Bertani et al. *EPJ Data Science* (2024) 13:20
https://doi.org/10.1140/epjds/s13688-024-00452-7

EPJ Data Science
a SpringerOpen Journal

EPJ.org

RESEARCH Open Access

# Large-scale digital signatures of emotional response to the COVID-19 vaccination campaign

Anna Bertani[1,2*], Riccardo Gallotti[2], Stefano Menini[3], Pierluigi Sacco[4,5] and Manlio De Domenico[6]

*Correspondence:
anna.bertani@unitn.it
[1]Department of Information Engineering and Computer Science, University of Trento, Via Sommarive, 9, 38123 Povo (TN), Italy
[2]CHuB Lab, Fondazione Bruno Kessler, Via Sommarive, 18, 38123 Povo (TN), Italy
Full list of author information is available at the end of the article

**Abstract**

The same individuals can express very different emotions in online social media with respect to face-to-face interactions, partially because of intrinsic limitations of the digital environments and partially because of their algorithmic design, which is optimized to maximize engagement. Such differences become even more pronounced for topics concerning socially sensitive and polarizing issues, such as massive pharmaceutical interventions. Here, we investigate how online emotional responses change during the large-scale COVID-19 vaccination campaign with respect to a baseline in which no specific contentious topic dominates. We show that the online discussions during the pandemic generate a vast spectrum of emotional response compared to the baseline, especially when we take into account the characteristics of the users and the type of information shared in the online platform. Furthermore, we analyze the role of the political orientation of shared news, whose circulation seems to be driven not only by their actual informational content but also by the social need to strengthen one's affiliation to, and positioning within, a specific online community by means of emotionally arousing posts. Our findings stress the importance of better understanding the emotional reactions to contentious topics at scale from digital signatures, while providing a more quantitative assessment of the ongoing online social dynamics to build a faithful picture of offline social implications.

**Keywords:** Computational social science; Socio-technical systems; Exceptional events; COVID-19 vaccination; Emotions

## 1 Introduction

Face-to-face interaction is notoriously important to facilitate civilized exchange and social cooperation [1, 2]. Through their nonverbal language [3], interacting humans send complex arrays of signals (of dominance, trust, composure etc.) [4] that favor mutual alignment [5] and even elicit behavioral mimicry [6]. For these reasons, face-to-face interactions are normally expected not to escalate into violent and confrontational behaviors [7], and even function as a driver of social cohesion [8] with distinctive psycho-physiological signatures [9]. With the advent of online interactions, however, this carefully evolved package of socio-cognitive skills has been put to a hard test. In digital interaction, the moderating



Springer

role of nonverbal cues is largely lost [10], and subjects must learn how the affordances of digital communication enable alternative ways to signal pro-sociality and affection [11]. However, developing alternative systems that work on a large social scale is challenging, and as a result it is widely observed that digitally based exchanges have higher chances to become vitriolic and prone to escalation than face-to-face ones [12].

In this context, a particularly crucial role is played by emotions. The once widely held conviction that human emotions were biologically programmed, universal 'natural kinds', spanned by a core group of six basic emotions and characterized by specific, interculturally readable sets of bodily cues, has gradually given way to an alternative paradigm that identifies a whole spectrum of emotional states mapped by a high-dimensional set of verbal and nonverbal signals [13] whose socio-cognitive indexing depends on social learning and cultural frames [14]. In view of the central role of emotions in social cognition [15], the use and interpretation of such emotionally-related signals is pivotal in human interaction. In particular, the expression and decoding of emotions is likely to be one of the most critical aspects related to the shift from face-to-face to online interactions [16].

If emotions are not 'natural kinds', a radical change in the socio-cognitive environment such as that brought about by the digital shift may not only affect how emotions are interpreted, but also how they are defined [17], expressed [18], and socially transmitted [19]. Despite that the affordances of the digital environment facilitate emotional over-reaction [20], it is also true that, unlike face-to-face interaction where nonverbal signals are observed at a millisecond scale and are largely automatic and not controlled [21], in online interaction the emotional response may be distilled and even constructed on the scale of seconds and even minutes or hours [22]. On the other hand, nonverbal cues are largely substituted by verbal equivalents that intentionally amplify the affective content of communication to compensate for the lower-dimensional nature of the signal [23]. Moreover, in online interactions some of the participants may be nonhuman (bots) which are explicitly designed to implement a certain strategy of affective communication to elicit types of emotional responses from users [24].

There is then a basic difference between emotions expressed in face-to-face interaction and 'emotions' as constructed in online exchanges [25]. And such difference is likely to be more substantial and relevant the more the topic of the exchange concerns socially sensitive, polarizing issues [26, 27]. In particular, fake or misleading content that embeds certain emotional references (which can be positive such as anticipation or trust, or negative such as anger, according to cases) is more likely to go viral and has a longer expected lifetime [28].

Moving from this premise, the aim of our paper is to investigate how online 'emotional' responses change when comparing two different kinds of exchanges: one regarding a baseline, non-polarizing topic and one regarding one of the most polarizing topics of today: vaccination [29]. It is intuitive to conjecture that more socially controversial issues incite more emotional reaction (and vice versa) online than less controversial ones [30], not unlike what happens offline, although specific features of the online interaction environment may make a significant difference [31]. However, it is of interest to understand what kind of emotional reaction is incited in what circumstances, depending on the characteristics of the users involved. The emotional response patterns that are found give us additional insight on the nature of the social contentiousness of the issue, and at the same time help us understand better how online 'emotions' are constructed to pursue specific

social goals. In this regard, vaccination is clearly a relevant test bed [32], and we are in particular interested in studying social exchanges on Twitter with special attention to the verified/un-verified user and human/bot dyads. To what extent the emotional communication of verified users will distinguish itself from that of un-verified ones? Will there be major differences between human and artificial 'emotional' responses when the topic is potentially more inflammatory, and in which sense?

## 2 Related Work

There is a rapidly growing literature that is exploring various methodological approaches to the measurement of emotions and of their socio-behavioral effects in online interactions, with special emphasis on contentious and polarizing topics. The work presented in [33] introduces a method using change point detection and topic modeling to measure online emotional reactions to offline polarizing events such as the COVID-19 pandemic and racial justice protests, effectively dis-aggregating different topics to measure the emotional and moral reactions of the public opinion.

