility of these crises¹⁵. There are a variety of laudable survey initiatives, such as the Weekly COVID-19 Snapshot Monitoring in Germany¹⁶, to track the evolution of risk perception using cross-sectional national surveys. However, such traditional survey measurements are prohibitively expensive at the

global scale and usually suffer from limited coverage, insufficient sampling frequency and substantial delays¹⁷. Social media data have offered a valuable complement to track the affective aspect of subjective well-being⁹. Sentiment analysis, which uses natural language processing (NLP) and computational linguistics, allows standardized quantification of emotional states from text¹⁸. Expressed sentiment indices built on people's posts on social media platforms have been validated to be meaningfully correlated with subjective well-being measured by conventional surveys (for example, the Gallup–Sharecare Well-Being Index survey, which is currently the most definitive measurement of subjective well-being and is widely used in well-being research⁹). Recent literature has applied social media expressed sentiment indices to estimate the effects of temperatures^{19,20}, local air pollution²¹, natural disasters^{22,23} and other environmental stressors²⁴ on subjective well-being.

Here we build a global dataset that tracks expressed sentiment at the national and subnational (state/province) levels with high temporal and spatial granularity using anonymized and aggregated data from the two largest social media microblogging platforms (Twitter and Weibo (the Chinese equivalent of Twitter)). The data contain more than 600 million geotagged social media posts on all topics published by 10.56 million individuals during the first wave of the COVID-19 pandemic (from 1 January to 31 May 2020) (Fig. 1). Since the sample of COVID-19-related discussions might not be a good representation of the affective state of the general population and could be polluted by political campaigns, we exclude tweets directly related to COVID-19 when building out the main sentiment indices. We then apply the state-of-the-art Bidirectional Encoder Representations from Transformers (BERT)²⁵ NLP technique to compute daily sentiment measures in over 100 countries standardized across 65 languages (Methods). Unlike dictionary-based

¹State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. ²Center for Real Estate, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA, USA. ³Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁴Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany. ⁵These authors contributed equally: Jianghao Wang, Yichun Fan. ✉e-mail: zhouch@leis.ac.cn; sqzheng@mit.edu

sentiment analysis such as Linguistic Inquiry and Word Count (LIWC)²⁶, deep-learning-based BERT algorithms allow word representations to be enriched with contextual information and enable multilingual computations²⁷.

On the basis of our measures of expressed sentiment and under the assumption that the existing evidence on the correlation between sentiment and affective well-being is valid, we conduct two inter-related empirical exercises to evaluate the global affective impacts of the COVID-19 pandemic and policy responses. The first exercise estimates the overall expressed sentiment alterations associated with COVID-19. We employ reduced-form econometric methods to measure the sentiment drops related to the advent of COVID-19 human-to-human transmission and estimate the recovery time needed for sentiment to return to the baseline levels. Our second exercise applies synthetic control methods (SCM) to explore how social-media-based sentiment measures can be used by countries and international organizations to evaluate alterations in affective states after policy interventions or events, using lockdown policies as an example. To facilitate comparisons across countries, our estimates of sentiment alteration are all measured in the unit of a country's own magnitude of sentiment variation (that is, the standard deviation of sentiment time series before COVID-19). We describe our approach in more detail in the Methods.

Results

Expressed sentiment alterations during the COVID-19 pandemic. The advent of COVID-19 was followed by a sizable drop in global expressed sentiment (Fig. 1b), especially after the World Health Organization (WHO) declared COVID-19 a global pandemic on 11 March 2020. Figure 2a highlights the universality of the sentiment change associated with the COVID-19 pandemic: all countries in our sample sequentially suffered sentiment alterations around the beginning of the pandemic, with varying magnitudes and durations. Sentiment gradually recovered after the shock, showing a similar trend with survey measurements of risk perception (for example, COVID-19 Snapshot Monitoring conducted in Germany²⁸). To measure the patterns of sentiment alterations created by COVID-19, we develop two global indices (Methods, 'Modelling of sentiment dynamics'): sentiment drop and recovery half-life.

We define the sentiment drop as a country's sentiment decline from the level before COVID-19 to its lowest value during the first wave of COVID-19. To estimate it, we separately fitted a sentiment trend before the date sentiment started to decline and after the date it reached its lowest value using local linear regressions; we then applied regression discontinuity design (RDD) to quantify the gap (see Methods, 'Sentiment drops', for the details). RDD is a quasi-experimental design commonly adopted to measure the impacts of abrupt and exogenous events^{29–31}, which allows us to separate the structural shock from daily fluctuations in sentiment. We measured the magnitude of a country's sentiment drop relative to the standard deviation of the country's sentiment before COVID-19 (that is, before the detected date of sentiment decline) for comparability across countries. We find that the sentiment impact of the COVID-19 pandemic is negative for all countries, with the average drop equivalent to 0.85 s.d. ($P < 0.001$; 95% confidence interval (CI),

(0.60, 1.10); Fig. 2b). The sentiment changes are statistically significant at the 5% level for the vast majority (91.5%) of the countries and present large heterogeneity across countries (see Supplementary Table 5 and Supplementary Fig. 12 for the country-specific results). The largest sentiment drops took place in Australia (coefficient = -3.308 ; $P < 0.001$; 95% CI, $(-3.656, -2.960)$), Spain (coefficient = -2.927 ; $P < 0.001$; 95% CI, $(-3.204, -2.650)$), the United Kingdom (coefficient = -2.354 ; $P < 0.001$; 95% CI, $(-2.521, -2.186)$) and Colombia (coefficient = -2.112 ; $P < 0.001$; 95% CI, $(-2.326, -1.899)$), while Botswana, Tunisia, Oman, Bahrain and Greece had effect sizes smaller than -0.15 s.d.

