Article

new media & society
2024, Vol. 26(1) 389–404

DOI: 10.1177/14614448211062070

Does Godwin's law (rule of Nazi analogies) apply in observable reality? An empirical study of selected words in 199 million Reddit posts

Gabriele Fariello
Harvard University, USA

Dariusz Jemielniak
Kozminski University, Poland

Adam Sulkowski
Babson College, USA

Abstract

As Godwin's Law states, "as a discussion on the Internet grows longer, the likelihood of a person being compared to Hitler, or another Nazi reference, increases." However, even though the theoretical probability of an infinitely long conversation including any term should approach 1.0, in practice, conversations cannot be infinite in length, and this long-accepted axiom is impossible to observe. By analyzing 199 million Reddit posts, we note that, after a certain point, the probability of observing the terms "Nazi" or "Hitler" actually decreases significantly with conversation length. In addition, a corollary of Godwin's Law holds that "the invocation of Godwin's Law is usually done by an individual that is losing the argument," and, thus, that comparisons to Nazis are a signal of a discussion's end. In other words, comparing one's interlocutor to Hitler is supposed to be a conversation-killer. While it is difficult to determine whether a discussion on a given topic ended or not in a large dataset, we observe a marked increase

Corresponding author:
Dariusz Jemielniak, Kozminski University, Warsaw 03301, Poland.
Email: darekj@kozminski.edu.pl

in conversation length when the words "Hitler" or "Nazi" are newly interjected. Given that both of these observations challenge widely accepted and intuitive truisms, other words were run through the same set of tests. Within the context of the initial question, these results suggest that it is not inevitable that conversations eventually disintegrate into *reductio ad Hitlerum*, and that such comparisons are not conversation-killers. The results moreover suggest that we may underestimate, in the popular imagination, how much conversations may actually become narrower and therefore may tend to have a more impoverished or limited vocabulary as they stretch on. All of these observations provoke questions for further research.

Keywords
Computational linguistics, digital culture, online conversations, Reddit

Introduction

Godwin's law, sometimes called Godwin's rule of Hitler analogies (Godwin, 1994), is a popular Internet adage stating that "as an online discussion grows longer, the probability of a comparison involving Nazis or Hitler approaches one" (Godwin, 1995). In addition, a corollary of Godwin's law is that whoever invokes Hitler is losing the argument:

> There is a tradition in many groups that, once this occurs, that thread is over, and whoever made a reference to Nazis has automatically lost whatever argument was in progress. Godwin's Law thus practically guarantees the existence of an upper bound on thread length in those groups. (Neiwert, 2016: 240)

Originally coined in the 1990s on Usenet, attorney and author Michael Godwin intended his seemingly self-evident eponymous aphorism as an admonition to think twice before employing inappropriately extreme comparisons in debates, as well as a warning against a *reductio ad Hitlerum* fallacy. The original phrasing has spawned multiple variations and corollaries (Godwin, 1995), is still one of the most evoked rules on the Internet to this day (Ohlheiser, 2017), and is widely accepted as a universal truth.

Godwin's law has gained widespread popularity because, in Western culture, Hitler and Nazis are often considered to be the ultimate reference point for evil (Burke and Goodman, 2012; Johnson, 2010), and Internet culture is often perceived as toxic, divisive, and driven by conflicts (Aswath et al., 2020; Reagle, 2015). In fact, conflict is considered to be one of the driving forces behind some peer production websites, such as Wikipedia (Jemielniak, 2014): people are much more likely to engage in collective knowledge creation if adding sourced information is the only way for them to win a dispute. Conflicts are so embedded in online folklore that inciting division is even considered one form of online custom (Phillips, 2015), and occasionally perceived as a specific form of art or social activism (Hodge and Hallgrimsdottir, 2020; Sanfilippo et al., 2018).

Yet, even though Godwin's Law is often treated as an axiom, it was not developed upon the basis of observations, nor has it been verified through large-scale and

systematic research. In our study, we show that, in fact, it is not observable within a vast dataset of Reddit conversations, such that we postulate it is quite unlikely to be more broadly observable.

Our study fits within a rich literature devoted to better understanding the evolving role of social media within society (Fuchs, 2011; van Dijck, 2013; Webster, 2014). We decided to focus on Reddit, "the front page of the internet," arguably the largest online platform dedicated solely to discussion, and also memes and other media sharing. There is a large universe of literature dedicated to the study and understanding of conversations on Reddit (Medvedev et al., 2017). Reddit threads in particular have been studied to see if hierarchies observed on the platform parallel those known to exist in other conversational contexts (Weninger, 2014; Weninger et al., 2013), and to see if threads can be modeled (Zayats and Ostendorf, 2018). Reddit has also been a platform in which trolling has been studied (Merritt, 2012). Most relevant to the analysis described below, subreddits have been studied to determine the nature of threads that are more likely to keep users engaged (Choi et al., 2015). In another study, researchers used modeling to predict what Reddit threads would be popular (He et al., 2016). Other studies have compared various methods, including clustering, to categorize and make sense of large numbers of threads (Curiskis et al., 2020).

Reddit conversations can be somewhat intimate, but always have "peripheral audiences of inactive conversational partners" and other eavesdroppers (Shelton et al., 2015). The occasional authentic invocation of fascist sentiments and purported reasoning to support those sentiments should perhaps not be shocking, due to the phenomenon of online disinhibition and use of pseudonyms and temporary accounts, including on Reddit (Gagnon, 2013).

One reason to doubt that Godwin's Law would be substantiated is the increased shallowness of communication, which, as documented in an article published in *Nature Communications*, appears to be driven by social media effectively shortening collective attention spans for any given topic (Lorenz-Spreen et al., 2019). In other words, it is possible that conversational threads are ending "too early" for Godwin's Law to manifest itself—and that, had they continued, eventually participants would have mentioned Hitler or Nazis.