Guo et al. [34] use natural language processing methods to highlight a surprising rise in positive affect during the early stages of the COVID-19 pandemic, likely reflecting the role of social media as a tool for emotion regulation and reducing fear of the unknown during the pandemic, while also revealing a partisan divide in emotional and moral reactions. Vemprala et al. [35] also develop a natural language processing approach to the study of public negative emotional responses during the COVID-19 pandemic to discern how emotional patterns react to the prevailing focus of the public discussion being placed upon health vs. economic concerns, with fear dominating in health-related conversations, and a more complex mix of emotions, mainly anger and fear, emerging in economics-related ones. Wang et al. [36] also develop a natural language processing approach based upon a BERT transformer model for sentiment classification to show that the COVID-19 pandemic caused a significant drop in expressed sentiment globally, followed by an asymmetric, slower recovery, with minor effects of sentiment expression related to lockdown policies, although with significant variation across countries. Zhang et al. [37] use instead a machine learning classification model based upon deep learning language models to identify positive and negative emotions in online conversations about COVID-19, and including an additional, so far not analyzed, ambivalent emotional expression (joking), finding a rapid burst and a slower decline in the online conversations in all of the 6 languages they analyze.

Although a comprehensive review of this literature is beyond the scope of the present paper, these few examples suffice to show the richness of the methodological and analytical contributions of computational social science approaches to the emotional analysis of online interactions on polarizing topics and especially of COVID-19 related ones.

In this rapidly evolving field of research, the specific contribution of our paper is making use again of a natural language processing approach and of the same lexicon used by [35], NRC Word-Emotion Association Lexicon, to analyze specifically the differences in emotional expression between a baseline non-contentious topic and a highly polarizing one such as COVID-19 vaccination, and testing specifically for differences related to whether online participants are humans or bots.

The NRC Lexicon considers eight 'emotions', four of which with positive valence and four with negative valence. With this choice, we intentionally move from the six basic emotions

of Ekman [38] not because we agree with the 'natural kinds' framework (within which there is a significant lack of consensus about the specific list of what emotions are basic; see [39]), but because this may be a simple benchmark for the analysis of their online counterparts. Specifically, the four 'basic' negative emotions (fear, anger, sadness, and disgust) and one of the positive ones (surprise) are kept in the NRC Lexicon, whereas the other (happiness) is unpacked into two positive emotions (anticipation and joy). Finally, trust has been added as a final positive emotion. Despite that trust is technically not considered an emotion [40], it can be interesting to consider a specific emotional signal related to trust in the relatively unfavorable conditions given by social signaling through a digital medium. This set of 'emotions' as encapsulated in the NRC Lexicon should therefore be seen as a useful first benchmark rather than as an invitation to consider them as more basic or foundational than others.

Within this framework, we find a significant difference in emotional reactions between no-contentious and polarizing conversations, and moreover we find that verified users, no matter whether humans of bots, exhibit more positively valenced emotional responses than unverified ones. In this regard, our paper provides fresh insights on specific aspects that have not been covered in the previous literature, while at the same time connecting to a solid ad growing stream of studies both in methodological and thematic terms.

Getting deeper insights into these issues is crucial for the future design of environments that favor more civilized online interaction. As argued by [41], civilized need not amount to 'polite' according to pre-digital standards of social etiquette. Online discussion may be more emotionally charged than offline ones for reasons that mostly concern the different affordances of the social environments [42]. However, as the experience of the pandemic has taught us, the social implications of massive uncivilized exchange on issues of primary public interest may be devastating [43] and may offer ample opportunities for manipulation by malevolent parties [44]. Therefore, a deeper understanding of the socio-emotional grammar of online interactions is a key issue for both computational social science and public policy.

## 3 Method

### 3.1 Overview of the data set

We collected social media data through the special Twitter's endpoint dedicated to COVID-19 research,[1] which allowed researchers to study the comprehensive, public conversation about COVID-19 in real-time. We focused our attention on data captured by the filter of the Twitter Firehose on COVID-19 in 18 among the most represented languages on Twitter. More specifically, we focused on terms related to the vaccination, to anti-vax campaigns but also to the most known vaccine brands, such as Pfizer, Astrazeneca, Johnson&Johnson, Moderna, Sputnik V (see further details in Additional file 1 Appendix).

In addition, we considered only a small fraction of the overall data, the about 1% of tweets that are geotagged, to guarantee an accurate information also as to their location, as signalled by the user's device. The data collected covers the period between August 31, 2020 and July 15, 2021, that is from the announcement of the availability of the first COVID-19 vaccines up to the peak of the vaccine campaign in Europe and in the United States. We compare our sample of tweets related to the vaccination topics with a 10 million

[1] https://blog.twitter.com/developer/en_us/topics/tools/2020/covid19_public_conversation_data

**Table 1** Statistics about the datasets. The table shows the number of messages, unique users and the timeframe considered for both the datasets

| Dataset | Messages | Unique Users | Timeframe |
|---|---|---|---|
| Vaccine | 9.6 million | 3 million | 31/08/2020-15/07/2021 |
| Baseline | 10 million | 6 million | 21/04/2021-24/04/2021 |

of messages posted on Twitter as baseline sample in which no specific contentious topic dominates, as shown in Table 1. The baseline sample was obtained using the stream API without specific filtering or keyword searches. As a result, the dataset encompasses a diverse range of languages, including but not limited to English, Japanese, Spanish, Korean, Portuguese, Thai, Indian languages, French, German, and Italian, among others.

### 3.2 Verified accounts

Verified users are those having the blue verified badge, a blue check mark, that defines those accounts that are of public interest because they are considered authentic, notable and active on Twitter. The verification process of users is given by the blue check mark that can be found next to the username, while unverified accounts do not have this distinctive signal. However, this definition pertains to the period before Musk's takeover, during which accounts were required to undergo request verification. Starting from April 1st, 2023, there has been a change in the rule. For users to acquire the verification badge now, they are required to subscribe to Twitter Blue [45]. Regarding this research, we adhere to the initial definition since the new rule was not in effect during the considered time frame.