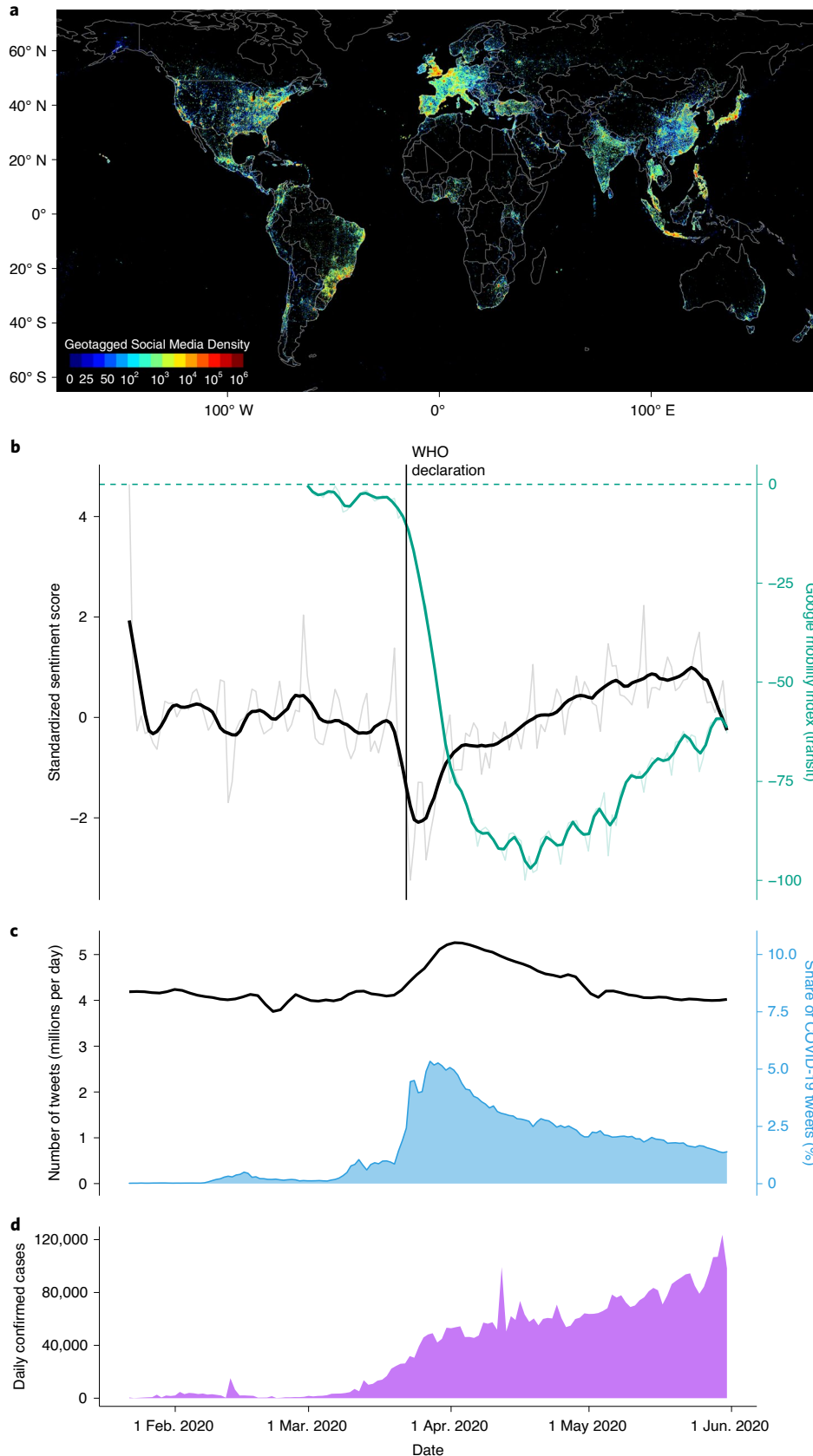
To contextualize our results, we first examined the average sentiment variations over the course of a week before COVID-19. We find that people have a higher expressed sentiment on weekends (Supplementary Fig. 13). The average difference in sentiment between Sunday and Monday (that is, the unhappiest day) was 0.18 s.d. across countries, which has a similar magnitude with findings in previous work¹⁹. The effect size of COVID-19 (that is, 0.85 s.d. across the globe) is more than 4.7 times as large as this weekly sentiment drop from Sunday to Monday. In addition, according to a previous study¹⁹, the difference in sentiment between days with maximum temperatures above 40 °C and days with the most comfortable maximum temperatures of 21–24 °C is 0.21 s.d. (25% of our estimated sentiment drop during COVID-19); nearby hurricanes cause a reduction in expressed sentiment of around 0.4–0.7 s.d. (47–82% of our estimated sentiment drop during COVID-19). This suggests that the acute impact of COVID-19 on sentiment is potentially more pronounced than that of extreme hot temperatures and climate disasters.

Besides the onset of sentiment drop, we estimate how long it took for people's expressed sentiment to recover. Our second index, sentiment recovery half-life, measures the days it took for a country to recover from the lowest sentiment to half of its stationary state of recovered sentiment (that is, the convergence value in the calibrated sentiment recovery model). It is important to mention that the recovery time not only reflects the emotional resilience towards the pandemic itself. This measure should be interpreted as a combined effect of pandemic severity and regulatory policies, and it may be influenced by other events happening around the first wave of the pandemic within each country. Following best practices proposed by previous studies³², we characterize the sentiment recovery process of each country in our sample as an exponential function starting from its minimum sentiment using equation (3) (Methods, 'Sentiment recovery', and Supplementary Fig. 10). The estimated indices show that the recovery half-life varies substantially across countries (Fig. 2c), ranging from 1.2 days (Israel) to 29.0 days (Turkey). Meanwhile, the new stationary state of recovered sentiment until 31 May 2020 also varies across countries: as shown in Fig. 2c, 18% of countries had sentiment recovered to a lower level (below -1.00 s.d.), 35% of countries recovered to the normal value (between -1.00 s.d. and 1.00 s.d.) and 46% of countries recovered to a higher level (above 1.00 s.d.). These results suggest a longer-term alteration in expressed sentiment in countries that show large discrepancies between their recovered sentiment and their average sentiment before COVID-19.

Fig. 1 | Global geotagged social media data coverage and sentiment, mobility and epidemic evolution. **a**, The spatial distribution of over 600 million geotagged social media posts (from Twitter and Weibo) from 10.56 million individuals from 1 January to 31 May 2020. The density of posts is higher (red) in large cities and developed regions and lower (blue) in developing cities and rural areas. **b**, The time series of the global standardized daily sentiment score (grey; seven-day smooth line in black) and the Google mobility index at transit (light green; seven-day smooth line in green). The sentiment index is obtained by applying NLP algorithms to tweets (Methods). The mobility index is from Google (<https://www.google.com/covid19/mobility/>), with trends at the 'transit stations' being used. **c**, The daily number of geotagged tweets (black) and the share of COVID-19 tweets (blue). **d**, The time series of daily new confirmed cases (purple), which was collected from the COVID-19 dashboard by the Center for Systems Science and Engineering at Johns Hopkins University (<https://coronavirus.jhu.edu/map.html>).

Finally, we explore how the magnitude of sentiment alterations by country (sentiment drop and recovery half-life) correlates with countries' pandemic severity, governance and cultural traits

(see Methods, 'Data', for the full list of variables collected for testing). These results are intended to motivate exploration for future studies, as our country-level correlation analysis cannot pin down



the causal mechanisms. The results show that countries with more confirmed COVID-19 cases experienced a larger sentiment drop (Supplementary Fig. 8a; Pearson correlation, $\rho=0.250$; $P=0.007$; 95% CI, (0.070, 0.414)). Moreover, we find that the governance efficiency index from the World Bank, a comprehensive measure of public sectors' performance proven to predict a country's capacity to control the COVID-19 pandemic³³, is positively correlated with fast recovery (Supplementary Fig. 8b; $\rho=-0.259$; $P=0.011$; 95% CI, (-0.437, -0.061)). Beyond objective characteristics, cultures usually play an important role in how people perceive and react to collective threats. Previous studies have shown how nations with loose cultures (that is, having lenient norms and punishments for deviance) had more difficulty coordinating in the face of the pandemic^{33,34}. Consistently, we find a positive correlation between cultural looseness and sentiment drops (Supplementary Fig. 8c; $\rho=0.447$; $P=0.001$; 95% CI, (0.187, 0.649)). We also conduct correlation tests for other dimensions (such as a country's development stage, health security and other cultural constructs from previous studies³⁵), and none of them pass the 5% significance threshold after family-wise adjustment for multiple hypothesis testing (Supplementary Table 2).

Impacts of lockdowns on expressed sentiment. Given the absence of a vaccine during the first wave of the pandemic, many governments implemented a series of non-pharmaceutical interventions to contain the spread of the virus, with lockdowns being the most stringent ones. Lockdowns aim at minimizing physical contact among citizens, which deprives individuals of their freedom to undertake a wide range of daily activities and creates financial risks linked to job loss. Nevertheless, lockdowns could also generate a sense of security regarding virus control and curb public concern about the pandemic¹. Given these circumstances, the direction and magnitude of sentiment change after lockdowns are likely to be context-specific, depending on the timing of implementation and public attitudes towards the policy.

