Title
A lurking bias: Representativeness of users across social media and its implications for sampling bias in cognitive science

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Vuorio, Valtteri
Horne, Zachary

Publication Date
2023

Peer reviewed

A lurking bias: Representativeness of users across social media and its implications for sampling bias in cognitive science

Valtteri Vuorio
University of Edinburgh
Department of Psychology
Edinburgh EH8 9JZ

Zachary Horne
University of Edinburgh
Department of Psychology
Edinburgh EH8 9JZ

Abstract

Within internet there exists the 90-9-1 principle (also called the 1% rule), which dictates that a vast majority of user-generated content in any specific community comes from the top 1% active users, with most people only listening in. When combined with other demographic biases among social media users, this casts doubt as to how well these users represent the wider world, which might be problematic considering how user-generated content is used in psychological research and in the wider media. We conduct three computational studies using pre-existing datasets from Reddit and Twitter; we examine the accuracy of the 1% rule and what effect this might have on how user-generated content is perceived by performing and comparing sentiment analyses between user groups. Our findings support the accuracy of the 1% rule, and we report a bias in sentiments between low- and high-frequency users. Limitations of our analyses will be discussed.

Keywords: 90-9-1 principle; participation bias; cognitive science; data science; social media

Introduction

Psychological science suffers from unprincipled sampling procedures, and, correspondingly, unrepresentative populations. Undergraduate participant pools are less diverse and more educated compared to the general population (Henrich et al., 2010; Nielsen et al., 2017; Rad et al., 2018), and even though participants from crowd-sourcing platforms like Mechanical Turk are more diverse than students, they do not represent broader population in either political orientation, education, or age (Sheehan, 2017). Further issues arise from the US-centricity of psychological research, which often leads to researchers comparing their subject pool against U.S. population and attempting to generalise their findings on the basis of this comparison (Cheon et al., 2020).

These biases have led for calls to use larger, more naturalistic datasets—and with new developments in computational methods, social scientists now have the tools and resources to examine human behaviour, particularly by scraping data from social media sites. Yet, even in large-scale naturalistic datasets, there exists vast demographic differences: For example, Twitter users tend to be younger as well as more urban and educated, and Reddit users are predominantly 18- to 34-year-olds and twice as likely to be men than women; likewise, while 46% of female U.S. adults report using Pinterest, only 16% of men report doing the same (Pew Research, 2021). This means larger sample sizes do not necessarily lead to a decrease in error, but can even increase it when the representativeness of the data deviates substantially from the broader population (Kaplan et al., 2014). These findings suggest large data sources may not be as representative as we might have hoped; perhaps even less so in some ways than now lamented convenience sampling procedures (e.g., Hauser et al., 2022).

Much of cognitive science remains focused on conducting controlled laboratory studies, so concerns about the demographic skew on social media sites may seem like a problem for computational social science broadly, but not cognitive science. However, cognitive scientists (e.g., Low et al., 2020; Priniski & Horne, 2018; 2019; Dayter & Messerli, 2022) and adjacent fields like natural language processing (Tadesse et al., 2019; Tan et al., 2016; Turcan & McKeown, 2019) rely extensively on these sources and some of the most influential findings in political and social psychology are based at least in part on data from a single domain like Twitter or Reddit (e.g., Jones & Silver, 2019; Mosleh et al., 2021). Rarely do studies "pair" data with laboratory experiments to address the generalisability of their work (e.g., Priniski & Horne, 2018).

To make matters worse, skewed demographics are not the entirety of the problem. How content is *created* on these sites may also pose problems for the validity of psychological claims made based on social media data alone. Natural language corpora harvested in the early days of social media appeared to follow a 90-9-1 principle, sometimes known as the 1% rule (Carron-Arthur et al., 2014; Nielsen, 2006; van Mierlo, 2014). This principle states that in any internet community, participants are divided in such a way that 90% of users (Lurkers) consume content while rarely creating it, 9% (Contributors) modify existing content (in the form of upvotes, likes, or retweets) and sometimes create it, and 1% (Superusers) are the most active participants, responsible for vast majority of contributions in their selected platform. Originally observed in the so-called "blogosphere"—where the divide was estimated to be an even larger 95–5–0.1 split—the distribution of content creators to users extends well-beyond this domain (Nielsen, 2006). For example, Wikipedia (2023) states it has had 130,218 active contributors in January even though the number of total page views in English Wikipedia amounts to 10 billion from 814 million unique devices. Still, little research has been conducted to verify whether this principle holds on contemporary social media ecosystems or investigate how it could impact our understanding of distinctly psychological questions (van Mierlo, 2014). In three studies, we set out to examine the distribution of content creation on social media sites used and examine the impact possible distributional issues could have on commonly deployed techniques used in social science.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

Datasets

We first conducted secondary data analysis on an existing natural language corpora gathered from Reddit (Baumgartner, 2015; Gaffney & Matias, 2018). We use data from January 2015 (Studies 1 and 2) and October 2014 (Studies 1 and 3). After removing the most clearly labeled bot accounts (used for statistical reporting as well as entertainment purposes) and observations where either the author or the body of comment were missing, we were left with $49,186,418$ comments from $2,500,848$ users from the January 2015 dataset and $42,315,878$ comments from $2,209,348$ users from the October 2014 dataset. We conducted another secondary data analysis on a set of Twitter data (Baoi, 2014) which contains $316,669$ tweets from October 21-23, 2014, in which the tweets mined included hashtags #Gamergate or #NotYourShield.

Study 1: Testing the 1% rule's accuracy on Reddit and Twitter

Methods

We first set out to confirm the validity of the 1% rule; that is, we examined what proportion of users participate at different topical and community levels on Reddit, a commonly used data source for cognitive scientists and social psychologists. We ranked users based on their post frequency, log-transforming both values to predict participant submissions on the basis of their rank. Thereafter, we sampled sixteen subreddits and used them to fit a regression model predicting submissions based on subreddit, rank, and the interaction of these predictors. This allowed us to examine how a model's fit changes depending on subreddit size, a possible predictor of misinformation bubbles or echo chambers (Priniski & Holyoak, 2020). Across all subreddits, we expected that user participation would be well approximated by Zipf's law (Carron-Arthur et al., 2014), a power law, which if an accurate description of the data, should indicate that the frequency of user submissions is inversely proportional to their rank.

Results

Reddit Table 1 shows how top 1% of most active users for January 2015 make up a total of 25% of all comments on the website, with the number of comments made by people classified to this hypothesised group ranging from 262 to 13,830. The top 10% of users are responsible for 68% of comments, whereas 27% of our users are 'singleton-posters' – those who left a comment once only during the observation period, replicating early findings from Whittaker and colleagues (1998).

We calculated the Pearson correlation coefficient between user-rank and post frequency r = -.96. A regression model predicting post frequency on rank accounted for 92% of the variance. Including subreddits in the model improved model fit, accounting for 96% of the variance. Figure 1 shows a sample of subreddits from January dataset and their adherence to Zipf's law even though the number of participants in any given subreddit varies greatly. For example, during this time period, r/AskReddit had $603,442$ active participants whereas r/HouseOfCards had $1,501$. Still, it should be acknowledged that including subreddit in the model improved its fit, suggesting that different communities may not abide by Zipf's law to precisely the same degree.

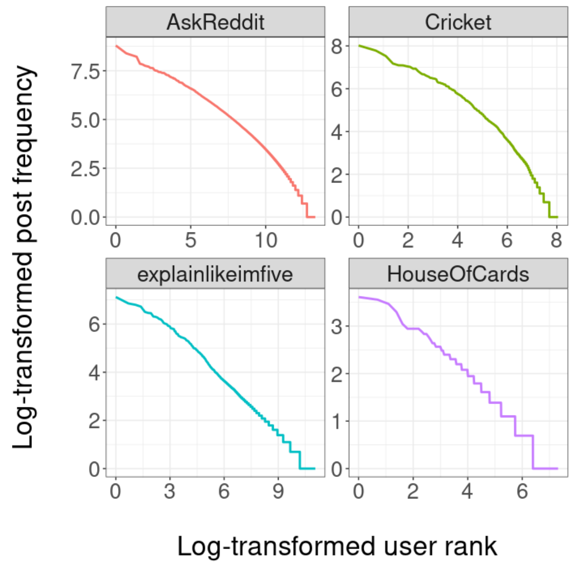

Figure 1: Log-transformed user-rank vs post frequency across different subreddits: The plot shows how user-participation consistently follows Zipf's law regardless of subreddit size.