### 3.3 News reliability and political leaning

In this analysis, we also considered some metadata associated to the textual content of the messages. We collected manually checked web domains from various publicly accessible databases, encompassing scientific and journalistic sources. In particular, we examined data provided by the MediaBiasFactCheck [46]. This is an organization that provide a huge database continuously updated whose methodology is to systematically evaluate the ideological leanings and factual accuracy of media and information outlets through a multi-faceted approach that incorporates both quantitative metrics and qualitative assessments. We found a total of 4988 domains, reduced to 4417 after removing hard duplicates across databases. Given the nature of our multilingual and multicultural analysis, we evaluated the language coverage of the web-domains classified, taking into account the English centric nature of the web. Building upon [47], we gathered statistics from Amazon Alexa (www.alexa.com/topsites/countries) about web traffic (the top 50 most visited websites) for all countries across the globe, matching these lists with the list of domains used in our analysis. For 127 countries we found at least one domain in the reliable top-50 news source, and for 21 (iso2 codes: AE, AR, BB, BE, CA, DK, FR, KE, MX, NG, PA, PE, PH, PR, PT, QA, SD, SE, TT, US and VE) they have at least one domain in the top-50 websites labelled as unreliable. In fact, this is a lower bound, because Alexa provided only major domains, disregarding subdomains that we instead classified as well. This large presence among the very top tier of websites suggests that the results are robust for multilanguage/multicultural analysis.

We considered the URLs contained in messages, and labeled each URL according to the political leaning (left, left-center, neutral, right- center and right) of its media source and

the type of source (political, satire, mainstream media, science, conspiracy/junk science, clickbait, fake/hoax) as manually classified by external experts. In particular, building on [47], we have classified our news sources as *reliable* (when belonging to the Science, Mainstream Media categories) and *unreliable* (when belonging to the Satire, Clickbait, Political, Fake or Hoax, Conspiracy and Junk science and Shadow categories). Finally, as a third category we consider tweets not containing any url to be classified (i.e. mere opinions without a source). We excluded all the web-domains classified as 'Other' as they point to general content that cannot be easily classified, such as videos on YouTube or post on Instagram.

### 3.4 Bot detection

Social bots are automated accounts that mimic human behavior, create posts, comments and likes on social networks. Analyzing the role of social bots in the emotional response to controversial topics such as those studied in this paper is important, given that they are being systematically used for the manipulation of the public opinion through social media [48]. In particular, bots have proven successful in spreading low-credibility content by strategically targeting influential human users [49]. Their typical mode of operation also includes amplification of inflammatory contents by human users [50], whereas actual instigation of emotions is rarer. However, cases of successful bot-to-human transmission of negative emotions like anger have been documented in online COVID-19 related conversations [51]. Bots can therefore be considered a significant threat to public health [52], whose mode of operation also includes emotional manipulation. It is therefore important to investigate the potential role of bots in the social dynamics of emotional responses on controversial topics.

To distinguish a bot from a human, we chose some criteria associated to some forms of unusual social behavior. In particular, based on [44], automated accounts tend to show an important productivity on social media by posting excessive number of content and, especially, in their concentration in particular moments during the day (e.g. overnight or all day long). We identified automated accounts using a machine learning algorithm designed to classify Twitter accounts as humans or bots and used in previous research [50, 53] The classification of users into "human" or "bot" is based on ten features that yield the best classification accuracy according to several authors [50, 54]. The features are (1) statuses count; (2) followers count; (3) friends count; (4) favorites count; (5) listed count; (6) default profile; (7) geo enabled; (8) profile use background image; (9) protected; and (10) verified, following the same prescriptions of previous studies [50, 53–55]. The models undergo training using 80% of the data and are subsequently validated on the remaining 20%. The division between these two sets is performed while ensuring a balance between bots and humans at the level of each individual original dataset. The models gave us the highest accuracy (>90%) and precision in identifying bots (>95%) [54].

In a previous work [53], we tested the ability of the algorithm to generalize the classification out of the data sample used for training and validation by applying the model on an independent data set [56], consisting of labeled information about 8,092 humans and 42,446 bots. The results of the classifier are satisfactory, with an accuracy of 60 %, an F1-score higher than 70 %, and a recall of 58 %. In this specific work, we did not manually inspect the users but we assume that our results woould be consistent with the ones found in the above mentioned work.

### 3.5 Emotional detection

The NLP pipeline used to process the tweets allows processing of text and emotions in multiple languages. For the emotions we rely on the NRC Word-Emotion Association Lexicon [57], containing 14,182 English words associated with one or more of eight basic emotions (*anger*, *fear*, *anticipation*, *trust*, *surprise*, *sadness*, *joy*, and *disgust*) and two sentiments (*negative* and *positive*). These words are then translated into several languages, including the 18 used in this work. A second resource used in the pipeline is the NRC-VAD Lexicon [58], containing 19,971 English words associated with three scores representing respectively valence, arousal, and dominance. As in the previous resource the NRC-VAD is also including the translation of each word in all the 18 needed languages. Taking as input the text of the tweet and its language, the pipeline returns the list of words associated with emotions, the amount of each emotion contained in the message and the total values of valence, arousal, and dominance. Being based on a lexicon, the emotions extraction is preceded by two preprocessing steps in order to increase the amount of emotions retieved from the text. The first preprocessing involves elements that are relevant in social media as hashtags and emoji, often carrying an important part of the content of a tweet. To process the hashtags we expand the Ekphrasis library [59], originally developed for English, to cover additional languages. This allows us to identify hashtags in tweets and to split them in the words composing it, e.g. #staysafe into 'stay safe'. To detect emotions related to emojis we adopted the strategy of replacing them with their textual descriptions in the language of the tweet. The description can then be used to search for matches in the NRC lexicons, e.g. "worried face". Since the NRC lexicons doesn't contain all the inflected forms of annotated words, we preprocess the tweets lemmatizing them (including the text extracted from hashtags and emoji), to increase the number of matches with the words in the lexicon. The lemmatization is done using the Spacy library. The pipeline also has the possibility of using a list of words that need to be excluded from the emotions extraction, for instance when working on tweets about Covid, we can decide to exclude words as 'virus' being present in data, as topic, regardless of the emotions expressed by the messages.

### 3.6 Emotional aggregation

Various emotional algorithms have been already tested on different kind of data. One of them is surely the Valence-Arousal-Dominance model [60]. We found that the three sentiments are highly correlated among them in our dataset, as shown in Additional file 1. Based on the purpose of this research, we decide to adopt another algorithm with eight emotions, four of which with positive valence (Trust, Anticipation, Joy, Surprise) and four with negative valence (Anger, Sadness, Fear, Disgust) in order to better capture each single emotion. Before starting the analysis, a preprocessing phase has been fundamental to understand how to normalize the data, since the emotional range of each emotion does not follow the same scale. In particular, we found the total emotional value of the messages posted on Twitter by summing each singular emotion. Then, we divided each emotion by the total emotional value found in order to obtain that the sum of each emotion should be equal to 1. We grouped the positive emotions (Anticipation, Surprise, Joy and Trust) and the negative emotions (Sadness, Anger, Fear, Disgust) into two different categories (positive vs negative emotions). After this procedure, we normalized the valence of each tweet across the emotional range of −1 (completely negative valence) and +1 (completely positive valence). In the emotional analysis of the three reliability-ranked categories of

news Fig. 1 and Fig. 2, we also decided to delete the emotional values equal to 0 because our interest was to observe the distribution of emotional response and the 0 value might indicate the absence of any emotion in the message posted.