The critical empirical challenge is that governments tend to impose lockdown measures in response to uncontrolled virus surges, challenging the construction of proper comparisons for the lockdown countries. Researchers can easily fall into the trap of comparing countries severely struck by COVID-19 and having a worsening sentiment trend with countries in a better situation, thus leading to false conclusions that lockdown itself worsens sentiment. Here we apply SCM³⁶ to construct suitable comparisons for each lockdown country. SCM allows comparisons of a treated country's sentiment after lockdown with the weighted average sentiment constructed from a pool of control countries that have no or late lockdown. Weights are assigned according to the similarity of the control countries with the lockdown countries of interest in pre-lockdown sentiment, pandemic severity and development indicators (Methods, 'Impacts of lockdowns on expressed sentiment').

We find that, on average, lockdown policies are followed by a small and positive sentiment change when comparing the average sentiment change across all locked-down countries with that of their synthetic controls in the first week of their implementation (Fig. 3a). Of the 52 countries that have over 500 daily geotagged social media

posts, that implemented nationwide stay-at-home orders and for which we can construct valid synthetic controls, 34 (65%) show a positive sentiment impact of lockdown policy, and 18 (35%) display a negative effect (Fig. 3b; the country-specific results of lockdowns are summarized in Supplementary Tables 7 and 8). The sentiment change is rather subtle compared with the reduction in mobility in the first week after lockdown policies are implemented, estimated using the same empirical strategy (Supplementary Fig. 15a). In contrast, as expected, we find overwhelmingly (89% of countries) negative effects of lockdown on mobility in the first week post-lockdown (Supplementary Fig. 15b), suggesting that our method is effective in picking up the changes following lockdown.

Although the sentiment change after lockdown is small in magnitude for most countries, we do see notable dispersion in effect size, ranging from -1 s.d. to $+1.2$ s.d. Statistical inferences constructed through permutation tests also show countries having both significantly positive and negative effects (Supplementary Note 5). We find suggestive evidence that for countries having significantly negative sentiment change post-lockdown (Supplementary Fig. 4a), the negative effect is concentrated in the unhappiest (the bottom sentiment quartile) social media posts within a country, compared with the happiest ones (the top sentiment quartile) (Supplementary Fig. 4b,c). These results suggest that lockdown policies could have disproportionate emotional impacts on the unhappiest people. Due to the macro nature of this study, the distributional effects will need future research to validate.

Discussion

Timely monitoring of the affective aspect of subjective well-being is essential for public policy design and management⁷. Survey methods usually have limited samples within developing economies and require considerable time to execute, leading to a lack of generalizability and time delay when faced with catastrophes. Using high-frequency social media data and state-of-the-art NLP algorithms, we construct a comprehensive database of expressed sentiment covering over 100 countries worldwide (74% of the world population). Our method applies a state-of-the-art sentiment metric using lexical expressions of social media data to measure the changes in emotional states, which is validated to correlate with traditional survey measures of subjective well-being^{9,15,37,38} (see Supplementary Note 2 for expanded discussions on this topic).

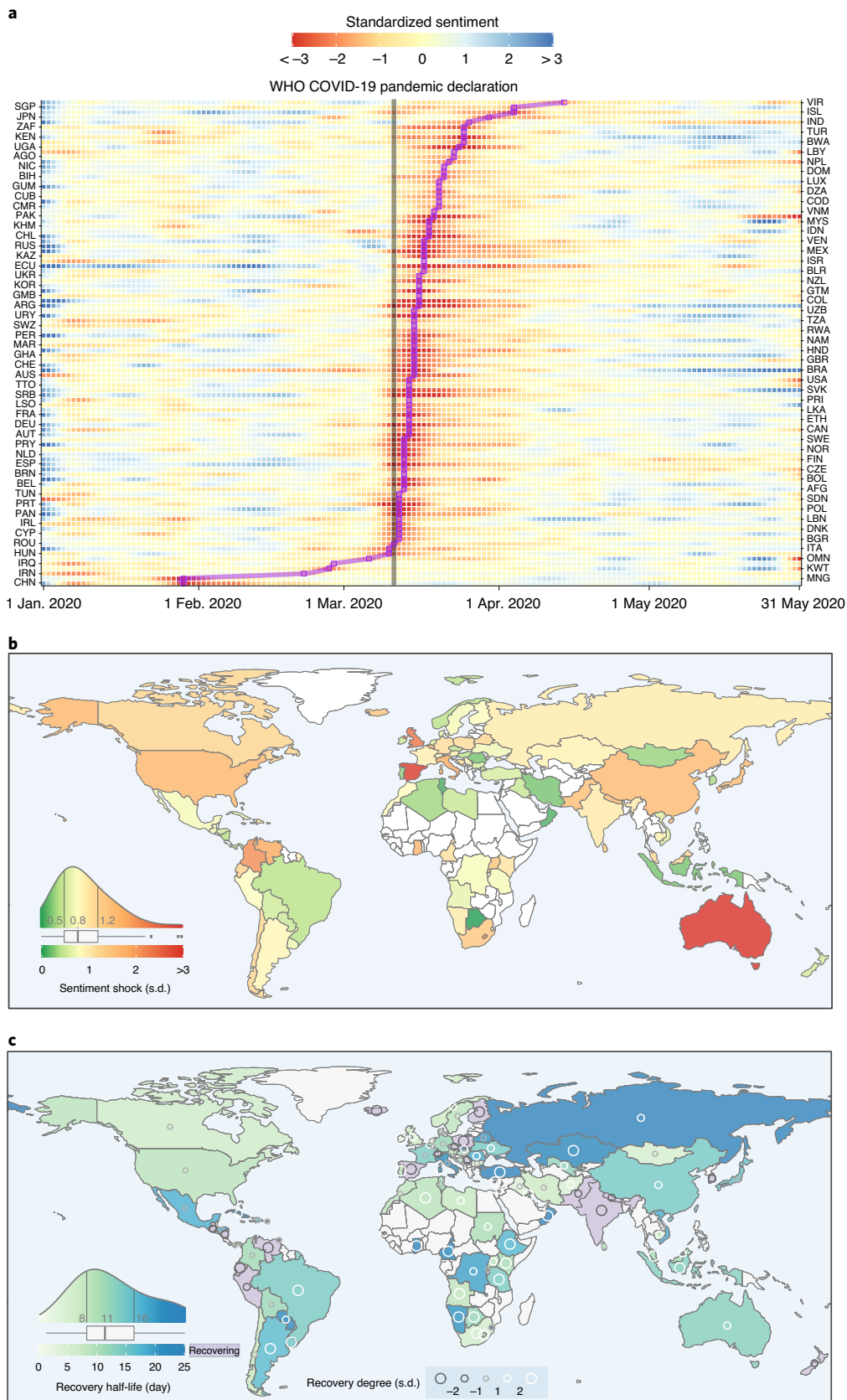
Leveraging this database, we provide empirical evidence at the global scale of the alterations in expressed sentiment associated with COVID-19. We find a remarkable consistency in the way COVID-19 induced sentiment alterations across countries. Though taking place at different time points, almost all countries showed an abrupt and statistically significant sentiment decline around the beginning of the COVID-19 pandemic, followed by an asymmetric and slower recovery. Despite the similarity in the shapes of sentiment response curves, sentiment drops were larger in countries having more confirmed COVID-19 cases or looser cultures, while the recovery was faster for countries with efficient governments.