Material and methods

We obtained a large dataset ($n = 1.35 \times 10^9$) containing *Reddit.com* posts for the year 2018 from the *PushShift.io* data repository (Baumgartner et al., 2020) for both Reddit Submissions (RS files) and Reddit Comments (RC files) and uncompressed them individually. Using the *jq* command-line tool, we extracted the subreddit for each posting and determined the frequency of activity in each subreddit, selecting the top 12 subreddits ($n = 1.99 \times 10^6$) comprising approximately 15% of all posts while removing duplicates found in the dataset (Figure 1).

We analyzed each of the *JSON* records ($n = 1.35 \times 10^9$) to determine the various data types and values present in the data set (results available in *PushShift.io-Reddit-2018-Attribute-Summary.txt*) and used the information to normalize attributes common to both Submissions and Comments ensuring that data types and values were consistent across

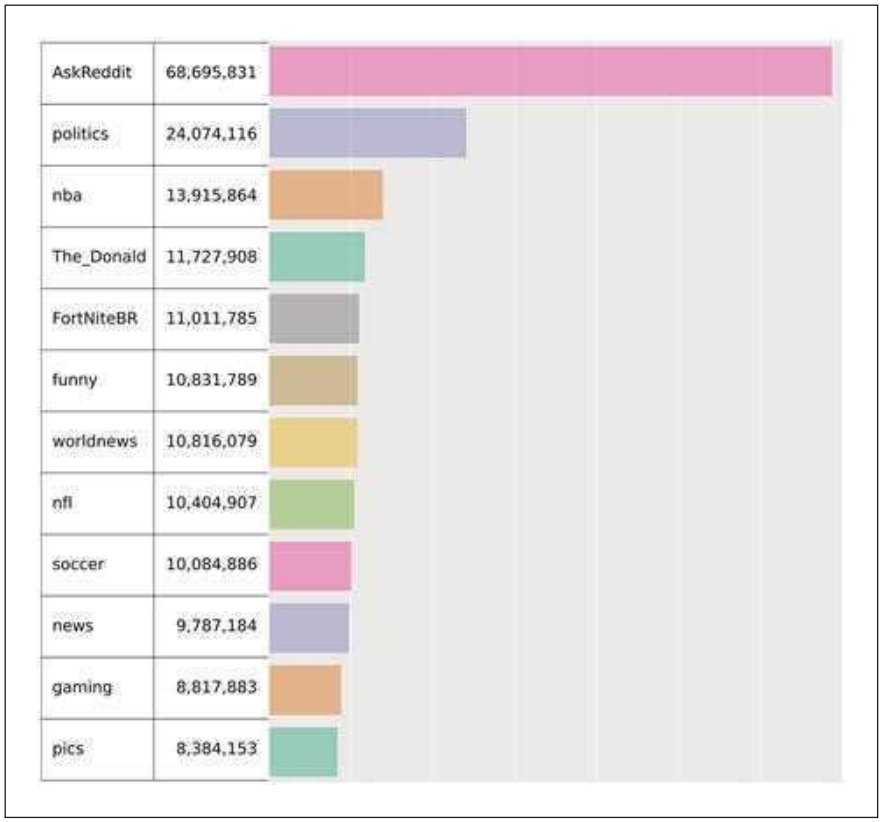

Figure 1. Distribution of posts in top 12 subreddits.

posts over time and post types. We then created comma separated value (CSV) files for each subreddit and post type by month. Here, we provide a detailed summary of all 288 files for the top 12 subreddits. We provide these files for the top 150 subreddits (3600 total files) for other researchers to use, but do not include them in the analysis.

Within the top 12 subreddits of 2018, 1,384,050 (0.7%) posts were deleted and 917,909 (0.4%) removed (1.2% aggregate). Although it is not possible to determine if there are posts missing from specific subreddits, we assume fewer than 0.5% consistent with previous Reddit-wide findings for 2017 (Medvedev et al., 2017). A total of 18,216,231 (9.2%) of posts originated from accounts that were no longer in existence at the time of download. The 100 most prolific authors[1] accounted for 1.8% of the remaining posts and 1,527,801 (0.85%) authors had only one post.

Using the normalized data, we created CSV files for analysis including information on the number of times case-insensitive regular expression matched the body or title of a post for a given set of search terms (e.g. “Hitler” or “Nazi”). In addition to the terms relevant to testing Godwin’s Law, we chose five additional contemporary political terms (“Clinton,” “Democrat,” “Obama,” “Republican,” “Trump”) and two high-frequency

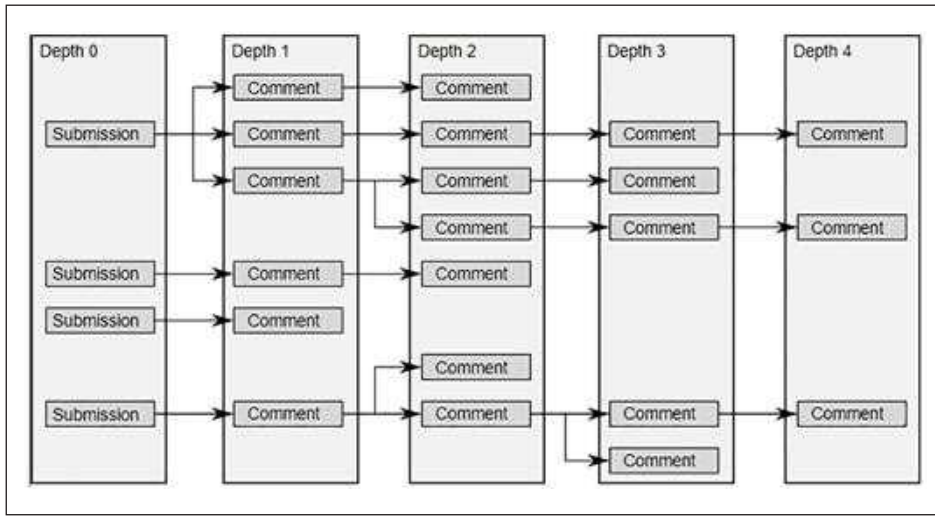

Figure 2. We determine the depth of a given post by counting the number of direct parents. Conversely, all subsequent posts that have a given post as an upstream parent are counted as that post's children.

common vulgar filler terms ("Fuck" and "Shit") in order to ascertain any similarities or differences with "Hitler" or "Nazi." The regular expressions were designed to match various permutations of each of the terms (e.g. "nazi," "nazism," "nazis"). For each post, we calculated the depth of the post within its conversation and the number of posts it had replying to it (children).