## 4 Results

### 4.1 Emotions distribution among the Covid-19 vaccination and the baseline sample

We consider online discussions in which no specific contentious topic dominates consisting of more than 10 millions messages posted to a popular microblogging platform, Twitter, between 21/04/2021 and 24/04/2021 from 6 millions of unique users. We also consider the online discussions concerning the massive COVID-19 vaccination campaign between August 31, 2020 and July 15, 2021, that is from the announcement of the availability of the first COVID-19 vaccines up to the peak of the vaccine campaign in Europe and in the United States, consisting of 9.6 million of posts from 3 million of unique users. We find that the conversation related to a contentious topic such as vaccination generate a spectrum of emotional response that differs from that of the baseline, as illustrated in Fig. 1.

In particular, the vaccine sample shows an over representation of four emotions with respect to the baseline: Fear, Anticipation, Sadness and Trust. Indeed, these three emotions have a mean value greater than the threshold value. If we imagine an ideal benchmark in which each of the eight emotions is equally represented, each of them should be by a frequency of 0.125. We therefore draw this level as a dotted line in the figure to make it more readable which emotions are actually over- vs. under-represented with respect to the benchmark. We can notice that the baseline distribution is more evenly distributed than the vaccination one. In particular, for the baseline we see that Joy, Sadness and Anger are very close to the benchmark value. On the contrary, the distribution is considerably less uniform for the vaccination sample, having an over-representation of Fear, Anticipation, Sadness and Trust. This suggests that a contentious topic may lead to the selective amplification of certain emotions with respect to others. A Kolmogorov-Smirnov test has

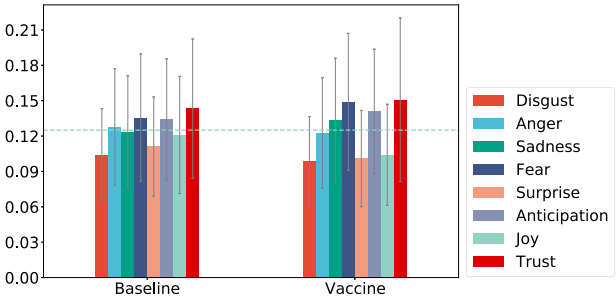


**Figure 1** Emotion Mean Value. For each sample (baseline (left panel) and vaccine (right panel)), we quantified the mean value of each emotion: Anger, Sadness, Surprise, Fear, Anticipation, Trust, Disgust, Joy. The horizontal line is placed at the 0.125 value to better visualize which emotions are over- vs under-represented with respect to our benchmark value. The grey vertical lines show the standard deviation of each distribution. To better understand the statistically significance of each emotions, we performed a Kolmogrov-Smirnov Test with the result that all the distribution are statistically different at the 95% of confidence level with a p-value lower than 0.01 both across emotions and among the baseline and vaccine samples. In addition, we calculated the standard error of the mean, resulting to be negligible among all the emotion distribution

been performed to evaluate the statistical significance of the two samples. We find that the distributions are actually statistically different between the two samples.

### 4.2 Emotional responses to different levels of news reliability

At this point, however, it may be reasonable to conjecture that this selective amplification of certain emotions in the case of contentious events may be in turn further modulated by specific features. First of all, we can ask whether getting in contact with unreliable news related to the contentious topic is different than getting in contact with reliable ones as far as the emotional response is concerned.

To test this, Fig. 2 shows the distribution of emotions for the three reliability-ranked categories of news: reliable, unreliable and opinions. The eight emotions have been normalized in a range between −1 (the most negative emotional valence) and +1 (the most positive emotional valence) to improve readability.

Interestingly, we find that in the baseline sample there are no systematic differences in terms of emotional response to the three categories of news. In particular, the median is 0 for all the news categories, showing that users not only do not discriminate across news categories in terms of emotional response, but also present a balanced overall response to the news, since there is not a skew towards positive or negative responses, indicating a more balanced emotional response, specifically with respect to the vaccination sample.

Also in the case of the differences between reliable and unreliable news we performed a non-parametric Kolmogrov-Smirnov Test, finding that the two distributions are not statistically significantly different, with a p-value equal to 0.22.

In the case of the vaccine sample, the pattern is completely different, and the distributions of emotions are statistically different for each type of news considered.

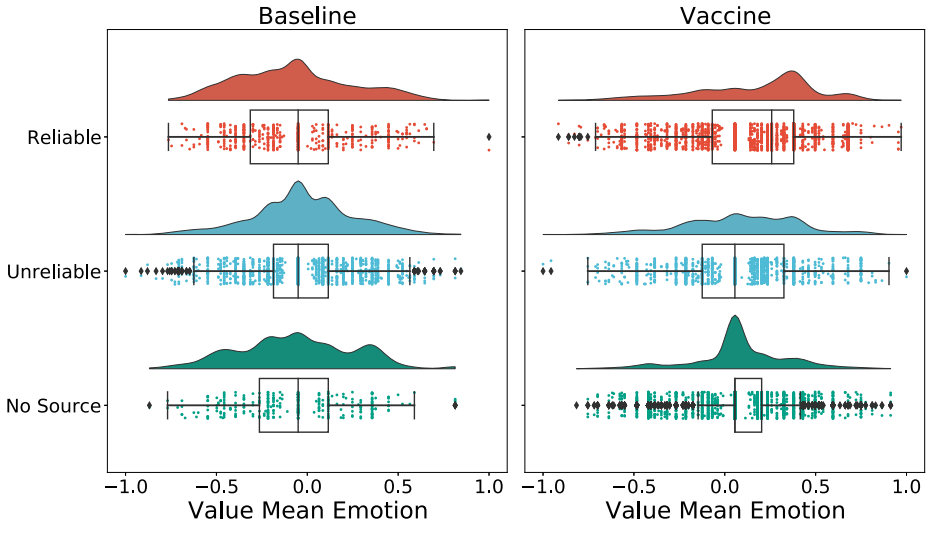


**Figure 2** Emotional range for different categories of news. For each sample (baseline (left panel) vs vaccine (right panel)), we quantified the distribution of emotional values from −1 to 1, respectively for tweets containing reliable news (top), unreliable news (middle) or just comments without sources (bottom). All the distribution have been tested through a Kolmogrov-Smirnov Test in order to evaluate the difference among the samples and across categories of news considered. The distribution are statistically significant at the 95% of confidence level with a p-value lower than 0.01 among the two samples and within the same sample, with the only exception of tweets containing unreliable news and reliable news in the Baseline sample, having a p-value equal to 0.22

Reliable news are characterized by being skewed toward the positive valence side of the spectrum when compared to unreliable ones. In other words, reliable news elicit in users more positive emotions in the case of a contentious topic with respect to unreliable ones.

### 4.3 Emotional responses of verified vs unverified users

In the Fig. 3 panel (A), we compare the emotional responses of verified vs. un-verified users. We find that verified users are characterized by a more positive emotional response. Once again, the difference between the two distributions is statistically significant as shown by the Kolmogrov-Smirnov Test. Moreover, also for the baseline sample we find more positive valenced responses for verified users (an average level of 0.31) in contrast with the unverified users (0.27), as shown in Table 2. Interestingly, in the case of the baseline the positive emotional response of the verified users is stronger than in the case of the vaccine sample. Even if verified users, as a consequence of their incentives to reputation management and accountability, tend to favor positively valenced emotional reactions, it turns out that the contentious nature of the vaccine topic influences their own mode of response and determines a less positively valenced response, although maintaining an overall positive emotional tone. Interestingly, for un-verified users the emotional response

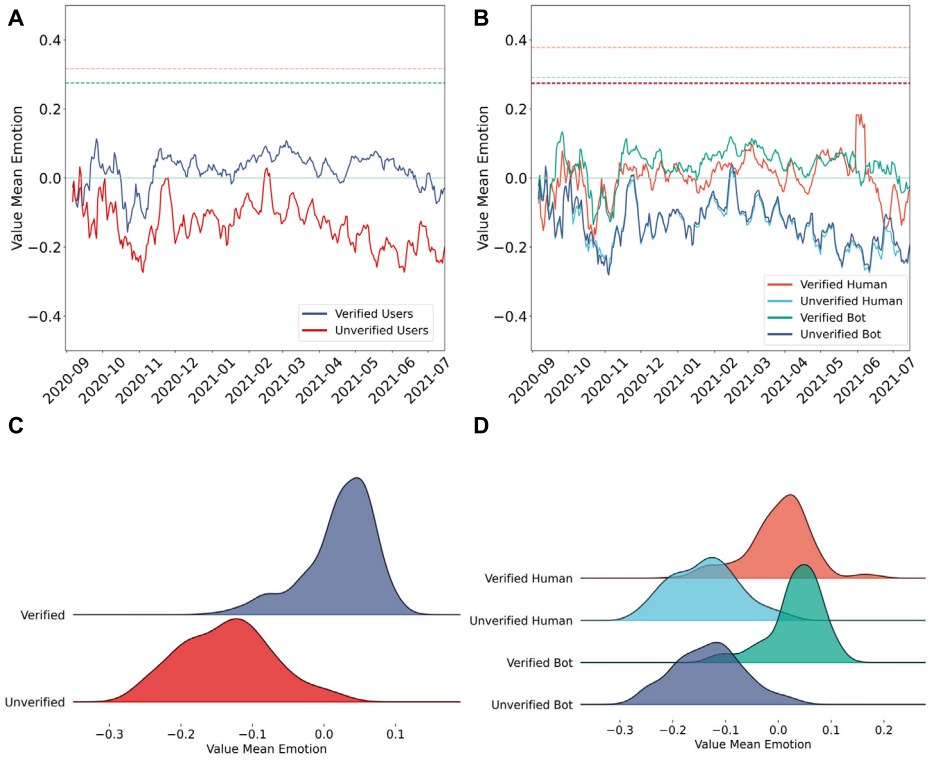


**Figure 3** Emotional Analysis for different types of user accounts. We quantified, ranging from −1 to +1, across (A) verified and unverified users, (B) both human and bot, over time. The dashed horizontal lines indicate the emotional mean values for each user type, based on a baseline of 10 million randomly sampled tweets. This highlights that the mean emotion of the random distribution is higher that of our specific interest. We present the distribution of emotional values, aggregated by day, for each category concerning the vaccine sample (C-D). We conducted a Kolmogorov-Smirnov Test across the entire distribution for each category. The results indicate statistical significance at the 95% confidence level, with a p-value below 0.01, not only within categories but also across the various samples considered

**Table 2** Summary Statistics of each type of users considered both in the Vaccine and Baseline sample. We show the mean value and the standard deviation of each distribution for the following categories: verified and unverified users, bots and humans. We find that on average the baseline sample is characterized by having higher values for each type of users and that the discourses carried out by verified accounts are marked by a more positive emotional features

| Type | Vaccine | Baseline |
|---|---|---|
| Users | Mean(SD) | Mean(SD) |
| Verified Users | 0.041(0.81) | 0.316(0.668) |
| Unverified Users | −0.131(0.8) | 0.275(0.744) |
| Verified Bot | 0.052(0.81) | 0.290(0.682) |
| Unverified Bot | −0.126(0.799) | 0.273(0.745) |
| Verified Human | 0.022(0.811) | 0.378(0.626) |
| Unverified Human | −0.134(0.8) | 0.276(0.743) |

is not only less positive, but particularly so in the first months of the vaccination campaign, when the level of contentiousness was particularly high.

In the second panel (B), we then find the emotional responses of the verified and unverified users as distinguished between bots and humans. Discerning bots from humans is usually carried out through the observation of the respective behaviors on social media platforms. More specifically, automated accounts are usually characterized by some forms of unusual behavior: very large volume of content created, or very high frequency of creation, for instance (see Method section for further details). Interestingly, the most significant distinction in terms of emotional expression in our sample is not that between bots and humans, but rather that between verified and un-verified users within each category. It turns out that when we compare the entire distribution of emotional responses between any pair of these four categories of users, each couple of distributions is significantly different in all cases. In particular, the C panel shows the emotional distribution across time respectively for verified and unverified users, while the D panel shows the same analysis discerning between bots and human users to better visualize the statistical distribution of each category.

When an user is verified, be it a human or bot, their emotional response is more positively valenced over time. Likewise, responses are less positively valenced for un-verified users, be them humans or bots. Comparing our results with the baseline sample, we notice an important difference. Whereas verified humans tend to present a more positive response in the baseline, for verified bots the response is more positive in the vaccine sample. The differences between the four categories of users tend to stabilize as the vaccination campaign unfolds, whereas in the very first months the responses are less predictable (see further details in Additional file 1 Appendix).

### 4.4 Emotional responses based on the political orientation

Another important aspect to consider is the role of news media sources according to their political orientation. In Fig. 4, we analyzed the political orientation of the sources in relation to the verified/un-verified status of the users and their emotional response. As always we have normalized the valence of the tweets across the emotional range (−1, +1) for each of the categories mentioned. In particular, each URL appearing in the messages has been manually inspected by experts (see Method section), classifying the various sources in one of the following categories: left, left-center, center, right-center, right.

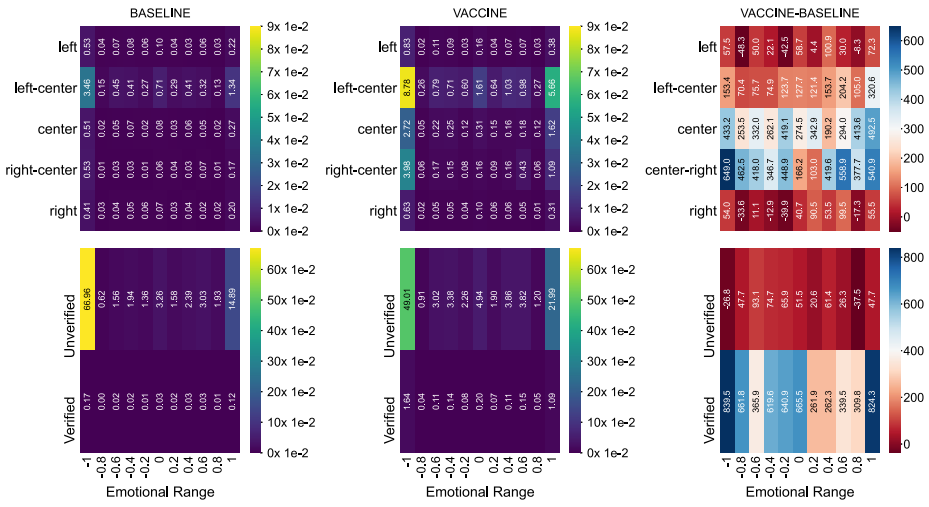


**Figure 4** Emotional range for messages posted on Twitter by verified-unverified accounts and with URLs associated to five different political leaning. For each panel, we quantified statistics of tweets across the emotional range (−1, +1) according to the political leaning of the sources and the verified/un-verified status of the users. The left panel reports the 10M random sample of tweets as a baseline, while the central one is about the vaccine tweets. The right panel shows the percentage difference between vaccine and baseline data for each cell. We show how emotional responses can became a sort of marker for different ideological position, especially for the left-center, center and center-right categories. An interesting result is shown by the verified accounts through which the emotional responses vary at different scales, reaching high negative and positive values

We quantified the distribution of the political orientation of the sources and the status of the users both for the baseline and the vaccination samples. We obtain an interesting result: the variation between the two samples is apparent, and the difference is statistically significant among all the political orientations, whether the users are verified or not. The first panel of the Figure indicates the statistics for tweets from the baseline sample, showing that a higher number of tweets has a left-center orientation compared to the others. However, in the vaccine sample the incidence of political orientations changes, and also the emotional responses are pretty different, especially for the left-center, center and center-right categories. The difference is substantial in the third panel where we can observe significant variations in emotional responses especially for the left-center, center and center-right categories. This clearly shows how emotional responses become markers of a differentiation between different ideological positions in the case of a highly contentious topic such as vaccination. Interestingly, the positions where we observe the most heterogeneity in response are the relatively moderate ones, and not the most extreme, as the former are the ones that need to differentiate more from their relatively closer analogs. Extreme positions are already well differentiated and their emotional response patterns are less characteristic.

In the bottom panel, where we further differentiate between verified and un-verified users, we find that again the emotional responses are different between the two categories. For unverified users, there is scarce differentiation in terms of dominant emotions. However, in the case of verified users, the emotional response changes significantly, with the emotional valence that tends to radicalize toward highly positive or highly negative levels, respectively (see Fig. 4), similarly to what we observed in 3.

## 5 Discussion

Are online 'emotions' an important source of social cognition, which is modulated by the status and by other specific characteristics of the subjects? Our analysis shows that this is the case, especially when the topic of the discussion is contentious. As we have seen, a contentious topic such as vaccination elicits stronger and more nuanced emotional responses than baseline conversations. Moreover, users who are characterized by a specific status, that of verified users, have an emotional response whose valence is more positive than that of unverified users. The higher social status related to the verification, and the consequent incentive to reputation management and higher accountability, implies that such users feel more pressured toward providing constructive, inspiring emotional responses rather than dismissive and confrontational ones. Interestingly, this is the case whether the users are humans or bots. However, given the recent changes in the conditions of access to a verified status on Twitter, it is possible that such social pressure effects on emotional responses would not be found under the new regime, as verified accounts now do not signal social status any longer but only mark the purchase of a specific service.

On the other hand, also political orientations of the sources cited have an important implication for emotional response, and in the case of contentious topics there is a general tendency to differentiate the emotional response so that to better differentiate political identity accordingly. However, unlike what could be expected, it is especially in the case of the differentiation between relatively moderate positions that emotional responses are used as a differentiating factor, whereas in the case of more extreme positions this effect is less marked, as the positioning at an extreme side of the political spectrum already suffices to ensure differentiation. This confirms that emotional responses may be strategically inflated in online interactions to construct a specific, and optimally differentiated, social identity [61]. This may be especially the case for contentious topics, where emotional arousal is pursued by the parties involved also to enhance the salience of intentional signals of false information that function as demonstrations of commitment to the group's cause, aiding in the strengthening of group solidarity against opponents in polarizing discussions [62].

This implies, in particular, that users with different political orientations might develop their own specific 'emotion playbook' which is characteristic of their political position and that might allow very complex forms of social coordination and even synchronization with their base through a suitable, skillful use of emotional signals. A political base that is particularly sensitive to fear or disgust could therefore be activated by any kind of content where such emotions are balanced with others in a certain, characteristic proportion. It is possible that this kind of strategy has been already experimented by populist leaders worldwide, where the emphasis on the emotional component often overrides that of the specific content [63]. This is certainly a topic that deserves further analysis and that could prove of great importance in understanding future social responses to major contentious topics and events.

## Supplementary information

**Supplementary information** accompanies this paper at https://doi.org/10.1140/epjds/s13688-024-00452-7.

**Additional file 1.** (PDF 6.5 MB)

**Acknowledgements**
We want to thank the Twitter COVID-19 working group for providing us the data stream. We also would like to acknowledge Sara Tonelli and Elisa Leornardelli for useful discussions.

**Funding**
Not applicable.

**Data availability**
Data analyzed in this work are available from the corresponding author upon reasonable request.

## Declarations

**Competing interests**
The authors declare that they have no competing interests.

**Author contributions**
AB, RG performed numerical experiments and analyzed the data. SM developed the algorithm. PS and MDD designed the study. AB, PS and MDD wrote the manuscript. All authors read and approved the final manuscript.

**Author details**
[1]Department of Information Engineering and Computer Science, University of Trento, Via Sommarive, 9, 38123 Povo (TN), Italy. [2]CHuB Lab, Fondazione Bruno Kessler, Via Sommarive, 18, 38123 Povo (TN), Italy. [3]Digital Humanities, Fondazione Bruno Kessler, Via Sommarive, 18, 38123 Povo (TN), Italy. [4]DiSFiPEQ, University of Chieti-Pescara, Viale Pindaro 42, 65127 Pescara, Italy. [5]metaLAB (at) Harvard, 42 Kirkland St, 02138 Cambridge, MA, USA. [6]Department of Physics and Astronomy, G. Galilei, University of Padua, Via Francesco Marzolo 8, 35131 Padua, Italy.

Received: 13 July 2023 Accepted: 9 February 2024 Published online: 08 March 2024

**References**

1. Behrens F, Kret ME (2019) The interplay between face-to-face contact and feedback on cooperation during real-life interactions. J Nonverbal Behav 43(4):513–528
2. Drolet AL, Morris MW (2000) Rapport in conflict resolution: accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. J Exp Soc Psychol 36(1):26–50
3. Kurzban R (2001) The social psychophysics of cooperation: nonverbal communication in a public goods game. J Nonverbal Behav 25:241–259
4. Burgoon JK, Wang X, Chen X, Pentland SJ, Dunbar NE (2021) Nonverbal behaviors "speak" relational messages of dominance, trust, and composure. Front Psychol 12:624177
5. Jiang J, Dai B, Peng D, Zhu C, Liu L, Lu C (2012) Neural synchronization during face-to-face communication. J Neurosci 32(45):16064–16069
6. Feese S, Arnrich B, Tröster G, Meyer B, Jonas K (2012) Quantifying behavioral mimicry by automatic detection of nonverbal cues from body motion. In: 2012 international conference on privacy, security, risk and trust and 2012 international confernece on social computing. IEEE, Los Alamitos, pp 520–525
7. Baker J (2019) The empathic foundations of security dilemma de-escalation. Polit Psychol 40(6):1251–1266
8. Dunbar RIM (1996) Grooming, gossip, and the evolution of language
9. Sariñana-González P, Romero-Martínez Á, Moya-Albiol L (2018) Cooperation between strangers in face-to-face dyads produces more cardiovascular activation than competition or working alone. J Psychophysiol
10. Halpin J, Wilson C (2022) How online interaction radicalises while group involvement restrains: a case study of action zealandia from 2019 to 2021. Polit Sci 74(1):18–33
11. Antheunis ML, Schouten AP, Valkenburg PM, Peter J (2012) Interactive uncertainty reduction strategies and verbal affection in computer-mediated communication. Commun Res 39(6):757–780
12. Lidsky LB (2011) Incendiary speech and social media. Tex Tech Law Rev 44:147
13. Cowen A, Sauter D, Tracy JL, Keltner D (2019) Mapping the passions: toward a high-dimensional taxonomy of emotional experience and expression. Psychol Sci Public Interest 20(1):69–90
14. Lindquist KA, Siegel EH, Quigley KS, Barrett LF (2013) The hundred-year emotion war: are emotions natural kinds or psychological constructions? comment on lench, flores, and bench (2011)
15. Gallese V, Keysers C, Rizzolatti G (2004) A unifying view of the basis of social cognition. Trends Cogn Sci 8(9):396–403
16. Garcia D, Kappas A, Küster D, Schweitzer F (2016) The dynamics of emotions in online interaction. R Soc Open Sci 3(8):160059
17. Beneito-Montagut R (2015) Encounters on the social web: everyday life and emotions online. Sociol Perspect 58(4):537–553
18. Beneito-Montagut R (2017) Emotions, everyday life, and the social web: age, gender, and social web engagement effects on online emotional expression. Sociol Res Online 22(4):87–104
19. Chmiel A, Sienkiewicz J, Thelwall M, Paltoglou G, Buckley K, Kappas A, Hołyst JA (2011) Collective emotions online and their influence on community life. PLoS ONE 6(7):22207
20. Fox J, Moreland JJ (2015) The dark side of social networking sites: an exploration of the relational and psychological stressors associated with Facebook use and affordances. Comput Hum Behav 45:168–176
21. Jahng J, Kralik JD, Hwang D-U, Jeong J (2017) Neural dynamics of two players when using nonverbal cues to gauge intentions to cooperate during the prisoner's dilemma game. NeuroImage 157:263–274
22. Phirangee K, Hewitt J (2016) Loving this dialogue!!!!: expressing emotion through the strategic manipulation of limited non-verbal cues in online learning environments. In: Emotions, technology, and learning, pp 69–85