We also display how this global sentiment database can be used to model sentiment changes after lockdown policies. Though severe emotional costs of lockdown policies are widely assumed, we found

Fig. 2 | Global patterns of COVID-19-induced sentiment change. **a**, Standardized sentiment index by country (labelled with ISO three-letter country codes). The vertical line shows the time the WHO declared COVID-19 a global pandemic, and the purple line represents the sentiment nadir (that is, the lowest point) of each country. The countries are labelled with ISO 3166-1 alpha-3 country codes. **b**, Mapping of sentiment drop variations across countries. Sentiment drop is defined by the magnitude of sentiment decline from each country's average sentiment before COVID-19 to its lowest sentiment value and measured by RDD (Methods, 'Sentiment alterations during the COVID-19 pandemic'). The magnitude of drops is measured in standard deviations before COVID-19. Red represents large shock; green represents small shock. The box plot shows the median (interquartile range) of sentiment shock sizes in s.d. **c**, Mapping of recovery half-life variations across countries. Light colours correspond to quick recovery, blue indicates long recovery and purple indicates that the country is still in the recovering stage by 25 May 2020 (that is, recovery degree is below -1 s.d.). The circle sizes further display the recovery degrees by country. The box plot shows the median (interquartile range) of recovery half-life in days.

little evidence supporting this hypothesis (at least in the short term), when comparing countries that had implemented lockdown policies with their synthetic controls. This seemingly surprising

result does not indicate that the social and financial risks created by lockdowns are trivial; instead, it suggests that for countries with severe pandemic situations, letting the virus spread without



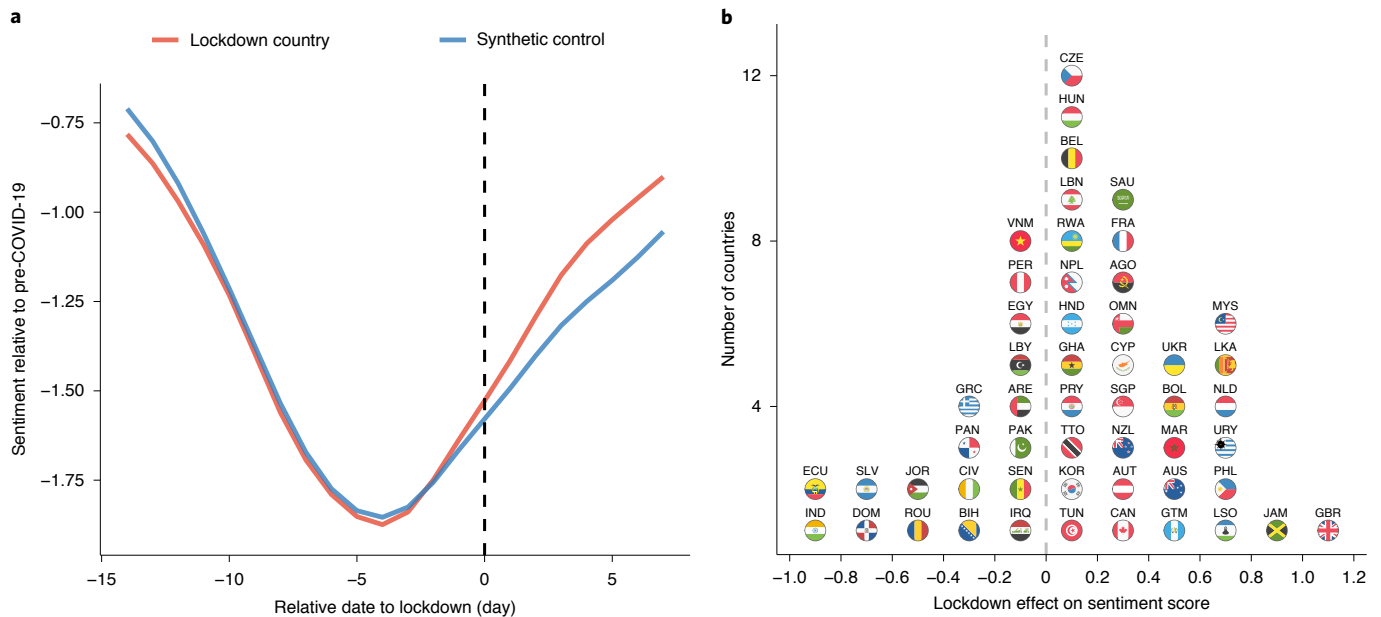


Fig. 3 | Impact of lockdown policy on sentiment. **a**, Average changes in sentiment scores associated with lockdown policies. **b**, Distribution of changes in sentiment associated with the enforcement of a national lockdown for the 52 countries in our sample that enforced a lockdown during our sample period (see Methods, ‘Impacts of lockdowns on expressed sentiment’, for the definition). Supplementary analyses at the subnational level for the United States are included in Supplementary Note 6. The effect is measured by comparing the standardized sentiment index with the pre-COVID-19 sentiment average for each country. The countries are labelled with ISO 3166-1 alpha-3 country codes. The flag SVG assets, used under the CC-BY 4.0 license, are taken from the Emojitwo set: <https://emojitwo.github.io/>.

imposing stringent anti-contagion policies would lead to similar or even larger emotional distress.

Several previous studies have also documented complex emotional responses towards lockdown, and studies showing a negative association between lockdown and sentiment usually have not removed the impacts of the pandemic itself from lockdown measurements (see Supplementary Note 7 for a literature summary). Our analysis implies that lockdown policies do not necessarily entail a trade-off between physical health and emotional well-being—at least not for the average population of a country. It is worth noting that COVID-19 policy is not a clean setting for causal identification, since the anticipation effect before lockdown interventions and the spillover effects in sentiment across treated and control countries could bias our estimates. In addition, we do see substantial dispersion in sentiment change after lockdown across countries. Understanding the specific, contextual factors that produce these variations in effect sizes is an important avenue for future work.

Social media sentiment analysis provides complementary merits to those of survey measures for subjective well-being surveillance, but it has several limitations. First, the internet and social media penetration rates vary across countries and across different income and age groups within countries. Our analysis can only be used to understand the patterns of those who use Twitter or Weibo to communicate and lacks explanatory power for the least developed regions and elderly populations. Second, although social media expressed sentiment correlates with the affective aspects of subjective well-being, it cannot reliably measure the life satisfaction dimension of subjective well-being^{39,40}. Due to the limitations in representativeness and measurement, social media sentiment analysis should serve as a complement rather than a substitute for self-reported measures of subjective well-being. And more research is needed to understand the relationship between NLP-based sentiment and survey-based well-being in developing

countries. Third, as sentiment analysis using digital trace is still a nascent research area, we do not have enough evidence to judge whether our expressed sentiment measurements can be used to diagnose clinically meaningful mental disorders. More psychometric validations with self-reported mental health status will be required to understand to what extent expressed sentiment on social media can be used in psychiatric epidemiology. Fourth, our study mainly focuses on sentiment changes for the average population of social media users within a country and how country-level characteristics and policies moderate the effect. While these measures are meaningful at a macro level to understand global heterogeneity, we cannot measure the moderating effects of individual-level socio-demographics, beliefs and preferences, which limits our capacity to speak to disparities and potential tailored interventions for particular population subgroups. The technological progress in demographic inference tools based on social media data⁴¹ could enable further heterogeneity analysis at the individual or subgroup level, which could be an important research direction for future studies. Finally, our study covers only the first wave of the COVID-19 pandemic, and there are countries still recovering from their first waves. Rather than directly extrapolating our empirical results to inform future pandemic strategies, we recommend careful evaluations using our method and extended datasets for future waves.