Table 1: Reddit, January 2015

Percentile (Group)	Comments, %	Range
1 (Superusers)	25	262-13,830
2-10 (Contributors)	43	42-262
11-100 (Lurkers)	32	1-42

The same procedure was performed on the October 2014 dataset. Top 1% most active users make up a total of 25% of all comments, with the number of comments ranging from 256 to 11,217, and top 10% were responsible for 68% of comments, replicating the effects reported above. The Pearson correlation coefficient between user-rank and their post frequency was r = -.96. A regression model predicting post frequency on rank accounted for 92% of the variance. Including subreddits in the model improved model fit, accounting for 96% of the variance.

Twitter Due to known biases in certain platforms, we replicated the preceding findings using an existing Twitter dataset from the same time period. Again, user rank and post frequency followed Zipf's law, with Pearson's r = -.97. Most active 10% of users were responsible for 79% of tweets marked with hashtags #Gamergate and #NotYourShield (Table 2).

Confirming prior results of natural corpora data, our re-

Table 2: Twitter, October 2014

Percentile (Group)	Comments, %	Range
1 (Superusers)	37	144-2,480
2-10 (Contributors)	42	10-144
11-100 (Lurkers)	21	1-10

sults show how user-participation follows Zipf's law and that vast majority of engagement on social media sites commonly used in cognitive and social science at any level comes from a small fraction of the user base. These findings support the claim that the 1% rule is a fairly generalisable principle of user activity in an online environment. However, our Reddit data shows the users we hypothesised to be 'Lurkers' to have quite the impact, making 32% of all comments. Furthermore, the range of comments suggest many of these users are anything but idle observers. Hence, while the rule is a good generalisation, it is not a precise fit, and a classification model would be a more prudent method of labeling users based on their activity (e.g., Pew Research, 2019).

Study 2: Participation inequality and its impact on perceived public discourse: A case study using sentiment analysis following the Charlie Hebdo attack

Methods

How might a vocal minority producing majority of content on social media sites impact researchers's psychological conclusions? To address this question and generalise our findings beyond the initial dataset we've considered thus far, we used a dataset which was collected prior to and immediately following the Charlie Hebdo attack – a terrorist attack on the French satirical weekly newspaper in January 2015. We chose this event due to it being the focus of multiple previous studies on how Islam is portrayed in the social media (e.g. Cervi et al., 2021; Giglietto & Lee, 2017; Marzouki et al., 2020), with Aydin and colleagues (2022) saying social media contributions have been integral in framing how Islam is perceived online. Using our January 2015 dataset, we filtered user comments to those which included any of the following keywords (or their derivatives): 'islam', 'muslim', or 'arab'; we did this to focus on the sentiment of users toward minority groups the attack prompted. We were left with 151,445 comments from 64,104 users. Sentiment scores were calculated using BING-dictionary (Hu & Liu, 2004), which determines a word to be either positive or negative, which we then transform into a score of 1 or -1. Due to aforementioned finding that our data does not neatly fit the 90-9-1 paradigm, user activity is henceforth treated as a continuum[1], and any group differences are analysed as four equal-sized quartiles.

[1]Thank you to an anonymous reviewer for making this point.

Results

We compiled descriptive statistics (Table 3) to illustrate the impact the most active users have on sentiment counts, with the top 25% most active users creating 91% of positive mentions *and* 92% of negative mentions. Overall, 29% of the non-neutral words analysable by our sentiment analysis were labeled positive.

We performed a Bonferroni-adjusted contrast analysis to test the difference in mean sentiment ratings for quartiles across four different temporal categories: *Pre-attack* (January 1st-6th), *During* (7th-12th), *Post-attack* (13th-18th), and *Aftermath* (19th-31st). We calculated mean sentiment score per submission for all users while controlling for their overall activity. There were differences in mean sentiment scores between quartiles in all temporal categories at $p < .01$, most notably between first quartile (Q1) when compared to the fourth and least active quartile. In all significant findings, sentiment estimates followed a linear progression, with the more active quartile showing less negative sentiments than the less active one. General trend is shown in Figure 2.

In the *Pre-attack* temporal condition, the most active quartile was found to be less negative in comparison to the least active quartile ($t(103{,}916) = 3.3$, CI = [.047, 2.7]), and the same was true for Q2 ($t(103{,}916) = 3.7$, CI = [.25, 2.9]), and Q3 ($t(103{,}916) = 3.9$, CI = [.35, 3.3]). In the *During* timeframe, when the Paris attacks were unfolding, the first quartile held less negative sentiments than Q2 ($t(103{,}916) = 4.0$, CI = [.091, .74]), Q3 ($t(103{,}916) = 3.2$, CI = [.017, 1.2]), and Q4 ($t(103{,}916) = 3.3$, CI = [.73, 2.4]). Lastly, they held

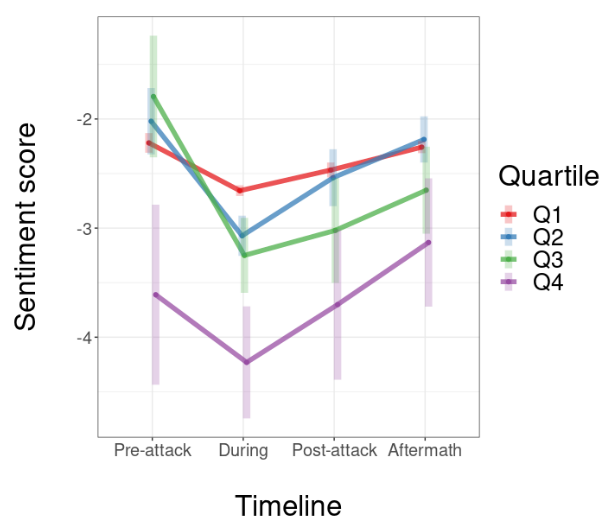

Figure 2: Mean sentiment scores across four activity quartiles throughout January, 2015. There is a significant difference in user sentiment based on their overall activity, with the least active conversationalists showing more negative sentiments in the context of the keywords 'islam', 'muslim', or 'arab'. Overall, user sentiments decrease during the Paris attacks, but return to their baseline value by the end of the month.

Table 3: Descriptive statistics from sentiment analysis.

Quart.	Positive (%)	Negative (%)	Comments (%)
Q1	253,538 (91)	624,522 (92)	138,260 (91)
Q2	17,304 (6.2)	40,108 (5.9)	9,346 (6.2)
Q3	4,613 (1.7)	10,661 (1.6)	2,660 (1.8)
Q4	2,322 (.84)	5,594 (.82)	1,179 (.78)

less negative sentiments than Q4 in the *Post-attack* condition (t(103,916) = 3.4, CI = [.11, 2.4]) after police operations had ended. In the *Aftermath* condition, when discussion around the keywords had returned to its baseline rate, the sentiments between different quartiles converged, with only Q2 showing different sentiment to Q4 (t(103,916) = 3.3, CI = [.047, 1.8]).

There were, also, differences in sentiment scores within groups between different temporal categories. In the *Pre-attack – During* comparison, mean sentiment scores were different for Q1 (t(103,916) = 8.7, CI = [.28, .60]), Q2 (t(103,916) = 6.5, CI = [.54, 1.6]), and Q3 (t(103,916) = 5.0, CI = [.54, 2.4]). We observed no difference in mean sentiment scores across groups at the *Pre-attack – Aftermath* comparison, which coincides with work from Silva (2018), who found no difference in public opinion after the attacks across a wide range of issues.

These findings indicate that discussion around the Paris attacks and in the context of keywords 'islam', 'muslim', or 'arab' was dominated by a very vocal minority of users. We found groups to differ not just on activity but on measures of sentiment, too – by only focusing on global differences in sentiment scores we could fail to have both an accurate measure of attitudes in general, but also focusing on the most active users could obscure the effects of events on a less vocal, majority group.

Study 3: Understanding the source of sentiment: An investigation into Gamergate related subreddits

Methods

To better understand the differences between user groups and how their behaviour affects the perceived sentiment of communities, we investigate user comments surrounding Gamergate made on Reddit during October 2014. Gamergate was a manifestation of a culture war within gaming community, which quickly turned into a misogynistic hate campaign where female developers and Gamergate critics were targeted and harassed online and offline. The event has been widely used in research on political psychology (e.g., Massanari, 2016; Mortensen, 2016). We used the following keywords and their derivatives to filter relevant comments: 'anita', 'sarkeesian', 'zoe', 'quinn', 'quinnspiracy', 'feminist frequency', 'gamergate', 'kotaku', and 'game journalism'. Exploratory analyses showed largest proportion of comments containing these keywords coming from r/KotakuInAction (a pro-gamergate subreddit) with with $47,240$ comments out of $210,870$ (22%) including at least one of the keywords. Likewise, r/GamerGhazi (an anti-gamergate subreddit) contained $8,283$ keyword-related comments (28%). Together with these two subreddits, we included r/AskReddit, a "control" subreddit which contains $85,213$ topical comments (2.5%), giving us a total of $140,736$ keyword related comments to analyse. We calculate mean sentiment per user comment and perform a contrast analysis to investigate differences between groups and subreddits.

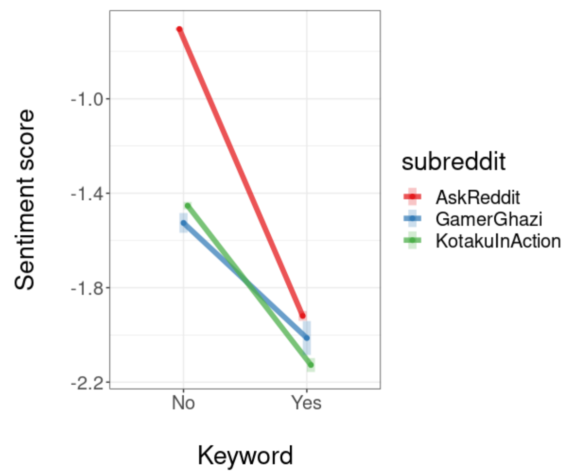

Figure 3: Mean sentiment scores per submission in different subreddits, with users in topic-specific subreddits holding more negative sentiments than those in the more general purpose subreddit r/AskReddit.

Results

We found the most active quartile to produce 87% of all discussion in the context of the keywords, and we found converging lines of evidence to Study 2, whereby low-frequency users produce more negatively labelled language when engaging in a conversation, whether the submission contained one of the keywords or not. There is, also, a strong positive correlation between the baseline sentiment score in relation to how user groups react to keywords being present, with Pearson's r = .75. This exacerbates differences between high and low frequency users and may make it hard to make generalisable statements about even specific subreddits – not to even mention the site as a whole. In all subreddits, there exists a statistically significant difference in sentiment, with discussion being less negative without keywords present in r/AskReddit (t(1,022,975) = 110, CI = [1.2, 1.2]) as well as the Gamergate specific subreddits r/GamerGhazi (t(1,022,975) = 12, CI = [.38, .60]), and r/KotakuInAction (t(1,022,975) = 40, CI = [.63, .72]), visualised in Figure 3. Bonferroni corrected contrast analysis yielded no differences in sentiment between the two Gamergate related subreddits in the presence of a keyword, with each submissions containing on average two negatively labelled words.

Discussion

We sought to understand how users interact with and produce online content and how this can inform cognitive and social scientists's understanding of the lurking biases in their datasets. In Study 1, we observed that the 1% rule accurately characterises contemporary online user activity, with vast majority of content being produced by an overactive and—to our surprise—a *less* negative minority of users. One reason most active users could behave in a cordial manner is due to them being *a priori* required to act as such for them to communicate their thoughts effectively; a user who repeatedly says distasteful things is unlikely to have others want to engage in a conversation with them.

In Study 2, we found different activity groups tending to similar sentiment use in the context of the keywords mentioned as the events unfolded, and a similar effect was observed by Jones and Silver (2019) during the Hawaii false missile alert, whereafter anxiety levels regressed to the mean among Twitter users, even though their pre-alert anxiety levels were vastly different. This might just reflect a temporary agreement between interlocutors, who tend to copy each other's word use. While this might seem to only be of relevance to political or social psychology, data from sites like Reddit is frequently used in cognitive science research, with a study (Jung et al., 2021) investigating language use in loneliness-related subreddits identifying linguistic markers on how lonely individual express their feelings. However, as we can see from our data, users differ in how they converse their feelings, especially with regards to sensitive topics, depending on how their activity level. Moreover, in Study 3 we found discussion to vastly differ between subreddits – hence, while some laud the opportunity to study special interest groups like subreddits focused on ADHD or depression, researchers run the risk of analysing discussions which do not generalise across individuals or subreddits.

These findings add to already existing differences within social media users, who have been found to differ in Big Five personality traits, , with openness to experience and extraversion predicting frequency of use well as engagement with other users (Correa et al., 2010; Gosling et al., 2011). Thus, while participants from social media sites might be more diverse than university students, the individuals studied might be especially skewed in their personality traits, as they represent only a small minority of all social media users – which, as was mentioned in the Introduction, when combined with large sample sizes might lead to an increase rather than decrease in error (Kaplan et al., 2014).

One major difference between our work and that of Carron-Arthur et al. (2014) and van Mierlo (2014) are that whereas they studied "closed systems", sites like Reddit or Twitter do not require users to have a registered account in order for them to observe or engage with the content. This is a significant limitation in our analyses, as we cannot analyse null users - those who consume the user-generated content but do not themselves participate in the process. This leads to our estimates being especially conservative: While our January dataset includes comments from $2,500,852$ users, Reddit is estimated to have had 120 million monthly active users in 2015. Supposing, then, that there exists another hundred million silent observers, these findings (which show how users differ in their conversational style both in frequency and sentiment) only highlight how minute and extraordinary community those participating in social media platforms really are—and what power they have in shaping discussion and research.

References

Aydin, Z., Fuess, A., Förster, M., & Sunier, T. (2022). When Birds of a Feather Instagram Together: Debating the Image of Islam in Echo Chambers and Through Trench Warfare on Social Media. *Social Media + Society*, *8*(3).

Baoi, A. (2014). 72 hours of gamergate: Digging through 316,669 tweets from three days of twitter's two-month-old trainwreck.

Baumgartner, J. (2016). *I have every publicly available Reddit comment for research. 1.7 billion comments at 250 GB compressed. Any interest in this?*

Carron-Arthur, B., Cunningham, J. A., & Griffiths, K. M. (2014). Describing the distribution of engagement in an Internet support group by post frequency: A comparison of the 90-9-1 Principle and Zipf's Law. *Internet Interventions*, *1*(4). doi: 10.1016/j.invent.2014.09.003

Castanho Silva, B. (2018). The (Non)Impact of the 2015 Paris Terrorist Attacks on Political Attitudes. *Personality and Social Psychology Bulletin*, *44*(6). doi: 10.1177/0146167217752118

Cervi, L., Tejedor, S., & Gracia, M. (2021). What Kind of Islamophobia? Representation of Muslims and Islam in Italian and Spanish Media. *Religions*, *12*(6). doi: 10.3390/rel12060427

Cheon, B. K., Melani, I., & Hong, Y.-y. (2020). How USA-Centric Is Psychology? An Archival Study of Implicit Assumptions of Generalizability of Findings to Human Nature Based on Origins of Study Samples. *Social Psychological and Personality Science*, *11*(7). doi: 10.1177/1948550620927269

Davis, J. L., & Love, T. P. (2019). Generalizing from social media data: a formal theory approach. *Information, Communication & Society*, *22*(5). doi: 10.1080/1369118X.2018.1555610

Dayter, D., & Messerli, T. C. (2022). Persuasive language and features of formality on the r/ChangeMyView subreddit. *Internet Pragmatics*, *5*(1). doi: 10.1075/ip.00072.day

Gaffney, D., & Matias, J. N. (2018). Caveat emptor, computational social science: Large-scale missing data in a widely-published Reddit corpus. *PLOS ONE*, *13*(7). doi: 10.1371/journal.pone.0200162

Giglietto, F., & Lee, Y. (2017). A Hashtag Worth a Thousand Words: Discursive Strategies Around #JeNeSuisPasCharlie After the 2015 *Charlie Hebdo* Shooting. *Social Media + Society*, *3*(1). doi: 10.1177/2056305116686992

Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information. *Cyberpsychology, Behavior, and Social Networking*, *14*(9). doi: 10.1089/cyber.2010.0087

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3). doi: 10.1017/S0140525X0999152X

Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining.* Seattle WA USA: ACM. doi: 10.1145/1014052.1014073

Kaplan, R. M., Chambers, D. A., & Glasgow, R. E. (2014). Big Data and Large Sample Size: A Cautionary Note on the Potential for Bias. *Clinical and Translational Science*, *7*(4). doi: 10.1111/cts.12178

Low, D. M., Rumker, L., Talkar, T., Torous, J., Cecchi, G., & Ghosh, S. S. (2020). *Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: an observational study* (preprint). PsyArXiv. doi: 10.31234/osf.io/xvwcy

Marzouki, Y., Barach, E., Srinivasan, V., Shaikh, S., & Feldman, L. B. (2020). The dynamics of negative stereotypes as revealed by tweeting behavior in the aftermath of the Charlie Hebdo terrorist attack. *Heliyon*, *6*(8). doi: 10.1016/j.heliyon.2020.e04311

Massanari, A. (2017). #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, *19*(3). doi: 10.1177/1461444815608807

Mortensen, T. E. (2018). Anger, Fear, and Games: The Long Event of #GamerGate. *Games and Culture*, *13*(8). doi: 10.1177/1555412016640408

Mosleh, M., Pennycook, G., Arechar, A. A., & Rand, D. G. (2021). Cognitive reflection correlates with behavior on Twitter. *Nature Communications*, *12*(1). doi: 10.1038/s41467-020-20043-0

Nielsen, J. . (2006). Participation inequality: Encouraging more users to contribute. *Jakob Nielsen's alertbox*.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*. doi: 10.1016/j.jecp.2017.04.017

Pew Research Center. (2019). National Politics on Twitter: Small Share of U.S. Adults Produce Majority of Tweets.

Pew Research Center. (2021). Social Media Use in 2018.

Priniski, J. H., & Holyoak, K. J. (2022). A darkening spring: How preexisting distrust shaped COVID-19 skepticism. *PLOS ONE*, *17*(1). doi: 10.1371/journal.pone.0263191

Priniski, J. H., & Horne, Z. (2019). *Crowdsourcing effective educational interventions* (preprint). PsyArXiv. doi: 10.31234/osf.io/edk3m

Priniski, Z., J.H. & Horne. (2018). Attitude Change on Reddit's Change My View. In T.T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), Proceedings of the 40th Annual Conference of the Cognitive Science Society (pp. 2276-2281). Austin, TX: Cognitive Science Society. In X. Z.-. C. W. K. T.T. Rogers M. Rau (Ed.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of *Homo sapiens* : Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences*, *115*(45). doi: 10.1073/pnas.1721165115

Shatz, I. (2016). Fast, Free, and Targeted: Reddit as a Source for Recruiting Participants Online. *Social Science Computer Review*, *35*(4). doi: 10.1177/0894439316650163

Sheehan, K. B. (2018). Crowdsourcing research: Data collection with Amazon's Mechanical Turk. *Communication Monographs*, *85*(1). doi: 10.1080/03637751.2017.1342043

Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of Depression-Related Posts in Reddit Social Media Forum. *IEEE Access*, *7*. doi: 10.1109/ACCESS.2019.2909180

Tan, N. V. D.-N.-M. C. . L. L., C. (2016). Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web.* Montreal, Canada.

Turcan, E., & McKeown, K. (2019). Dreaddit: A Reddit Dataset for Stress Analysis in Social Media. In *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019).* Hong Kong: Association for Computational Linguistics. doi: 10.18653/v1/D19-6213

van Mierlo, T. (2014). The 1% Rule in Four Digital Health Social Networks: An Observational Study. *Journal of Medical Internet Research*, *16*(2). doi: 10.2196/jmir.2966

Whittaker, T. L. H.-W.-. C. L., S. (1998). The dynamics of mass interaction. *CSCW '98 Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Network*.