23. Walther JB, Loh T, Granka L (2005) Let me count the ways: the interchange of verbal and nonverbal cues in computer-mediated and face-to-face affinity. J Lang Soc Psychol 24(1):36–65
24. Shum H-Y, He X-d, Li D (2018) From Eliza to xiaoice: challenges and opportunities with social chatbots. Front Inf Technol Electron Eng 19(1):10–26
25. Riordan MA, Kreuz RJ (2010) Emotion encoding and interpretation in computer-mediated communication: reasons for use. Comput Hum Behav 26(6):1667–1673
26. Sobkowicz P, Sobkowicz A (2012) Two-year study of emotion and communication patterns in a highly polarized political discussion forum. Soc Sci Comput Rev 30(4):448–469
27. Zhu Q, Weeks BE, Kwak N (2021) Implications of online incidental and selective exposure for political emotions: affective polarization during elections. New Media Soc 26(1):450–472
28. Pröllochs N, Bär D, Feuerriegel S (2021) Emotions explain differences in the diffusion of true vs. false social media rumors. Sci Rep 11(1):22721
29. Mønsted B, Lehmann S (2022) Characterizing polarization in online vaccine discourse—a large-scale study. PLoS ONE 17(2):0263746
30. Chen K, He Z, Chang R-C, May J, Lerman K (2023) Anger breeds controversy: analyzing controversy and emotions on reddit. In: International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. Springer, Berlin, pp 44–53
31. Jakob J, Dobbrick T, Freudenthaler R, Haffner P, Wessler H (2023) Is constructive engagement online a lost cause? Toxic outrage in online user comments across democratic political systems and discussion arenas. Commun Res 50(4):508–531
32. Tomljenovic H, Bubic A, Erceg N (2020) It just doesn't feel right–the relevance of emotions and intuition for parental vaccine conspiracy beliefs and vaccination uptake. Psychol Health 35(5):538–554
33. Guo S, He Z, Rao A, Jang E, Nan Y, Morstatter F, Brantingham J, Lerman K (2023) Measuring online emotional reactions to offline events. arXiv preprint. arXiv:2307.10245
34. Guo S, Burghardt K, Rao A, Lerman K (2022) Emotion regulation and dynamics of moral concerns during the early covid-19 pandemic. arXiv preprint. arXiv:2203.03608
35. Vemprala N, Bhatt P, Valecha R, Rao H (2021) Emotions during the Covid-19 crisis: a health versus economy analysis of public responses. Am Behav Sci 65(14):1972–1989
36. Wang J, Fan Y, Palacios J, Chai Y, Guetta-Jeanrenaud N, Obradovich N, Zhou C, Zheng S (2022) Global evidence of expressed sentiment alterations during the Covid-19 pandemic. Nat Hum Behav 6(3):349–358
37. Zhang X, Yang Q, Albaradei S, Lyu X, Alamro H, Salhi A, Ma C, Alshehri M, Jaber II, Tifratene F et al (2021) Rise and fall of the global conversation and shifting sentiments during the Covid-19 pandemic. Humanit Soc Sci Commun 8(1):1–10
38. Ekman P (1999) Basic emotions. In: Handbook of cognition and emotion 98(45-60), p 16
39. Tracy JL, Randles D (2011) Four models of basic emotions: a review of Ekman and cordaro, izard, levenson, and panksepp and watt. Emot Rev 3(4):397–405
40. Lahno B (2020) Trust and emotion. In: The Routledge handbook of trust and philosophy, pp 147–159
41. Papacharissi Z (2004) Democracy online: civility, politeness, and the democratic potential of online political discussion groups. New Media Soc 6(2):259–283
42. Asker D, Dinas E (2019) Thinking fast and furious: emotional intensity and opinion polarization in online media. Public Opin Q 83(3):487–509
43. Ng E (2020) The pandemic of hate is giving Covid-19 a helping hand. Am J Trop Med Hyg 102(6):1158
44. Sacco PL, Gallotti R, Pilati F, Castaldo N, De Domenico M (2021) Emergence of knowledge communities and information centralization during the Covid-19 pandemic. Soc Sci Med 285:114215
45. O'kane C (2023) Twitter is officially ending its old verification process on April 1. To get a blue check mark, you'll have to pay. https://www.cbsnews.com/news/twitter-blue-check-verification-ending-new-subscription-april-1-elon-musk/, CBS News. [Accessed 11-November-2023]
46. FactCheck M (2020) MediaBiasFactCheck. https://mediabiasfactcheck.com/
47. Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M (2020) Assessing the risks of 'infodemics' in response to Covid-19 epidemics. Nat Hum Behav 4(12):1285–1293
48. Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. Commun ACM 59(7):96–104
49. Shao C, Ciampaglia GL, Varol O, Yang K-C, Flammini A, Menczer F (2018) The spread of low-credibility content by social bots. Nat Commun 9(1):1–9
50. Stella M, Ferrara E, De Domenico M (2018) Bots increase exposure to negative and inflammatory content in online social systems. Proc Natl Acad Sci 115(49):12435–12440
51. Shi W, Liu D, Yang J, Zhang J, Wen S, Su J (2020) Social bots' sentiment engagement in health emergencies: a topic-based analysis of the Covid-19 pandemic discussions on Twitter. Int J Environ Res Public Health 17(22):8701
52. Allem J-P, Ferrara E (2018) Could social bots pose a threat to public health? Am J Publ Health 108(8):1005
53. González-Bailón S, De Domenico M (2021) Bots are less central than verified accounts during contentious political events. Proc Natl Acad Sci 118(11):2013443118
54. Ferrara E (2017) Disinformation and social bot operations in the run up to the 2017 french presidential election. arXiv preprint. arXiv:1707.00086
55. Stella M, Cristoforetti M, De Domenico M (2019) Influence of augmented humans in online interactions during voting events. PLoS ONE 14(5):0214210
56. Yang KC, Varol Onur, Varol Onur Hui PM, Menczer F (2020) Scalable and generalizable social bot detection through data selection. In: Proceedings of the AAAI conference on artificial intelligence pp 1096–1103
57. Mohammad SM, Turney PD (2013) Crowdsourcing a word-emotion association lexicon. Comput Intell 29(3):436–465
58. Mohammad SM (2018) Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In: Proceedings of the annual conference of the association for computational linguistics (ACL), Melbourne, Australia
59. Baziotis C, Pelekis N, Doulkeridis C (2017) Datastories at semeval-2017 task 4: deep lstm with attention for message-level and topic-based sentiment analysis. In: Proceedings of the 11th international workshop on Semantic Evaluation (SemEval-2017). Assoc. Comput. Linguistics, Vancouver, pp 747–754

60. Warriner AB, Kuperman V, Brysbaert M (2013) Norms of valence, arousal, and dominance for 13,915 English lemmas. Behav Res Methods 45(4):1191–1207
61. Graham J, Haidt J, Nosek BA (2009) Liberals and conservatives rely on different sets of moral foundations. J Pers Soc Psychol 96(5):1029
62. Petersen MB, Osmundsen M, Tooby J (2021) The evolutionary psychology of conflict and the functions of falsehood. In: The politics of truth in polarized America, p 131
63. Rico G, Guinjoan M, Anduiza E (2017) The emotional underpinnings of populism: how anger and fear affect populist attitudes. Swiss Polit Sci Rev 23(4):444–461

## Publisher's Note