Our data and methodology intend to provide a useful tool for tracking emotional well-being. This tool can support timely monitoring and decision-making by international and national policymakers.

Methods

Data. Social media data. We collected social media data from two large microblogging platforms, Twitter and Weibo (the Chinese equivalent of Twitter). The data cover the five months from 1 January 2020 (when COVID-19 spread was essentially restricted to China’s Wuhan region) to 31 May 2020 (when most countries had recovered from their first COVID-19 wave). Only geolocated social

media posts, for which users consented to share their location information, are of interest to our analysis. In the study period, 654 million geotagged Twitter and Weibo posts were collected globally (Fig. 1a).

Twitter is a global platform where users share content, or ‘tweets’, with their followers. As of 2019, Twitter had 330 million monthly active users. Users can give consent to share their location information by enabling background GPS collection or by tagging a location in their tweet. These geolocated tweets are encoded with latitude and longitude coordinates. We employed reverse-geocoding techniques to extract country information from geolocations. Twitter changed its geotagging approach in 2019 to enhance privacy protection⁴². However, because our analysis is conducted at the national or subnational level, this change in approach does not affect the state/country assignment of the individual. As Twitter’s counterpart in China, Weibo (Sina microblog) is one of the top social networking platforms, with 462 million monthly active users in 2020. The geotagged Weibo posts are a subsample of all Weibo posts for which users consented to share their location information. This location information is based on the user’s exact latitude and longitude when releasing the Weibo post from a smartphone or a computer. We filtered out the institutional accounts (including big ‘Vs’ (the most influential celebrities) in Weibo) and used only the individual’s original posts.

To measure people’s general emotional well-being rather than specific emotions towards COVID-19 itself, we excluded all COVID-19-related posts on the basis of an exhaustive list of COVID-19-related terms (Supplementary Fig. 19). During the implementation, we translated the COVID-19-related terms into the 30 most common languages to account for multilingual content. Posts where any one of these text patterns matched with the content were flagged as COVID-19 related.

Lockdown policy data. We collected and evaluated country-level lockdown policy data from two sources. The first was the Oxford Coronavirus Government Response Tracker (OxCGRT)¹⁴, and the second was the WHO Public Health and Social Measures (PHSM, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/phsm>). The former records 17 different government responses, including 8 related to containment and movement restrictions. Each government response is coded for response stringency (response-specific) and scope (‘targeted’ versus ‘general’). The latter joins seven policy databases together (including the OxCGRT) to provide a common taxonomy and comprehensive policy outlook at the national and subnational levels.

We consider the WHO PHSM to be comprehensive of all COVID-19 policy announcements because this dataset aggregates the efforts of seven separate COVID-19 policy databases. However, upon manual review of this dataset, we decided that its encoding of policy announcements into stringencies and measures was not sufficiently accurate for use as our final start and end order database. Instead, we used the WHO PHSM dataset to cross-validate the OxCGRT dataset by comparing all OxCGRT policy start dates, levels of enforcement and scopes with the WHO PHSM dataset’s announcements. When there was a discrepancy between the two datasets, we updated the OxCGRT dataset with the manually reviewed announcement from the WHO PHSM dataset.

Epidemiological data. The COVID-19 epidemiological data were collected from the COVID-19 dashboard by the Center for Systems Science and Engineering at Johns Hopkins University (<https://coronavirus.jhu.edu/map.html>).

Human mobility data. In 2020, Google provided a comprehensive COVID-19 Community Mobility Report (<https://www.google.com/covid19/mobility/>) showing the changes in mobility patterns in 135 countries from 15 February to 27 July 2020. The report displays the relative strengths of mobility indices in six different aspects (retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential) based on the baseline values, which are calculated by taking the median value of visits and length of stay in each type of place during five weeks before the COVID-19 pandemic.

Country-level indices. We collected a rich set of country-level indices to assist with the heterogeneity analysis of sentiment drop globally. The base years are different for each index due to differences in the most recent data availability. First, we collected a set of development indices. We collected gross domestic product per capita (2018), urbanization rate (2019) and unemployment rate (2019) from the World Bank; these are the commonly used indicators to represent a country’s development status⁴³. We also collected the 2017 Socio-demographic Index, a comparative summary socio-demographic metric synthesizing per capita income, educational attainment and fertility rate⁴⁴, which is commonly used to explain the disparities in countries’ burdens of disease⁴⁵. Second, we collected two indices related to a country’s management capacity: government efficiency index (2018) from the World Bank and the 2015 Global Health Security Index (<https://www.ghsindex.org/>). Government efficiency is a comprehensive measure of public sectors’ performance and has been proven to predict a country’s capacity to control the COVID-19 pandemic⁴⁶. The Global Health Security Index is the first comprehensive assessment of health security at the global scale⁴⁷. Finally, we collected culture-related indices from Awad et al.³⁵. The key culture variable of interest is cultural tightness, which measures the tightness of social norms and was found to explain a country’s capacity

to manage the COVID-19 pandemic⁴⁶. The cultural indices in this dataset also include individualism, religiousness and relational mobility (that is, the fluidity with which people can develop new relationships).

Sentiment analysis. We employed NLP sentiment imputation algorithms to analyse the million daily social media posts that make up our dataset. Different types of NLP methods have been used for sentiment classification of textual data in the existing literature. Dictionary-based approaches match the words that make up each text entry to sentiment-specific lists (or dictionaries). LIWC²⁶ and the Hedonometer project⁴⁸ offer two such dictionaries and have often been used in social media research⁴⁹. More recently, sentiment classification of text data has been successfully implemented using neural networks such as transformers or convolutional neural networks⁵⁰. These methods create high-dimensional representations of the text entries, usually based on pre-trained word vectors. For this study, we used a transformer that has achieved state-of-the-art results in text classification: BERT²⁵. Unlike traditional word2vec text representation models, BERT creates dynamic word representations informed locally by the neighbouring context. Our global study uses a pre-trained Multilingual BERT model, which creates representations broadly consistent across different languages²⁷. Sentence-BERT provides an additional document-level embedding. On the basis of Siamese-coupled neural networks, this model produces semantically meaningful representations of sentences that can be compared among themselves in the embedding space⁵¹. For our study, we created these high-dimensional representations for every social media entry in our dataset.

We trained a simple logistic-regression classifier on the first 100 principal component analysis dimensions of the Sentence-BERT social media post embeddings. The training data we used are a set of 1,600,000 tweets labelled as positive or negative²⁵. Since representations are consistent across languages, we were able to train our sentiment classifier in English and predict sentiment in the 104 languages supported by Multilingual BERT²⁷, which covers 65 identifiable languages in Twitter and Weibo. We evaluate the performance of this model in Supplementary Note 1 and find a classification accuracy of 0.84 for English content and 0.75 on average in other languages (see the details in Supplementary Table 1). We further compared the sentiment from our BERT-based algorithm with sentiment indices from the dictionary-based LIWC method using English tweets, and the results show high consistency (Supplementary Fig. 1). To enhance the transparency of our algorithm, we display how people changed their use of emotional words (defined by the LIWC English dictionary) accompanied with the decline in our sentiment index at the onset of the COVID-19 pandemic in Supplementary Note 3.

We averaged each social media post sentiment score daily at the national and subnational levels (for example, state or province; the largest subnational administrative unit of a country). To avoid oversampling individuals who post the most on social media, we first aggregated our sentiment data to the individual–date level and then averaged the individual sentiments on each day to the subnational (state/province) or national level. Moreover, we used the one-class classification approach⁵³ to detect and exclude Twitter bots.