Submissions and comments in Reddit posts take the form of an acyclic directed tree structure such as the following (Figure 2):

Where each conversation is started with a submission within a given subreddit, users can extend the conversation by replying via comments either to the original submission or to a comment. In the above example, we consider "Submission A" as being at depth 0 and having 7 children, "Comment A1" as being at depth 1, having 3 children, and one parent, "Comment A1a" as being at depth 2, having no children, and having two parents.

This is consistent with prior handling of Reddit conversation structure and similar graph structures for conversation threadings that exist in email, USENET News, Google Groups, and other forms of online conversations (Curiskis et al., 2020; Medvedev et al., 2018; Zayats and Ostendorf, 2018).

Attempting to test the validity of Godwin's law is made complicated by the various ways that Nazis and Hitler can be invoked. There is the issue of whether a thread dedicated to discussing the history of Germany in the 20th century should be disqualified. This prompted us to take further steps. In addition to analyzing general term frequency, we tested for the appearance of off-topic terms as implied by Godwin's Law. For the purposes of determining the probability of an off-topic term appearing within a thread, we define a set of terms as being "off-topic" to the conversation if the first post (Submission) and first reply (Comment) within a thread did not contain the search terms.

For those, the first time a term is encountered while descending into the thread, it is deemed "off-topic" while thereafter it is not.

Further investigation in future studies could include

(1) Performing a Jensen–Shannon and/or Kullback–Leibler divergence analysis,
(2) Estimating the change in probability via depth (similar to 1),
(3) Estimating the off-topic relevance (signal importance) based on depth (related to 1 and 2).

Results

Initially, we were interested in finding the percentages of conversations in the 12 studied subreddits that contained "Nazi" or "Hitler," and analyzing conversation depths. This percentage observed at each depth by subreddit is shown in Figure 1.

Depth of conversations was calculated by counting how many parents a given post has. The number of posts which either had (1) text which matched a case-insensitive search for "Hitler" or "Nazi" or (2) at least one parent who matched those same terms was summed.

The resulting sum was then divided by the total number of posts at the same depth, regardless of matching, to determine the percentage of posts that belonged to a conversation that had "seen" the terms at or above the current depth. The "drop off" of conversations permits the percentages to decrease or increase with depth.

Only depths with at least 100 posts are included (e.g. if there were only 87 posts at conversation depth 57 for "news," that depth is discarded to ensure that only meaningful means with reasonable confidence intervals were included). This results in depth gaps such as those observed in the "news" subreddit. The results of this percentage calculation are shared in Figure 3.

We then decided to study the cumulative probability density, that is, the percentage of conversations for which a post or any parent of the conversation contains "Nazi" or "Hitler" for all 12 threads.

Next, we applied the conversation extension method discussed in the proof. The cumulative percentage of conversations for each of the 12 top subreddits which had seen the terms "Nazi" or "Hitler" at or above each depth are shown in Figure 5.

In Figure 4, even with the confidence interval of 95%, the top boundary of the aggregate of the extended data does not exceed 0.06.

As represented in Figure 4(a), the rate of decline in conversations is not immediately obvious. However, taking the log base 10 (Figure 4(b)) and the log of the log base 10 (Figure 4(c)) of the number of posts per depth shows that the rate of decline is greater than many common exponential decay functions observed in nature.

Applying a third log base 10 (Figure 4(d)) results in a graph that shows remarkable fitness to a straight line. We only tested conversations of depths of 0–300 because, with conversations of depths greater than 300, further attempts to take the log base 10 results in errors due to the limitations of mathematical precision of the systems used.

We also analyzed other search terms, to check if these patterns could be observed in the context of other words. We chose "Trump," "Obama," "Hillary," "Clinton" (anticipating

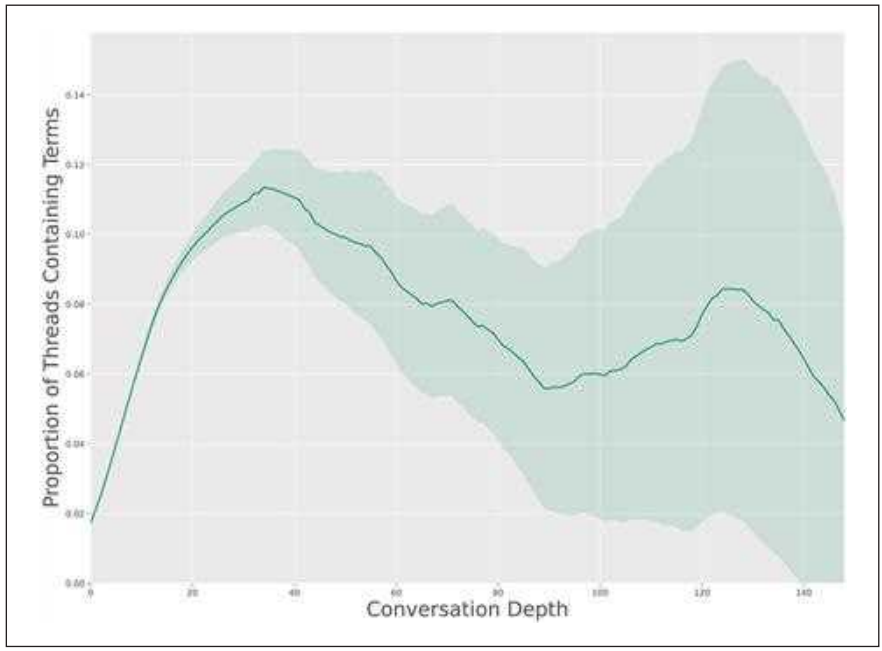

Figure 3. The proportion (percent expressed as a decimal) of threads that, at a given depth, had encountered the search terms "Hitler" or "Nazi" and various derivatives at that depth or at any previous depth using a rolling mean of width 10. We show the 95% confidence interval for each depth to reflect the relative uncertainty with the increase in depth.

that data might reveal one of the politicians to be "the new Hitler" for an updated version of Godwin's Law), as well as "Republican," "Democrat," "shit," and "fuck." In Figures 5 to 7, we show the results of the same tests (conducted with the search terms "Nazi" and "Hitler") for Hitler and the eight other chosen words.

The proportion of posts at a given depth which contained the search terms are represented above in Figure 5(a) as a percentage (expressed as a decimal). To normalize these values to 1, the values in Figure 5(a) were divided by the maximum value observed for each of the search terms such that the maximum for each is 1.0 and the minimum 0.0, as represented in Figure 5(b). The values in Figure 5(a) were averaged over a rolling window of 10 depths to better visualize the change in proportion over depths, depicted along with the 95% confidence interval per depth in Figure 5(c). Just as we normalized the values in 5(a)—as shown in Figure 5(b)—we normalized the values in Figure 5(c) to 1. It is interesting to note that for the first 25 depths, as highlighted by the green rectangle in Figures 5(b), the normalized series behave nearly identically, suggesting that there may be nothing particularly special about the terms "Hitler" or "Nazi" with respect to their frequency or tendency to appear in off-topic conversational tangents.

However, Figure 5(c) is particularly striking. With 95% confidence, we conclude that the lines representing the frequency of the appearance of "Fuck" and "Shit" and "Trump"

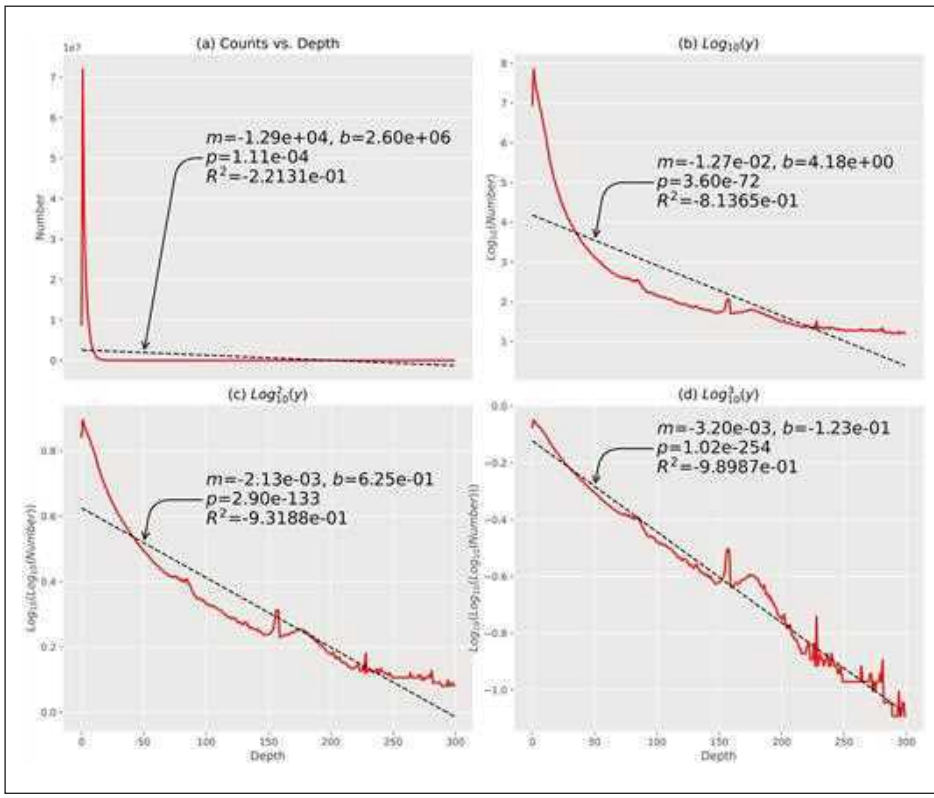

Figure 4. Number of posts at a given depth with the best fit straight line (in black). The rate at which conversations fall renders it impossible to visualize (a) taking the log base 10 and (b) the log of the log base 10 (c) of the number of posts per depth shows that the drop-off is well above many common exponential decay functions observed in nature. It is not until we take a third log base 10 (d) that a straight line demonstrates exceptional fitness.

are nearly the same, while, remarkably, "Hitler," "Nazi," and all other tested words are similar to each other in a band that is clearly distinct from the "Fuck Shit Trump" band.

The proportions of threads at a given depth that contained the search terms are represented in Figure 6(a). A thread is defined as containing the terms at a given depth if the current post or any parent post contained the search terms. As before, we normalized values to 1 by dividing by the maximum value observed for each of the search terms such that the maximum for each is 1.0 and the minimum 0.0, as shown in Figure 6(b). As before, we averaged the values in Figure 6(a) over a rolling window of 10 depths to better visualize the change in proportion over depths, as shown in Figure 6(c). Finally, the values of Figure 6(c) were normalized to 1.

Just as was the case with individual posts, we observe that for the first approximately 50 depths, all search terms behave nearly identically once normalized, as highlighted by the green square in Figure 6(b). Worthy of note is that the 95% confidence interval for all

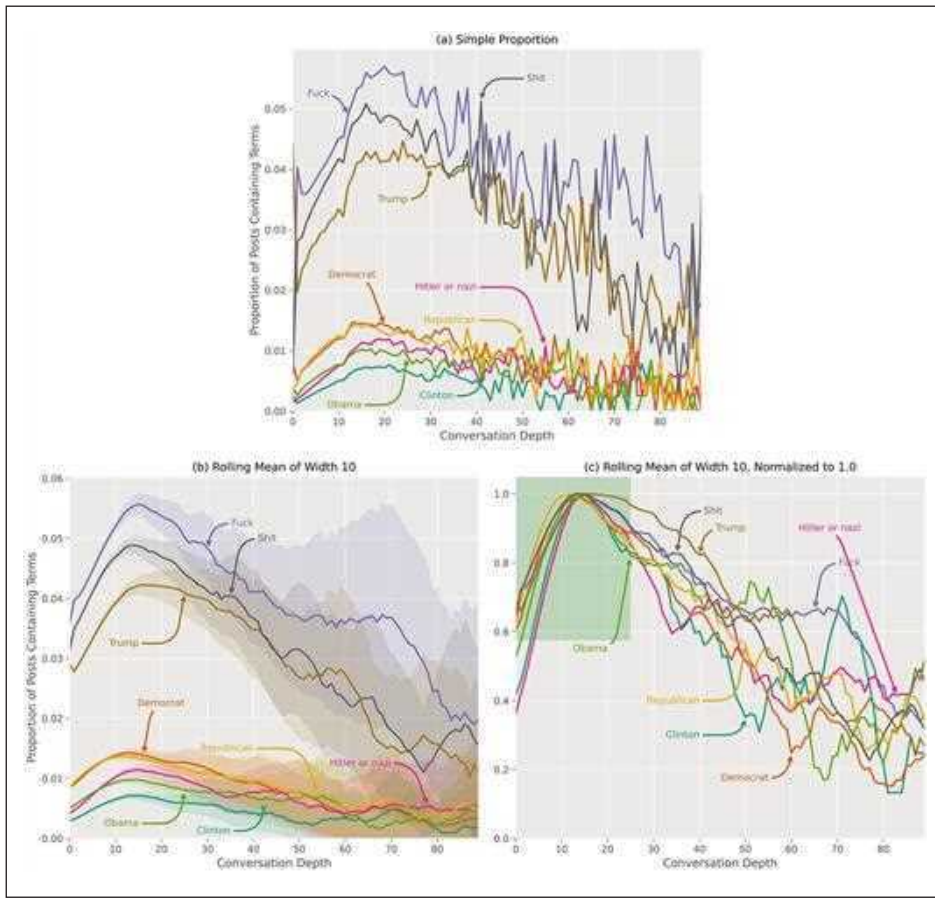

Figure 5. (a) The proportion of posts at a given depth which contained the search terms. (b) The values of (a) averaged over a rolling window of 10 depths to better visualize the change in proportion over depths. (c) The values of (b) normalized to 1.0 by dividing the maximum value for each series by the value at each depth. It is interesting to note that for the first 25 depths, as highlighted by the green rectangle, the normalized series behave nearly identically, suggesting that there may be nothing particularly special about the terms "Hitler" or "Nazi" with respect to frequency or off-topic tendencies.

search terms at depths greater than 113 extends to below the 0.0 on the y-axis but never extends above 1.0 for all depths with at least 100 posts. As will be explained in the "Discussion" section, this further undermines the key assertion and assumption of Godwin's Law.

Our final step was to test the corollary of Godwin's Law: that certain words—specifically, invoking Hitler or Nazis—end a conversation.

We assessed the average number of children, as a surrogate for conversation length, for posts containing or missing the given search terms. As above, we defined

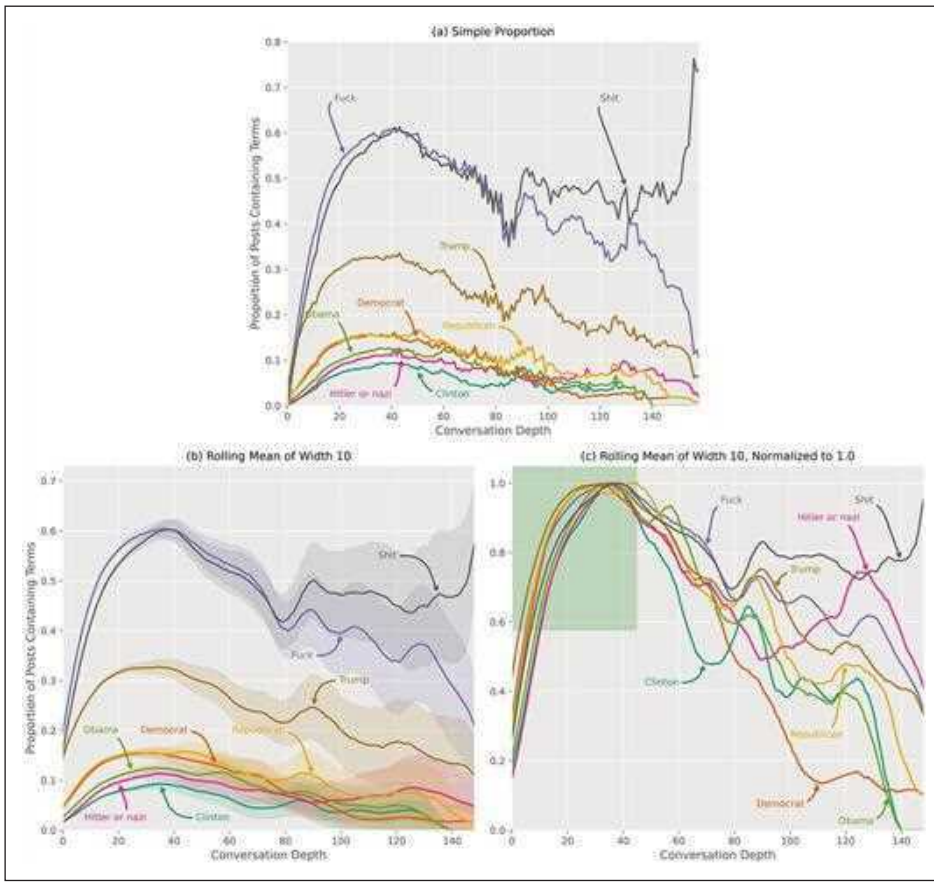

Figure 6. (a) The proportion (percent expressed as a decimal) of threads at a given depth which contained the search terms. A thread is defined as containing the terms at a given depth if the current post or any parent post contained the search terms. (b):The values of (a) averaged over a rolling window of 10 depths to better visualize the change in proportion over depths. (c) The values of (b) normalized to 1.0 by dividing the maximum value for each series by the value at each depth. In the same way as with individual posts, we observe that for the first approximately 50 depths, all search terms behave nearly identically once normalized. Worthy of note is that the 95% confidence interval for all search terms at depths greater than 113 extends to below the 0.0 on the y-axis but never extends above 1.0 for all depths with at least 100 posts.

the appearance of a search term as being off-topic if it had not appeared in any previous parent, while we consider submissions to be on-topic if the search terms appeared at depth 0 (the submission) or 1 (the comments responding directly to the submission); using this approach, only the first appearance of the search terms in a conversation thread at depths greater than 1 are considered off-topic.

Figure 7(a) shows that, at depths greater than 1, the introduction of off-topic (new) terms is strongly correlated with considerably longer conversations—and that this is true

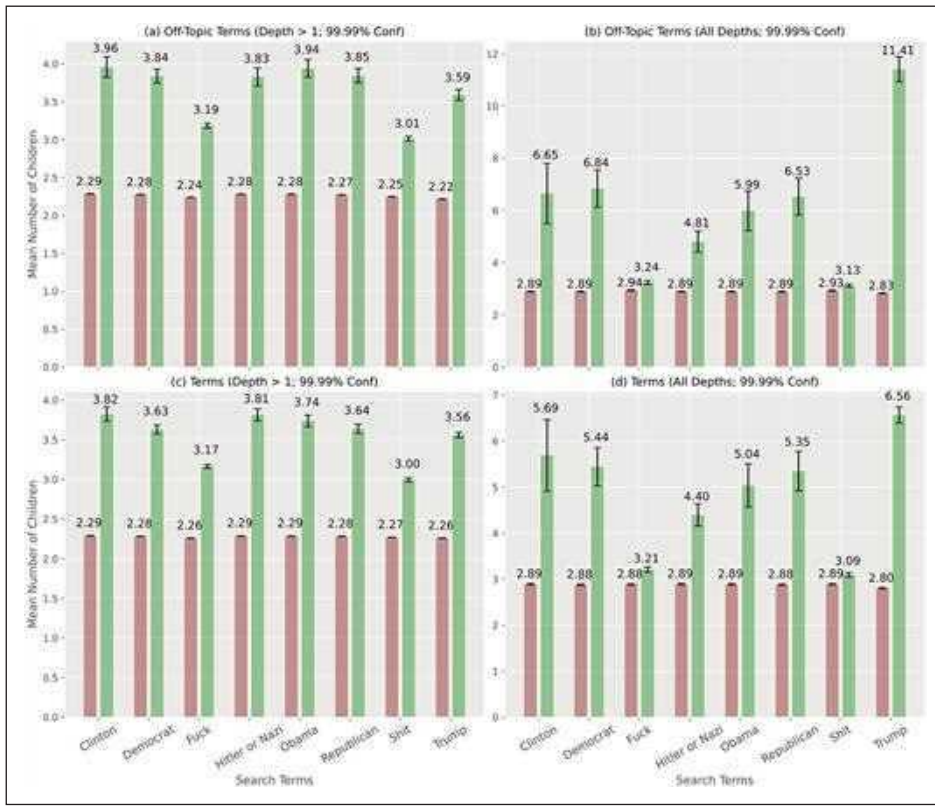

Figure 7. The appearance of off-topic terms in a conversation at depths greater than 1 is strongly correlated with considerably longer conversations, as shown in Figure 7(a). This tendency is present even when including depths 0 and 1, as shown in Figure 7(b). "Fuck" and "Shit" have significantly lower correlations (smaller effect size, although still significant at $p = .0001$). Removing the off-topic criterion has little effect at depths greater than 1 in Figure 7(c), but has a more dramatic effect for all depths, as shown in Figure 7(d).

in the context of all terms that we tested. In Figure 7(b), we show this broad tendency is observable, even when including depths 0 and 1.

The words "Fuck" and "Shit" have significantly reduced correlations (smaller effect size, although still significant at $p = .0001$), as shown in Figure 7(c). Finally, we show that removing the off-topic criterion has little effect at depths greater than 1, but has a more dramatic effect for all depths, as shown in Figure 7(d).

Discussion

If Godwin's Law manifested in reality, we should expect a steady growth of percentages of conversations that contained "Nazi" or "Hitler" as the depth increases in Figure 3. Even if, given the fact that the conversations are not infinite, the probability does not

rise to 100%, one would expect the trajectory of growth should resemble one approaching 100%.

In fact, we observe the opposite. The decline of the mention of "Hitler" or "Nazi" as conversations progress is visible when data from all of the posts are aggregated, and the percentage of conversations for which a post or any parent of the conversation contains "Nazi" or "Hitler" for all 12 threads is shown, as seen in Figure 3. Initially, as conversations progress, the mentions of "Nazi" or "Hitler" increase, but subsequently, the occurrence of these words declines. According to statistical orthodoxy, all well-defined Cumulative Probability Distribution Functions are expected to typically go to one, as depth approaches infinity. In the context of this observable data, however, the opposite appears to be true.

We observe in Figure 4 that conversations die off at a rate significantly faster than a normal exponential decay, rendering the ability to observe the continuation of a given conversation to substantial length extraordinarily improbable.

In Figures 5 and 6, we did not remove from the sample conversations in which Hitler or Nazis were mentioned in the first two generations—in other words, we did not filter-out conversations that most likely started-out as authentically about Nazis or Hitler. Logically, by retaining these conversations in the sample, we should have increased the likelihood that Godwin's Law could be observed. Even including conversations that started-out explicitly mentioning Nazis or Hitler, we do not see Godwin's Law manifested: the probability that Nazis or Hitler will be invoked does not at all approach 1.0. In fact, as is the case, the probability declines over time, at least in observed reality.

A comparison to seven other terms, including five other words related to politics and two multipurpose vulgarities, suggests that "Nazi" or "Hitler" do not appear to function differently from most other terms related to politics (Figures 5 and 6). It is interesting to note that "Shit," "Fuck," and "Trump" resemble each other, at least in terms of trendlines in Figure 5, across a surprising number of depths. We believe that the prevalence of "shit" and "fuck" is due to their being syntactic words, used casually in regular conversations to register everything from surprise to excitement to anger. Nevertheless, the analysis shows that "Hitler" and "Nazi" show a very close similarity to the invocation of other names of popular politicians and political parties—and, remarkably, appear to be unremarkable words in this context.

Finally, the graphs in Figure 7 show that the corollary to Godwin's Law—that the invocation of Hitler or Nazis is a conversation-killer—is not supported by observations. On the contrary, we find that introducing "Hitler" or "Nazi" into a conversation does significantly impact the longevity of the conversation by increasing discussion length. These words tend to prolong conversations similarly to other words that are more contemporary and that are also related to politicians or political parties. This is true even when controlling for conversations in which "Hitler" or "Nazi" were initially on-topic. As was the case in the tests whose results are represented in Figures 5 and 6, we find that "Hitler" and "Nazi" are words that are not unusual in terms of their observable impact on conversations. This result, of course, does not imply causation: the use of "Hitler" or "Nazi" may be a result of a heated discussion, rather than the key reason why a dialogue is protracted. Nevertheless, we suggest, based on intuitive phenomena documented in the

research cited above (on conflict provoking more posts), to assume that inflammatory language, in general, provokes replies.

Conclusion

We have shown that Godwin's Law in its narrow meaning cannot be observed empirically in a large sample of online conversations such as those found on Reddit. This is so for several reasons. First, while in theory and per *infinite monkey theorem*, any conversation running ad infinitum will produce any word at some point, conversations in real life do not continue forever. In fact, they "drop off" too quickly, and much faster than a simple exponential decay function.

Second, the probability of an off-topic word appearing also decreases with conversation length. One possible explanation is that, as each thread operates on a finite collective vocabulary of terms that are most likely to be used and that are related to the given topic, this vocabulary's coverage is exhausted as the conversation continues. As a result, the probability of observing any low-frequency off-topic term in "real life" approaching 1.0 is essentially impossible. In fact, it appears that in a large corpus of "real" online conversations such as on Reddit, the observable cumulative probability will fall considerably short of 1.0. The probability of a low-frequency off-topic term, such as "Nazi" or "Hitler" in conversations that are not about Nazis or Hitler, is nearly impossible to observe in the reality of large datasets. We would encourage others to compare more words and conduct further analysis. Based on our observations of multiple words, it seems extremely unlikely that someone would observe Godwin's Law in real life in conversations that are threaded, like on Reddit. It should be noted, however, that even though we believe that we may find similarities with threaded discussion systems, it is just a working hypothesis. We cannot extrapolate our results easily to all conversations in real life, or even all conversations online. Godwin's Law was coined in the context of discussions on Usenet, constructed similarly to Reddit, but not so to Facebook, Twitter, or any other discussion systems. In addition, we only explored 15% of posts, over a single year, and only from the 12 most popular subreddits, which may experience moderation policies different from those which are less popular. While we believe our results are valid, future research is needed to have our results reproduced across multiple years, threads, and platforms.

Third, beyond the initial depths of approximately 40, we observe that the probability of off-topic terms being injected into a conversation thread decreases with depth. In other words, off-topic tangents appear to become less likely as a conversation progresses and its focus increases. Incivility, insults, and hyperbolic comparisons, such as comparing one's interlocutor to Hitler or a Nazi, are most likely to happen early on. One possibility is that random trolling or insults are much less likely to happen later on. We have three possible explanations for this phenomenon. First, we may be observing specific local culture bubbles: popular threads develop intrinsic behavioral norms, and informal vocabularies or lexicons. This phenomenon is analogous to an echo chamber; however, it may manifest itself in the wording chosen in each thread rather than value and beliefs systems. Second, in many

conversations, there is also a dedicated core—a handful of people most frequently discussing or disagreeing, and they are most active in the thread. Third, most of the really long threads will, through natural evolution, further limit the number of participants in the conversation. We can observe a specific application of the escalation of commitment: by making a comment, one invests into a discussion and the chances are higher one will participate, which contributes to the fact that the number of discussants stabilizes or goes down.

Finally, we find strong statistical evidence that injection of the terms "Hitler" or "Nazi" into a conversation are correlated with longer conversations, not shorter, which is in stark contrast with the second variation of Godwin's Law. The observation that other politically related terms have a similar prolonging effect further diminishes a core tenet or assumption: we find that, in reality, "Hitler" or "Nazi" are not such special or remarkable words, at least in terms of their power to end a conversation.

As a closing remark, the authors have noted that for the studied terms, over a remarkable large number of depths of conversations, "Trump," "Shit," and "Fuck" cluster together in a common band separate-and-distinct from that of all the others, and that this phenomenon merits further investigation.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Dariusz Jemielniak's participation in this research was possible thanks for a research grant from Polish National Science Center , no. 2020/38/A/HS6/00066.

ORCID iD

Dariusz Jemielniak https://orcid.org/0000-0002-3745-7931

Note

1. These include non-human authors such as known "bots" (e.g. "AutoModerator").

References

Aswath S, Godavarthi D and Das B (2020) Analysing conflicts in online football communities of Reddit. In: *2020 international conference on emerging trends in information technology and engineering (ic-ETITE)*, Vellore, India, 24–25 February, pp. 1–6. New York: IEEE.

Baumgartner J, Zannettou S, Keegan B, et al. (2020) The Pushshift Reddit dataset. In: *Proceedings of the international AAAI conference on web and social media*, Palo Alto, CA , 2–3 November, pp. 830–839. Available at: https://ojs.aaai.org/index.php/ICWSM/article/view/7347

Burke S and Goodman S (2012) "Bring back Hitler's gas chambers": asylum seeking, Nazis and Facebook—a discursive analysis. *Discourse & Society* 23(1): 19–33.

Choi D, Han J, Chung T, et al. (2015) Characterizing conversation patterns in Reddit: from the perspectives of content properties and user participation behaviors. In: *COSN '15: Proceedings of the 2015 ACM on conference on online social networks*, November, pp. 233–243. New York: Association for Computing Machinery. Available at: https://doi.org/10.1145/2817946.2817959

Curiskis SA, Drake B, Osborn TR, et al. (2020) An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. *Information Processing & Management* 57(2): 102034.

Fuchs C (2011) *Foundations of Critical Media and Information Studies*. Available at: https://books.google.ca/books?hl=en&lr=&id=K2Zo8wAymHAC&oi=fnd&pg=PP1&ots=z62YpQSjOJ&sig=3F2fegW7C3rSOg1Hs2MGn2xZNaM

Gagnon T (2013) The disinhibition of Reddit users. Available at: https://writingandrhetoric.cah.ucf.edu/wp-content/uploads/sites/17/2019/10/KWS2_Gagnon.pdf

Godwin M (1994) Meme, counter-meme. *Wired*, October 1. Available at: https://www.wired.com/1994/10/godwin-if-2/

Godwin M (1995) Godwin's law (EFF mailing list). Available at: https://web.archive.org/web/20120829094739/http://w2.eff.org/Net_culture/Folklore/Humor/godwins.law

He J, Ostendorf M, He X, et al. (2016) Deep reinforcement learning with a combinatorial action space for predicting popular Reddit threads. *arXiv [cs.CL]*. Available at: http://arxiv.org/abs/1606.03667

Hodge E and Hallgrimsdottir H (2020) Networks of hate: the alt-right, "troll culture," and the cultural geography of social movement spaces online. *Journal of Borderlands Studies* 35: 563–580.

Jemielniak D (2014) *Common Knowledge? An Ethnography of Wikipedia*. Stanford, CA: Stanford University Press.

Johnson BS (2010) *"Just like Hitler": comparisons to Nazism in American culture*. PhD Dissertation, University of Massachusetts Amherst, Amherst, MA. Available at: https://scholarworks.umass.edu/open_access_dissertations/233/

Lorenz-Spreen P, Mønsted BM, Hövel P, et al. (2019) Accelerating dynamics of collective attention. *Nature Communications* 10(1): 1759.

Medvedev AN, Delvenne JC and Lambiotte R (2018) Modelling structure and predicting dynamics of discussion threads in online boards. *Journal of Complex Networks* 7(1): 67–82.

Medvedev AN, Lambiotte R and Delvenne JC (2017) The anatomy of Reddit: an overview of academic research. In: Ghanbarnejad F, Saha Roy R, Karimi F, et al. (eds) *Dynamics on and of Complex Networks*. Cham: Springer, pp. 183–204.

Merritt E (2012) An analysis of the discourse of internet trolling: a case study of Reddit.Com. Available at: https://ida.mtholyoke.edu/xmlui/handle/10166/1058

Neiwert D (2016) *Eliminationists: How Hate Talk Radicalized the American Right*. London: Routledge.

Ohlheiser A (2017) The creator of Godwin's law explains why some Nazi comparisons don't break his famous internet rule. *The Washington Post*, August 14. Available at: https://www.washingtonpost.com/news/the-intersect/wp/2017/08/14/the-creator-of-godwins-law-explains-why-some-nazi-comparisons-dont-break-his-famous-internet-rule/

Phillips W (2015) *This Is Why We Can't Have Nice Things: Mapping the Relationship Between Online Trolling and Mainstream Culture*. Cambridge, MA: MIT Press.

Reagle JM (2015) *Reading the Comments: Likers, Haters, and Manipulators at the Bottom of the Web*. Cambridge, MA: MIT Press.

Sanfilippo MR, Fichman P and Yang S (2018) Multidimensionality of online trolling behaviors. *The Information Society* 34(1): 27–39.

Shelton M, Lo K and Nardi B (2015) Online media forums as separate social lives: a qualitative study of disclosure within and beyond Reddit. In: *iConference 2015 proceedings*. Available at: https://www.ideals.illinois.edu/handle/2142/73676

van Dijck J (2013) *The Culture of Connectivity: A Critical History of Social Media*. Oxford: Oxford University Press.

Webster F (2014) *Theories of the Information Society*. London: Routledge.

Weninger T (2014) An exploration of submissions and discussions in social news: mining collective intelligence of Reddit. *Social Network Analysis and Mining* 4(1): 173.

Weninger T, Zhu XA and Han J (2013) An exploration of discussion threads in social news sites: a case study of the Reddit community. In: *2013 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM 2013)*, pp. 579–583. ieeexplore.ieee.org

Zayats V and Ostendorf M (2018) Conversation modeling on Reddit using a graph-structured LSTM. *Transactions of the Association for Computational Linguistics* 6: 121–132.

Author biographies

Gabriele Fariello is an instructor at Harvard University (DCE) in Machine Learning and Artificial Intelligence and an affiliated researcher (FAS). He is the Senior Advisor on multiple data science programs with the US Food and Drug Administration (CDRH). He has been formerly the Inaugural Assistant Dean for Computing and CIO of Harvard's John A Paulson School of Engineering and Applied Science, the Head of Neuroinformatics at Harvard's Faculty of Arts and Sciences (Center for Brain Science), and the Director of Clinical Research Informatics at Massachusetts General Hospital. His interests have ranged from systems engineering to computational genomics, from neuroimaging to EHR analysis, and, more recently, computational sociology.

Dariusz Jemielniak is Full Professor and head of Management in Networked and Digital Environments (MINDS) department, Kozminski University, and faculty associate at Berkman-Klein Center for Internet and Society, Harvard University. He is a corresponding member of the Polish Academy of Sciences. His recent books include Collaborative Society (2020, MIT Press, with A. Przegalinska), Thick Big Data (2020, Oxford University Press), Common Knowlege? An Ethnography of Wikipedia (2014, Stanford University Press). His current research projects include climate change denialism online, anti-vaxxer internet communities, and bot detection.

Adam Sulkowski is Associate Professor of Law and Sustainability at Babson College. His recent articles include *The Tao of DAO: Hardcoding Business Ethics on Blockchain* (Business and Finance Law Review, 2020), *Industry 4.0 Era Technology (AI, Big Data, Blockchain, DAO): Why The Law Needs New Memes* (Kansas Journal of Law & Public Policy Online, 2020), and *Blockchain, Business Supply Chains, Sustainability, and Law: The Future of Governance, Legal Frameworks, and Lawyers?* (Delaware Journal of Corporate Law, 2019). His current research focuses on how information technology can impact sustainability data reporting and usage.