Sentiment alterations during the COVID-19 pandemic. *Modelling of sentiment dynamics.* To measure the ability of sentiment recovery and each country’s recovery status, we adopted the following procedures:

- Select countries.** We only kept countries in which tweets were generated by more than 100 active users a day from 1 January to 31 May 2020, to ensure that extreme observations and insufficient data do not severely impact a country’s sentiment index.
- Detrend.** To extract the overall sentiment trend out of daily sentiment fluctuations, we implemented seasonal trend decomposition using a locally estimated scatterplot smoothing algorithm⁵⁴ with seven days as a feeding parameter for temporal cycles to remove seven-day periodical patterns from the raw data and average the fluctuations (Supplementary Fig. 9). The seven-day period captures weekly circular patterns and is further confirmed when applying the Fourier transformation algorithm to detect temporal cycles of sentiment.
- Identify drop-date and min-date.** Drop-date is defined as the day when a country’s sentiment started to be affected by COVID-19, while min-date is when a country reached its minimum sentiment. Though most countries have only one sentiment nadir caused by COVID-19, a few countries have sentiment drops due to other events coinciding with the COVID-19 period. To locate the period related to COVID-19, we introduced T-max, the date on which the share of COVID-19-related tweets reached the maximum proportion among all tweets. To get sentiment min-date, we searched for the date on which a country’s sentiment reached its minimum value within 30 days of the country’s T-max. For sentiment drop-date, we searched for the date on which the share of COVID-19-related tweets reached 20% of the maximum share achieved on T-max. This is the moment when people started to pay attention to COVID-19 topics within a country. The results from the cumulative sum⁵⁵ test on a country’s sentiment curve confirmed that this is the approximate time point when a country’s sentiment started to decline.

Sentiment drops. To estimate the effect of COVID-19 on sentiment drop, we exploited the ‘donut regression discontinuity’ design⁵⁶ at the timing threshold of min-date, t^* , by removing the confounding days between drop-date and min-date and measuring the sentiment discontinuity from the level before drop-date to that after min-date (Supplementary Fig. 11). We use time (daily) as the running variable, and the estimate for sentiment shock (τ_{RD}) we use is:

$$\tau_{RD} = \lim_{date \uparrow t^*} E[\text{sentiment}_i | \text{date}_i = t^*] - \lim_{date \downarrow t^*} E[\text{sentiment}_i | \text{date}_i = t^*] \quad (1)$$

In the equation, E is the expectation value. This RDD is equivalent to segmented regressions used for interrupted time series analysis in public health research^{57,58}. This quasi-experimental design is particularly useful in cases where an abrupt event causes all units to be treated, and there are potential confounding time trends pre- and post-treatment. To calculate the sentiment limit on both sides of the running variable, we used local linear regression to fit the sentiment curve on each side of the threshold t^* . Local linear instead of higher-order polynomial fit is recommended by previous literature⁵⁹. Fitting the general trends pre- and post-interruption can prevent one day with extreme sentiment values to substantially bias the magnitude of sentiment drop. In practice, we estimated the equation for each country separately with administration (province or state level) sentiment time series as input:

$$y_{sit} = \beta \text{COVID}_{it} + \gamma_1 \text{rel_date}_{sit} + \gamma_2 \text{rel_date}_{sit} \times \text{COVID}_{it} + \delta_{DOW} + \eta_s + \varepsilon_{sit} \quad (2)$$

In the equation, y_{sit} is the average sentiment index for province/state s in country i on date t , COVID_{it} is a binary variable equal to 1 when the time is after the country’s sentiment nadir and 0 otherwise, and rel_date_{sit} is the date measured in days from the minimum sentiment date t^* . The terms $\gamma_1 \text{rel_date}_{sit}$ and $\gamma_2 \text{rel_date}_{sit} \times \text{COVID}_{it}$ absorb the smooth relationship of sentiment trend within the bandwidth surrounding t^* . We used a bandwidth of 28 days on each side and the triangular kernel as our base specification. As robustness checks, we also tested the results for the uniform kernel as well as for different bandwidths, which yields similar results (Supplementary Table 6 and Supplementary Fig. 7). In all cases, we weighed the regression by the number of tweets so that larger provinces/states have higher weights. We added day-of-week fixed effects (δ_{DOW}) and state fixed effects (η_s) to control for weekly cyclical and time-invariant state-specific confounding factors; ε_{sit} is the error term. β is our coefficient of interest, the magnitude of which is divided by the standard deviation of sentiment pre-COVID-19 (that is, before the sentiment drop-date of each country) to make the results more comparable across countries. The standard errors are clustered within province/state to account for sentiment correlation.

Sentiment recovery. To characterize the recovery of each country, we established two indices: recovery half-life and recovery status. Recovery half-life represents a country’s recovery speed, while recovery status represents to what degree a country’s sentiment had recovered at the end of May 2020. Following Fan et al.³², we parametrized the sentiment recovery process through an exponential model and estimated the parameters u , v and γ with nonlinear least squares. More specifically, we regress the daily sentiment value $f(x)$ on x , which captures the number of days since the country achieved the minimum level of sentiment using the following exponential function:

$$f(x) = u + ve^{-\gamma x} \quad (3)$$

To ensure that the sentiment we measured was free from impacts of the Black Lives Matter campaigns, we set 25 May 2020 (that is, when the George Floyd event took place) as the end date of our sentiment analysis. We removed all countries with abnormal sentiment fluctuations on which the parameter calibration algorithm could not converge within 1,000 steps during the fitting process to ensure quality.

We then identified the two recovery indices using the fitted exponential model. To find the recovery half-life, we searched for the date on the fitted curve where the sentiment recovered 50% of the distance between a country’s sentiment nadir and its final sentiment on 25 May. To further understand the recovery status, we compared a country’s sentiment status on 25 May with the baseline level before the sentiment drop-date and defined it with pre-COVID-19 sentiment standard deviation as a unit to ensure comparability across countries (for example, -1 indicates recovering to a status 1 s.d. below the baseline sentiment).

Impacts of lockdowns on expressed sentiment. When defining lockdowns, we refer to the ‘stay-at-home requirements’ policy category of OxCGRT⁶⁰. A country is defined as a lockdown country if it has national-level requirements on not leaving the house except essential trips (that is, levels 2 and 3 of the C6 policy category in OxCGRT). Our cross-validation process was summarized in the ‘Data’

section. The lockdown dates compare with the sentiment drop, and min dates are summarized in Supplementary Fig. 14.

Here we applied SCM³⁶ to estimate sentiment alterations after lockdown interventions for each country separately. In the case that no single country alone provided a good comparison for the lockdown country of interest (that is, violating the parallel trend assumptions required for difference-in-differences or event studies), we constructed a combination of non-lockdown countries as a synthetic control group to best resemble the characteristics of the treated country before its lockdown. We then compared the sentiment of a treated country on days after the lockdown with the weighted average sentiment of the control countries at the same period to estimate the treatment effect. The weights assigned to each control country were calculated such that the simulated synthetic control best resembles the treated country of interest in the pre-lockdown period. Mathematically, the distance between the vector of pre-specified characteristics of the treated country and that of the weighted average controls is minimized before the lockdown time^{36,61}.

The validity of synthetic control relies heavily on the countries included in the ‘donor pool’ and the observable characteristics that the weights of synthetic control are built on. We allowed the ‘late adopters’ of lockdown policies to serve as controls for the ‘early adopters’ to increase the similarity between the treatment country and its donor pool. This approach does not change the results for late adopters and those early adopters for which appropriate comparisons can be constructed from the never adopters (for example, northern European countries rely heavily on Sweden as control); but it can provide a better simulation of the comparison scenario of early adopters with no comparable countries that never implemented lockdown policies. We only explored one-week post-lockdown in the SCM analysis so that late adopters implemented more than seven days after a specific treated country could be matched to enhance comparability between a treatment country and its synthetic control country. Meanwhile, we excluded countries with many subnational-level lockdowns, including the United States, China, Nigeria, Brazil and Germany. We also included four layers of socio-economic and pandemic-related variables in the covariates to ensure that higher weights were assigned to similar countries. Details about the variables included as covariates and the associated robustness checks are presented in Supplementary Note 4.

For the identification of SCM to be causal, we need to assume that the choice of which unit will be treated is random conditional on the choice of the donor pool, the observable variables included as predictors and the unobserved factors that can be captured by the pre-treatment path of the outcome variable⁶². It is hard to test this assumption directly, and COVID-19 policies might generate anticipation effects and spillover effects for some countries; care is thus warranted for interpreting these estimates as causal.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data used in this paper are available at https://github.com/Jianghao/Sentiment_COVID-19.

Code availability

The code used in this paper is available at https://github.com/Jianghao/Sentiment_COVID-19.

Received: 11 February 2021; Accepted: 20 January 2022; Published online: 17 March 2022

References

- Clark, A. et al. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. *Lancet Glob. Health* **8**, e1003–e1017 (2020).
- Gunnell, D. et al. Suicide risk and prevention during the COVID-19 pandemic. *Lancet Psychiatry* **7**, 468–471 (2020).
- Pfefferbaum, B. & North, C. S. Mental health and the COVID-19 pandemic. *N. Engl. J. Med.* **383**, 510–512 (2020).
- Sibley, C. G. et al. Effects of the COVID-19 pandemic and nationwide lockdown on trust, attitudes toward government, and well-being. *Am. Psychol.* **75**, 618–630 (2020).
- Brooks, S. K. et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *Lancet* **395**, 912–920 (2020).
- Hao, B., Li, L., Gao, R., Li, A. & Zhu, T. in *Active Media Technology* (eds Ślęzak, D. et al.) 324–335 (Springer International Publishing, 2014).
- Diener, E., Oishi, S. & Tay, L. Advances in subjective well-being research. *Nat. Hum. Behav.* **2**, 253–260 (2018).

8. Exton, C. & Shinwell, M. *Policy Use of Well-Being Metrics* (OECD, 2018); <https://doi.org/10.1787/d98eb8ed-en>
9. Jaidka, K. et al. Estimating geographic subjective well-being from Twitter: a comparison of dictionary and data-driven language methods. *Proc. Natl Acad. Sci. USA* **117**, 10165–10171 (2020).
10. Luhmann, M., Hawkey, L. C., Eid, M. & Cacioppo, J. T. Time frames and the distinction between affective and cognitive well-being. *J. Res. Pers.* **46**, 431–441 (2012).
11. Holmes, E. A. et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *Lancet Psychiatry* **7**, 547–560 (2020).
12. Lima, C. K. T. et al. The emotional impact of coronavirus 2019-nCoV (new coronavirus disease). *Psychiatry Res.* **287**, 112915 (2020).
13. Cheng, C., Barceló, J., Hartnett, A. S., Kubinec, R. & Messerschmidt, L. COVID-19 government response event dataset (CoronaNet v.1.0). *Nat. Hum. Behav.* **4**, 756–768 (2020).
14. Hale, T. et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat. Hum. Behav.* **5**, 529–538 (2021).
15. Luhmann, M. Using Big Data to study subjective well-being. *Curr. Opin. Behav. Sci.* **18**, 28–33 (2017).
16. Betsch, C. et al. Germany COVID-19 Snapshot MONitoring (COSMO Germany): monitoring knowledge, risk perceptions, preventive behaviours, and public trust in the current coronavirus outbreak in Germany. Preprint at *PsychArchives* <https://doi.org/10.23668/PSYCHARCHIVES.2776> (2020).
17. Ji, X., Chun, S. A., Wei, Z. & Geller, J. Twitter sentiment classification for measuring public health concerns. *Soc. Netw. Anal. Min.* **5**, 13 (2015).
18. Mohammad, S. M. in *Emotion Measurement* (ed. Meiselman, H. L.) 201–237 (Woodhead Publishing, 2016).
19. Baylis, P. Temperature and temperament: evidence from Twitter. *J. Public Econ.* **184**, 104161 (2020).
20. Wang, J., Obradovich, N. & Zheng, S. A 43-million-person investigation into weather and expressed sentiment in a changing climate. *One Earth* **2**, 568–577 (2020).
21. Zheng, S., Wang, J., Sun, C., Zhang, X. & Kahn, M. E. Air pollution lowers Chinese urbanites' expressed happiness on social media. *Nat. Hum. Behav.* **3**, 237–243 (2019).
22. Kryvasheyeu, Y. et al. Rapid assessment of disaster damage using social media activity. *Sci. Adv.* **2**, e1500779 (2016).
23. Miura, A., Komori, M., Matsumura, N. & Maeda, K. Expression of negative emotional responses to the 2011 Great East Japan Earthquake: analysis of big data from social media. *Shinrigaku Kenkyu* **86**, 102–111 (2015).
24. Roy, A. et al. A machine learning approach predicts future risk to suicidal ideation from social media data. *NPJ Digit. Med.* **3**, 78 (2020).
25. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. Preprint at *arXiv* <https://arxiv.org/abs/1810.04805> (2018).
26. Pennebaker, J. W., Francis, M. E., & Booth, R. J. *Linguistic Inquiry and Word Count: LIWC 2001* (Lawrence Erlbaum Associates, 2001).
27. Pires, T., Schlinger, E. & Garrette, D. How multilingual is Multilingual BERT? Preprint at *arXiv* <https://arxiv.org/abs/1906.01502> (2019).
28. Rattay, P. et al. Differences in risk perception, knowledge and protective behaviour regarding COVID-19 by education level among women and men in Germany: results from the COVID-19 Snapshot Monitoring (COSMO) study. *PLoS ONE* **16**, e0251694 (2021).
29. Sun, A. & Zhao, Y. Divorce, abortion, and the child sex ratio: the impact of divorce reform in China. *J. Dev. Econ.* **120**, 53–69 (2016).
30. Anderson, M. L. Subways, strikes, and slowdowns: the impacts of public transit on traffic congestion. *Am. Econ. Rev.* **104**, 2763–2796 (2014).
31. Hausman, C. & Rapson, D. S. Regression discontinuity in time: considerations for empirical applications. *Annu. Rev. Resour. Econ.* **10**, 533–552 (2018).
32. Fan, R. et al. The minute-scale dynamics of online emotions reveal the effects of affect labeling. *Nat. Hum. Behav.* **3**, 92–100 (2019).
33. Mula, S., Di Santo, D., Gelfand, M. J., Cabras, C. & Pierro, A. The mediational role of desire for cultural tightness on concern with COVID-19 and perceived self-control. *Front. Psychol.* **12**, 713952 (2021).
34. Van Bavel, J. J. et al. Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **4**, 460–471 (2020).
35. Awad, E., Dsouza, S., Shariff, A., Rahwan, I. & Bonnefon, J.-F. Universals and variations in moral decisions made in 42 countries by 70,000 participants. *Proc. Natl Acad. Sci. USA* **117**, 2332–2337 (2020).
36. Abadie, A. Using synthetic controls: feasibility, data requirements, and methodological aspects. *J. Econ. Lit.* **59**, 391–425 (2019).
37. Settanni, M. & Marengo, D. Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. *Front. Psychol.* **6**, 1045 (2015).
38. Kahn, J. H., Tobin, R. M., Massey, A. E. & Anderson, J. A. Measuring emotional expression with the Linguistic Inquiry and Word Count. *Am. J. Psychol.* **120**, 263–286 (2007).
39. Liu, P., Tov, W., Kosinski, M., Stillwell, D. J. & Qiu, L. Do Facebook status updates reflect subjective well-being? *Cyberpsychol. Behav. Soc. Netw.* **18**, 373–379 (2015).
40. Wang, N., Kosinski, M., Stillwell, D. J. & Rust, J. Can well-being be measured using Facebook status updates? Validation of Facebook's gross national happiness index. *Soc. Indic. Res.* **115**, 483–491 (2014).
41. Wang, Z. et al. Demographic inference and representative population estimates from multilingual social media data. In *WWW '19: The World Wide Web Conference* (eds Liu, L. & White, R.) 2056–2067 (ACM, 2019); <https://doi.org/10.1145/3308558.3313684>
42. Hu, Y. & Wang, R.-Q. Understanding the removal of precise geotagging in tweets. *Nat. Hum. Behav.* **4**, 1219–1221 (2020).
43. Chen, M., Zhang, H., Liu, W. & Zhang, W. The global pattern of urbanization and economic growth: evidence from the last three decades. *PLoS ONE* **9**, e103799 (2014).
44. Kraemer, M. U. G. et al. Mapping global variation in human mobility. *Nat. Hum. Behav.* **4**, 800–810 (2020).
45. Hu, K. et al. Global patterns and trends in the breast cancer incidence and mortality according to sociodemographic indices: an observational study based on the global burden of diseases. *BMJ Open* **9**, e028461 (2019).
46. Gelfand, M. J. et al. The relationship between cultural tightness-looseness and COVID-19 cases and deaths: a global analysis. *Lancet Planet. Health* **5**, e135–e144 (2021).
47. Razavi, A., Erondu, N. & Okereke, E. The Global Health Security Index: what value does it add? *BMJ Glob. Health* **5**, e002477 (2020).
48. Dodds, P. S. et al. Human language reveals a universal positivity bias. *Proc. Natl Acad. Sci. USA* **112**, 2389–2394 (2015).
49. Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H. & Eichstaedt, J. C. Detecting depression and mental illness on social media: an integrative review. *Curr. Opin. Behav. Sci.* **18**, 43–49 (2017).
50. Kim, Y. Convolutional neural networks for sentence classification. Preprint at *arXiv* <https://arxiv.org/abs/1408.5882> (2014).
51. Reimers, N. & Gurevych, I. Sentence-BERT: sentence embeddings using Siamese BERT-networks. Preprint at *arXiv* <https://arxiv.org/abs/1908.10084> (2019).
52. Sahni, T., Chandak, C., Chedeti, N. R. & Singh, M. Efficient Twitter sentiment classification using subjective distant supervision. Preprint at *arXiv* <https://arxiv.org/abs/1701.03051> (2017).
53. Rodríguez-Ruiz, J., Mata-Sánchez, J. I., Monroy, R., Loyola-González, O. & López-Cuevas, A. A one-class classification approach for bot detection on Twitter. *Comput. Secur.* **91**, 101715 (2020).
54. Rojo, J., Rivero, R., Romero-Morte, J., Fernández-González, F. & Pérez-Badia, R. Modeling pollen time series using seasonal-trend decomposition procedure based on LOESS smoothing. *Int. J. Biometeorol.* **61**, 335–348 (2017).
55. Grigg, O. A., Farewell, V. T. & Spiegelhalter, D. J. Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. *Stat. Methods Med. Res.* **12**, 147–170 (2003).
56. Barreca, A. I., Guldi, M., Lindo, J. M. & Waddell, G. R. Saving babies? Revisiting the effect of very low birth weight classification. *Q. J. Econ.* **126**, 2117–2223 (2011).
57. Kontopantelis, E., Doran, T., Springate, D. A., Buchan, I. & Reeves, D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *Brit. Med. J.* **350**, h2750 (2015).
58. Lopez Bernal, J., Cummins, S. & Gasparrini, A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int. J. Epidemiol.* **46**, dyw098 (2016).
59. Cattaneo, M. D., Idrobo, N. & Titiunik, R. A practical introduction to regression discontinuity designs: foundations. Preprint at *arXiv* <https://arxiv.org/abs/1911.09511> (2019).
60. Hale, T. et al. *Variation in Government Responses to COVID-19*. Version 6.0, Working Paper (Blavatnik School of Government, 2020).
61. Cavallo, E., Galiani, S., Noy, I. & Pantano, J. Catastrophic natural disasters and economic growth. *Rev. Econ. Stat.* **95**, 1549–1561 (2013).
62. Firpo, S. & Possebom, V. Synthetic control method: inference, sensitivity analysis and confidence sets. *J. Causal Inference* **6**, 1–26 (2018).

Acknowledgements

We thank Z. Cheng, A. Mino and E. Trieschman for their excellent research assistance. S.Z. acknowledges research support from the Massachusetts Consortium on Pathogen

Readiness (MassCPR). J.W. acknowledges research support from the National Natural Science Foundation of China (grant nos 41971409 and 41421001) and the Youth Innovation Promotion Association of the Chinese Academy of Sciences (grant no. 2020052). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

J.W., S.Z., Y.F., J.P. and C.Z. designed the project. J.W., Y.F., Y.C. and N.G.-J. collected and analysed the data. J.W., Y.F., J.P., Y.C. and N.O. interpreted the results. Y.F. and J.W. wrote the manuscript. J.P., Y.C., N.G.-J., N.O., C.Z. and S.Z. edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41562-022-01312-y>.

Correspondence and requests for materials should be addressed to Chenghu Zhou or Siqi Zheng.

Peer review information *Nature Human Behaviour* thanks Peter Craig, Anna Kruspe and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

We collected 654 million geolocated social media posts from two large microblog platforms, Twitter and Weibo (Chinese equivalent of Twitter). The data covers the five months from January 1, 2020 to May 31, 2020. Data used in this paper are available at https://github.com/Jianghao/Sentiment_COVID-19.

Data analysis

R 4.0, Python 3.7 and Stata 14.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data used in this paper are available at https://github.com/Jianghao/Sentiment_COVID-19.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Using tweets in over 100 countries, we develop a global sentiment dataset during the COVID-19. We conduct two inter-related empirical exercises to evaluate the global affective impacts of the COVID-19 pandemic and policy responses.
Research sample	We collected 654 million geotagged Twitter and Weibo posts to develop a global dataset of expressed sentiment indices to track national and subnational level affective states on a daily basis.
Sampling strategy	No statistical methods were used to pre-determine sample sizes. We try our best to collect the largest possible sample (654 million geotagged posts worldwide) to avoid sampling bias.
Data collection	We collected social media data from Twitter and Weibo using web crawler technology that continuously connected to the streaming application programming interface (API) during our study period.
Timing	The data covers five months, from January 1, 2020 to May 31, 2020.
Data exclusions	Since the sample of COVID-19 related discussions might not be a good representation of the affective state of the general population and could be polluted by political campaigns, we exclude tweets directly related to the COVID-19 when building out main sentiment indices.
Non-participation	NA
Randomization	NA

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging