ValentProjects

5G Conspiracy Theories and Anonymity

June 2020

info@valent-projects.com

Content

Executive Summary

This report concludes that Social Media Companies must take urgent and concerted action to rein in abuse of anonymity on their platforms - because right now it is a major risk factor in the spread of dangerous, false information about Coronavirus. Research into Twitter conversations involving conspiracy theories about 5G and Coronavirus, during the period when the virus was spreading rapidly and lockdown measures were being introduced, has found disproportionate levels of activity by anonymous users, who drove much of the conversation and pushed the most extreme content.

We analysed and compared general conversations about Coronavirus, conversations about the impact of Coronavirus on business, and conversations about Coronavirus and 5G. **We found that conversations about Coronavirus and 5G attracted more accounts with anonymous characteristics. Levels of anonymity were at least 48% higher compared to the conversation about the impact of Coronavirus on business.**

When we looked more closely at the Coronavirus-5G discussion, we saw that anonymous accounts were driving the conversation. Over the six-week period we examined, anonymous accounts rapidly increased their volume of daily tweets. **Anonymous accounts produced over five times the volume of tweets of attributed accounts, and were over four times (42% vs 9%) more likely to be promoting 5G conspiracy theories.**

When we looked more closely at the anonymous accounts promoting 5G conspiracy theories, **we found the that anonymous accounts were tweeting far more of the most extreme conspiracy content** - such as themes from the QAnon movement, which is seen as a potential domestic terrorism threat by US authorities. [1]

We found that identity concealment was a far better indicator for 5G conspiracy promotion than automated activity. Those seeking to address the spread of false information on social media tend to focus their efforts on the activities of automated accounts or "bots". We found that automated/bot accounts made up a very small portion of those promoting 5G conspiracy theories. This may reflect effective action by the platforms. However, this has not resolved the problem of false information -
suggesting the current approach is inadequate.

Our research indicates that anonymity should be seen as a risk factor for a user engaging in the spread of harmful false information. Our recommendations therefore focus on steps which social media platforms could take to manage that risk.

Facebook and Twitter currently only offer the option of identity verification to a very restricted proportion of their users. We suggest that this should be extended to all users. We suggest that all users should be able to tell if another user has chosen to use a pseudonym or use their real name, and whether or not that name has been verified. We also propose that users should be given more choice as to whether or not unverified and anonymous users are able to interact with them or appear in their feeds.

1. In 2019, the US Federal Bureau of Investigation made an internal assessment that Q-Anon conspiracy theorists posed a domestic terrorism threat

Executive Summary

There are several limitations to this study.

Firstly, as we explain in the Methodology section, it is extremely difficult to determine the true extent of anonymity and identity concealment on Twitter. In this study we have defined as "anonymous" only those accounts where identity information is not given, or is quite obviously some form of pseudonym. This means we are likely to have failed to classify as "anonymous" users who conceal their identity by providing false, but plausible, identity information. As such, our estimates of the prevalence and impact of anonymous accounts are almost certainly under-estimates.

Secondly, we have confined our analysis to users who identified as located in the UK. This gave the study focus, and ensured it can be considered relevant for UK policymakers. It also made the numbers of users more manageable for in-depth analysis. However, other users, either not based in the UK or not identifiable from their profiles as based in the UK, will also have been visible to UK users. We therefore have not examined all the users who have had a potential impact on the UK twitter conversation, whether anonymous or named, or the balance between the impact of UK vs overseas users on the UK conversation.

Finally, this study limits itself to Twitter, and doesn't examine other social media platforms which play a significant role in the spread of false information, especially Facebook. This is because, notwithstanding its limitations, Twitter is easier to analyse as its data is more open and available for independent analysis.

Specifically with regard to anonymity, Twitter explicitly permits identity concealment, whereas Facebook's "real name policy" means identity concealment is harder to detect. We have no reason to believe that identity concealment is not prevalent on Facebook - the platform's lack of robust identity verification hugely undermines the "real name policy", and recent studies have identified large numbers of suspicious looking accounts. Twitter has been singled out here because, of the larger platforms, it is easier to study - not because it is necessarily the worst offender.

There has been substantial resistance from the major platforms to the idea that anonymity and identity concealment are relevant factors when considering the spread of false information - let alone being a risk factor which they should actively manage. This resistance was criticised on 4 May 2020 by Julian Knight MP, chair of the DCMS Select Committee, who expressed frustration at repeated failures to offer factual answers to questions about the role of anonymous accounts.

We believe that these limitations do not detract from our conclusion that social media companies have a case to answer. **Our analysis of this snapshot of UK Twitter conversations suggests a clear positive correlation between anonymity and the spread of dangerous false information during a pandemic.** If Twitter wishes to challenge our findings, we would invite them to provide their own data on the prevalence and behaviour of accounts with concealed identities.

2. Davis, T., Livingston, S. and Hindman, M., (2019) "Suspicious Election Campaign Activity on Facebook: How a Large Networks of Suspicious Accounts Promoted Alternative Für Deutschland in the 2019 EU Parliamentary Elections"; School of Media and Public Affairs - George Washington University Suspicious Election Campaign Activity on Facebook

3. For an example, see question 54; https://committees.parliament.uk/oralevidence/329/html/

Background

It is widely accepted that there is a dangerous level of false information circulating about the Coronavirus pandemic. The World Health Organisation has described it as an "infodemic". In the United Kingdom, the Secretary of State for Digital, Culture, Media, and Sport warned of the "spread of falsehoods and rumours which could cost lives". The Coronavirus pandemic has thrown into sharp relief the way false information can degrade the ability of governments to develop consensus around policy and implement a public health response.

Conspiracy theorists around the world have linked 5G technology to the Coronavirus. This has caused particular concern because as well as inspiring scepticism towards government health messages, these conspiracy theories appear to have inspired dozens of attacks on phone masts. Telecoms engineers have also been attacked.

Social media platforms have so far responded to calls for them to take a greater role in tackling the spread of false information by focusing on the activities of "inauthentic" automated/bot accounts; by adding links to guide users to reputable sources; and by removing, in a reactive manner, particularly high profile pieces of misleading content. The continued prevalence of false information shows that this is not enough.

This study aims to look beyond the removal or labelling of specific bits of content, and beyond a narrow definition of "inauthenticity", focused on automation. We explore to what extent, if any, anonymity could be said to be a factor driving the spread of false information.

By better understanding the role of anonymity in the behaviour of those who create, produce and disseminate false information, we hope to be able to recommend courses of action that better address the problem.

Facebook and Twitter are perhaps the two most important platforms in the world for the exchange of ideas. Both have been identified by researchers as having been used to spread Disinformation - where false information is purposefully spread for some sort of real-world gain - and Misinformation - where individuals believe false information and share it believing it to be correct.

For this study, we decided to focus on Twitter rather than Facebook for two reasons. Firstly, Twitter expressly allows anonymity whereas Facebook does not. This means it is easier to recognise at least some of those using anonymity on Twitter, because there is no requirement from the platform for them to conceal their anonymity. There is much evidence that Facebook's so-called "real name policy" is widely flouted, but this policy means all anonymity on Facebook is concealed anonymity in the form of pseudonymity, making it much harder to identify at scale. Secondly, Twitter allows far greater open access to data. Twitter was therefore the easier platform upon which to conduct an initial assessment of the role of anonymity.

This should not be taken to imply that Twitter necessarily has the greater problem. It could well be that Facebook's under-enforced real name policy means an even greater prevalence of concealed anonymity.

Methodology

This research project was conducted in four stages, which are outlined here and described in greater detail in the following pages:

- **Stage 1:** Collecting the 16 million tweets that featured the hashtag #Coronavirus between March 1, 2020 and April 18, 2020, followed by sifting to find those based in the UK, and sorting into three sample sets
- **Stage 2:** Categorisation based on level of anonymity
- **Stage 3:** Investigating the most and least anonymous accounts in the 5G sample set to understand differences in the way users in each group engage with 5G conspiracy content
- **Stage 4:** Examining the activity and impact of the accounts tweeting about Coronavirus and 5G conspiracies

Stage 1: Collecting Tweets and developing three sample sets

The data team collected all tweets that included the hashtag "#Coronavirus" from March 1, 2020 to April 18, 2020. This resulted in 16 million tweets that were then sifted to include only those that originated from UK-identifying accounts. The team identified the geographical location of the accounts based on keywords used in the location section of their bios, and/or the presence of a geotag. As a result of this interrogation of the data set, we found that a total of 284,839 Twitter accounts in the UK had tweeted the hashtag #Coronavirus. However, it should be noted that the actual number of UK accounts tweeting on the subject is likely to be higher as some users without any geographical identifiers will be UK-based.

The data set of 284,839 UK Twitter accounts was used to draw out three smaller groups for further interrogation and comparison. The first group was of 2,560 accounts picked at random from amongst the data set to form a "UK #Coronavirus Twitter Accounts" data sample with an acceptable margin of error (about 1 percent). The second group was of 7,350 accounts that had mentioned #Coronavirus and #Business, which formed a "UK #Coronavirus and #Business Twitter Accounts" data sample. The third group was of 1,228 accounts that had mentioned #Coronavirus and "5G", which formed a "UK #Coronavirus and 5G Twitter Accounts" data sample.

The number of accounts identified which tweeted about #Coronavirus and 5G was a small proportion of the UK accounts. This is because linking 5G to Coronavirus was an extremely fringe position. We will not have captured all tweets on the subject visible to a UK Twitter user, as our sample only contains Twitter accounts identified as from the UK, and additionally some tweets on the subject may have used alternative hashtags. However, the proportions would be likely to remain similar. This reflects the fact that the 5G conversation was an example of a relatively small number of users achieving disproportionate impact through their energetic promotion of a fringe topic. The sample size of 1,228 accounts was large enough to enable statistically significant analysis.

Methodology

Stage 2: Categorisation on the basis of anonymity

The data team then categorised the accounts in the samples on the basis of how anonymous they appeared to be. The approach focused on four identity markers; the name used in the handle (e.g. @JohnDoe), the first and second name used in the username (i.e. "John" and "Doe"), and the photo included in the profile.

Each account was given a score based on the presence or absence of the four markers. Each missing anonymity marker earned the account 25% added to its anonymity score. As an example, the lead author of this study would have received a score of 25% since his profile includes a faceshot and his username contains his first name and surname ("Amil Khan"), but his handle is @Londonstani.

The scores received by each account were then aggregated in each sample set. This resulted in each receiving a percentage score that indicates the average level of anonymity of the accounts in the group. A score of 100% would signify that all the accounts in the sample lacked an identifying photo or a real (or apparently real) name in the handle and username. Alternatively, a score of 0% would mean that each account had a photo and an identifiable name in the username and handle.

It was clear that a significant number of accounts may have been using pseudonyms. In some obvious cases, we were able to apply our judgement that a superficially real-sounding name did not refer to a real person. However, in other cases where the names used were more generic, it was harder to tell whether a claimed identity was genuine or not. For the purposes of this study, we only categorised very obvious pseudonymity as concealing identity, such as the tribute account @robspark1 (see Fig. 4). We have therefore almost certainly under-estimated the number of accounts using a pseudonym.

This study seeks to examine the use of anonymity rather than uncover the precise extent of anonymity. As such, we believe this conservative approach to classifying pseudonyms was the most straightforward way of avoiding wrongly classifying attributed accounts as anonymous. This is a limitation of the study. However, it is one which is intrinsic to any examination of behaviour on social media platforms given the absence of any form of robust identity verification.

A future version of this study could seek to uncover more pseudonymous accounts, for example by running all profile photos through machine learning image analysis software to see if they have been appropriated from elsewhere on the internet. A 2019 study of German-language Facebook was able to take advantage of Germany's naming laws to identify a set of accounts with names which were very unlikely to be real – an option not available to a UK-focused project. [4] There simply is not currently a fail-safe method of distinguishing between pseudonymous and real named accounts. For as long as platforms resist offering genuine identification to more of their users, both researchers and genuine users will continue to struggle to determine the exact extent to which other users are and aren't using their real name

4. Davis, T., Livingston, S. and Hindman, M., (2019) "Suspicious Election Campaign Activity on Facebook: How a Large Networks of Suspicious Accounts Promoted Alternative Für Deutschland in the 2019 EU Parliamentary Elections"; School of Media and Public Affairs - George Washington University Suspicious Election Campaign Activity on Facebook

Methodology

Fig 1. Fully Attributed Account

As shown on the left, the Twitter account belonging to former MP George Galloway is fully attributed. The account shows Galloway's face in the profile picture, and the username (in white text below the profile photo). Also, the handle (the grey element starting with "@") includes his first and second names. The rating system developed for this study would score this account as 0% anonymous.

Fig 2. Semi Attributed Account

Although the account on the left contains a face shot in its profile photo and a first name, the lack of a second name and the generic nature of the handle earn it a rating of 50%. Although at first glance, it looks as it the owner of this account is easily identifiable, with just a common first name and a faceshot, it would be quite difficult to trace who is behind it.

Fig 3. Fully Anonymous Account

A fully anonymous account, as shown on the left, does not display a face in the profile photo, the username does not contain the user's first or second name, and the handle also does not contain a recognisable, real name. This account, having no identity markers present, scores 100% under the rating system developed for this study.

Methodology

Fig 4. Pseudonymous account Rob Spark

@robspark1 is a fan account for the actor Robert Pattinson. The username "Rob Spark" is likely a form of tribute rather than the user's name. As such the account was judged as pseudonymous for the purpose of this study and was assigned a score of 100% as each identity marker is likely inauthentic

Stage 3: Investigating the most and least anonymous accounts

Stage 3 used qualitative analysis to understand how anonymous accounts behave in comparison to non-anonymous (i.e. attributed) accounts. The analytical team examined each of the accounts within the "UK #Coronavirus and 5G Twitter Accounts" data sample that have all four identity markers present, and those that had none. The accounts with all markers present - i.e. the fully attributed segment - comprised of 223 accounts. Those with no identity markers present - ie the fully anonymous segment - comprised of 194 accounts.

The analytical team examined each account in both segments at length and sought to establish a basic analytical profile that included identity (e.g. professional, public personality, political tool etc), motivation (e.g. self promotion, political proselytising etc), the main themes the account focuses on and, crucially, its position on 5G conspiracies.

The team also ran each account through bot detection software. This allowed us to understand how prevalent 5G conspiracy promotion was within each segment and the kind of users who were engaging in it.

Stage 4: Examining the activity and impact of the accounts tweeting

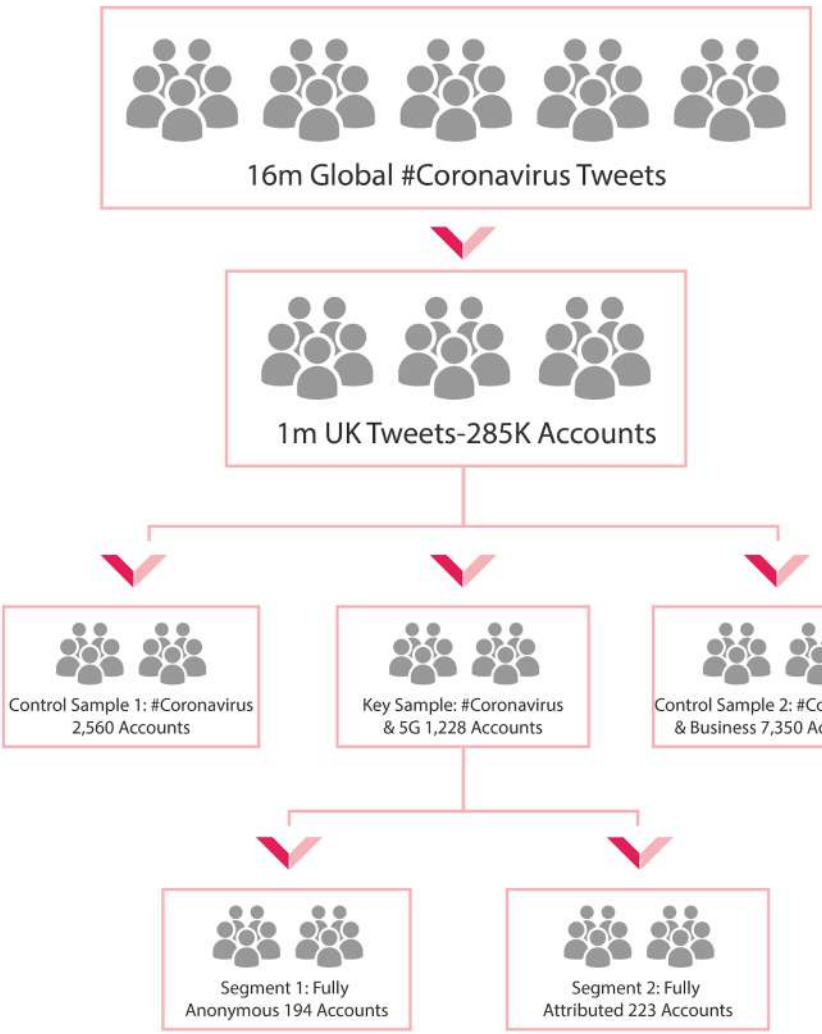
The analysis undertaken in the third stage provided us with insight into which accounts were turning up most commonly in 5G conspiracy conversations and also what specific content they were promoting.

We took these accounts and information sources and ran them through software that allowed us to map their spread across Twitter. Comparing them to the content most commonly shared by attributed accounts, allowed us to compare the reach of anonymous and attributed accounts.

Data Overview

The diagram below outlines how the research teams sifted and categorised the raw data in order to understand and compare how anonymity featured in different Coronavirus-related conversations. Samples 1-3 were drawn from a data set of 285,000 UK accounts. Sample 2, which included accounts that featured the hashtag #Coronavirus and “5G”, was subdivided further into “anonymous” and “attributed” groups.

Fig 5. Data Sift



Examining Anonymity

Anonymity Across Conversations

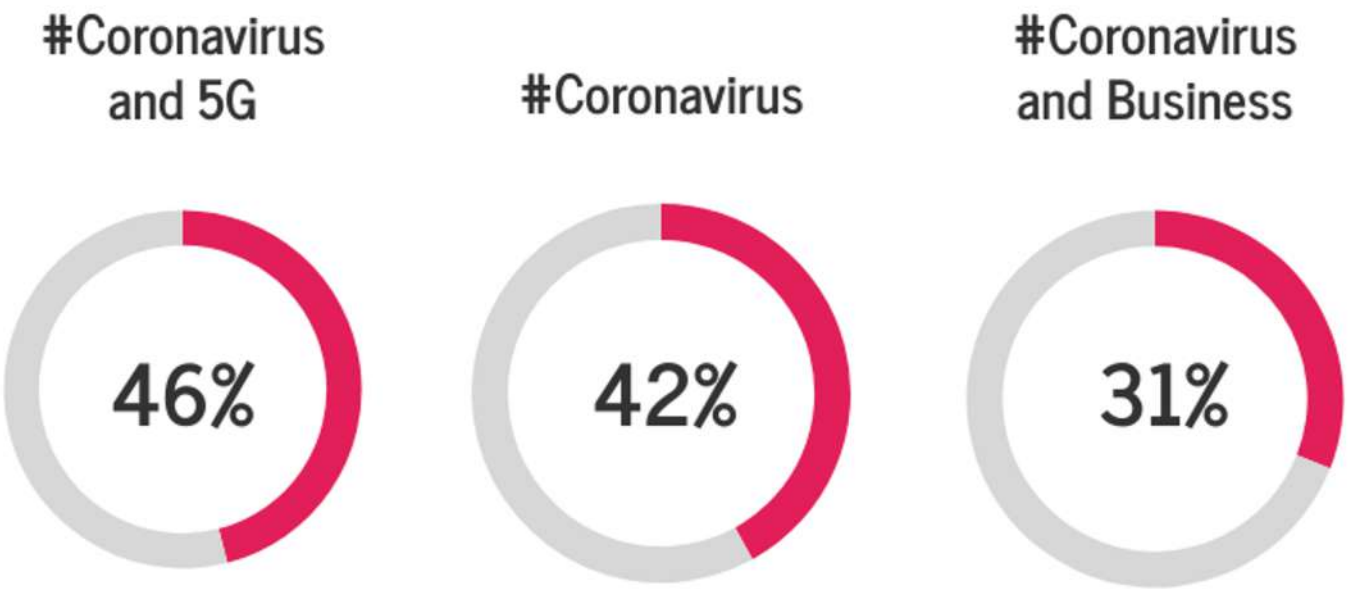


Fig. 6. Anonymity levels in different Coronavirus-related conversations

The graphs above show the level of overall anonymity within each of the conversations listed; #Coronavirus-5G, #Coronavirus and #Coronavirus-Business. The percentage related to the average level of anonymity of the accounts in the conversation. As such, it is clear that #Coronavirus-Business had the lowest level of anonymity. Whereas #Coronavirus-5G had the highest.

Examining Anonymity

Behaviour of Anonymous Users

The data team also looked at the features and activity of the accounts ranked as highly anonymous in the main sample set (UK #Coronavirus)

Fig 7. Location of anonymous users tweeting #Coronavirus and "5G"

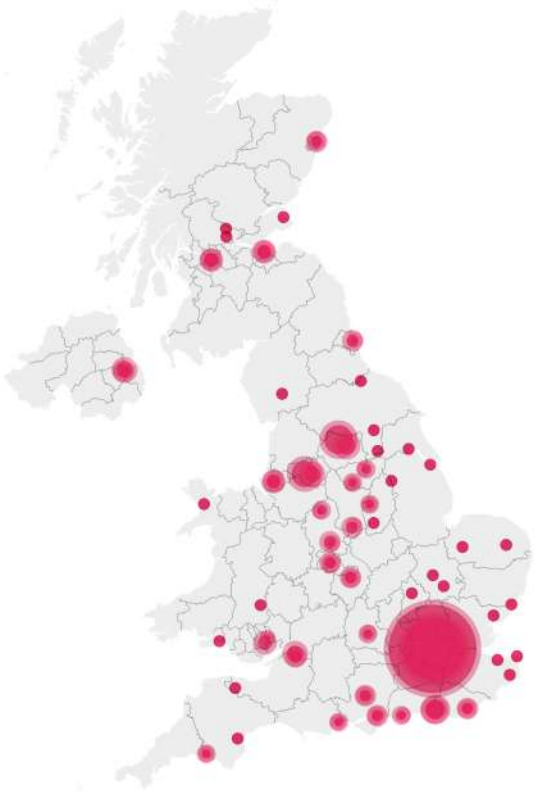
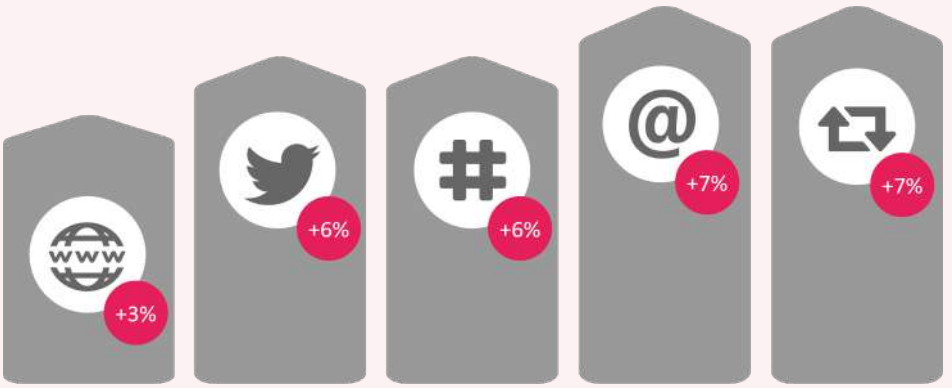


Fig 8. Top 10 hashtags used by anonymous accounts

#nicolasturgeon 1	#auspol 6
#panickbuying 2	#maga 7
#worldhealthorganization 3	#toiletpaper 8
#ccpvirus 4	#torygenocide 9
#world 5	#ukgoverment 10

Fig 9. Activities of anonymous accounts

A greater proportion of the accounts sharing urls, tweeting, hashtagging, tagging other users and retweeting were anonymous. This reflects the fact that they were more likely to be actively promoting the conspiracy theory, as opposed to merely discussing it, and meant they had disproportionate reach and impact within the conversation



Examining Anonymity

Anonymous vs Attributed Users

To understand the scale and impact of the anonymous accounts required a deep dive into the accounts themselves. The analytical team conducted a qualitative analysis within the Coronavirus-5G sample. The team looked at those accounts where all four identity markers are present, and those where none are present. These two segments allowed us to compare the behaviour of accounts where the user's identity is fully known and those where is is completely obscured.

Fig 10. Conversation themes amongst anonymous accounts

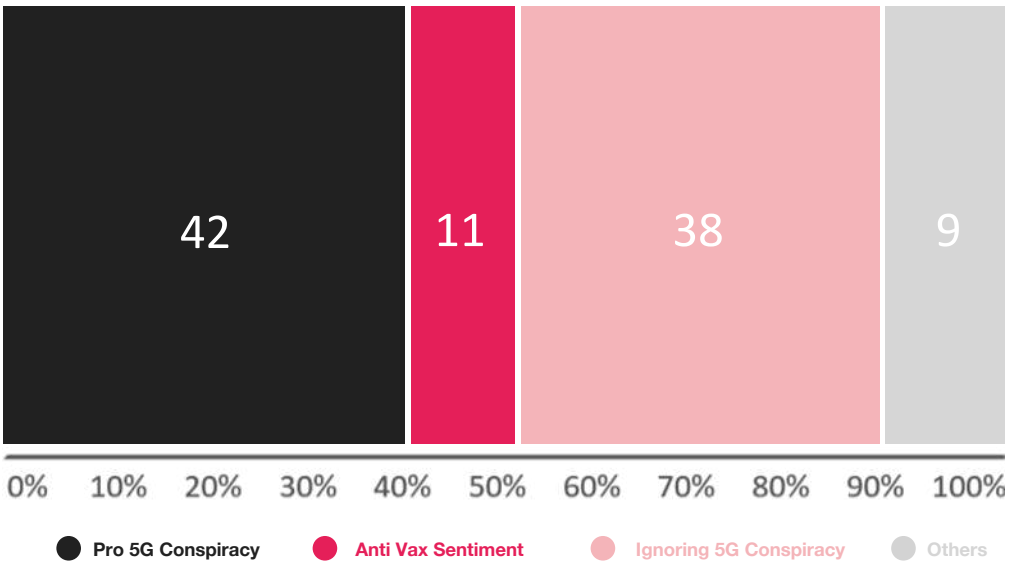
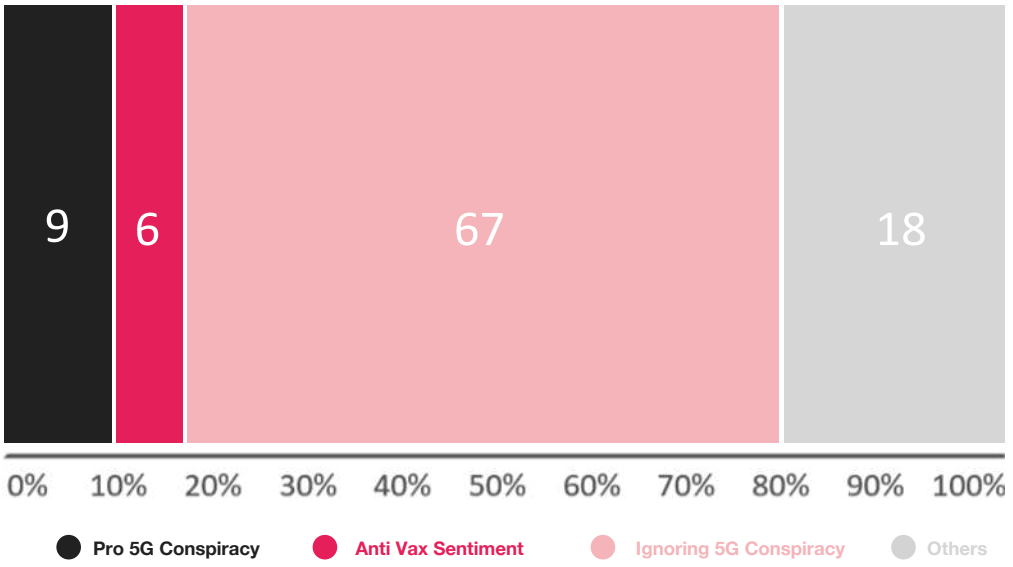


Fig 11. Conversation themes amongst attributed accounts



Comparing Behaviours

Fig 12. Comparing proportion of anonymous and attributed users promoting 5G conspiracies

The graphs make plain just how much more anonymous accounts are active in promoting 5G conspiracy theories. When accounts are anonymous there are over four times as many (42% vs 9%) promoting 5G conspiracy theories.

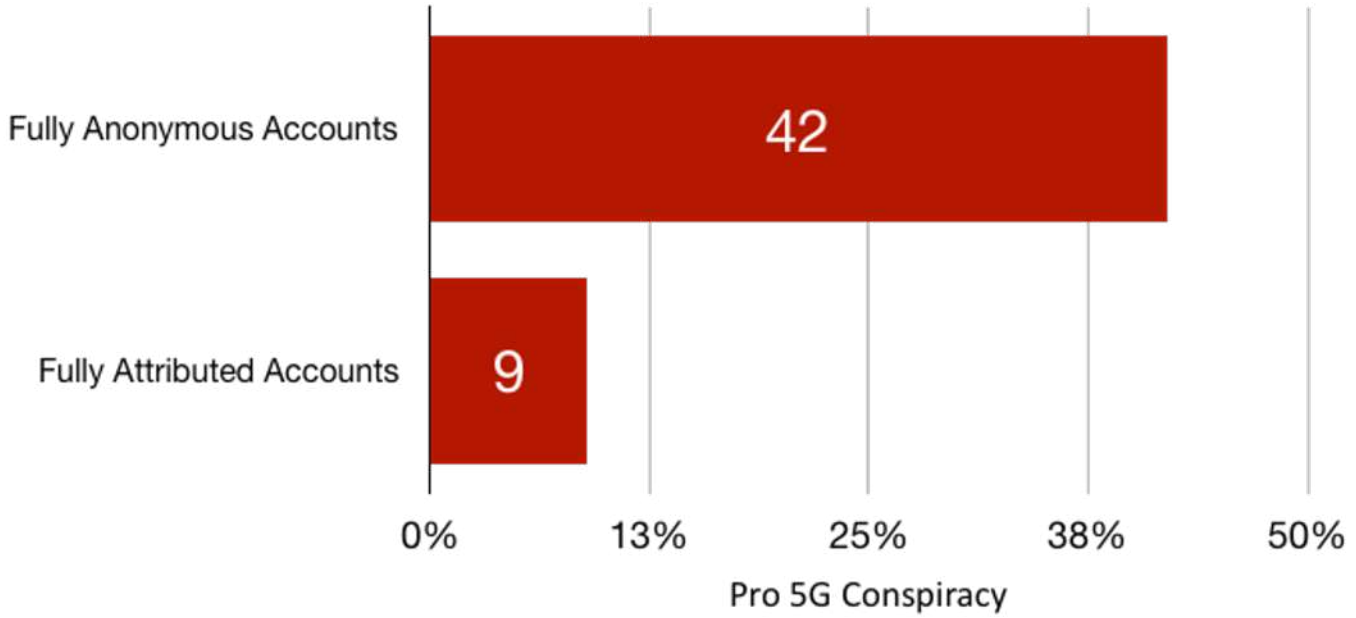
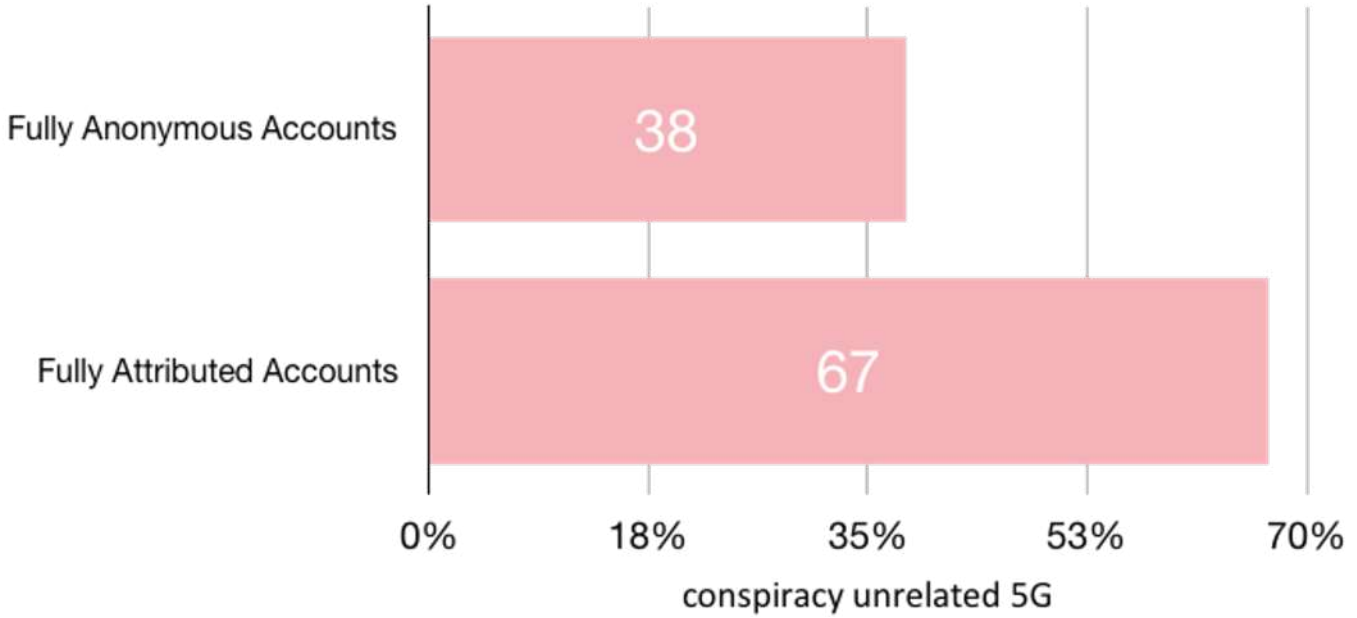


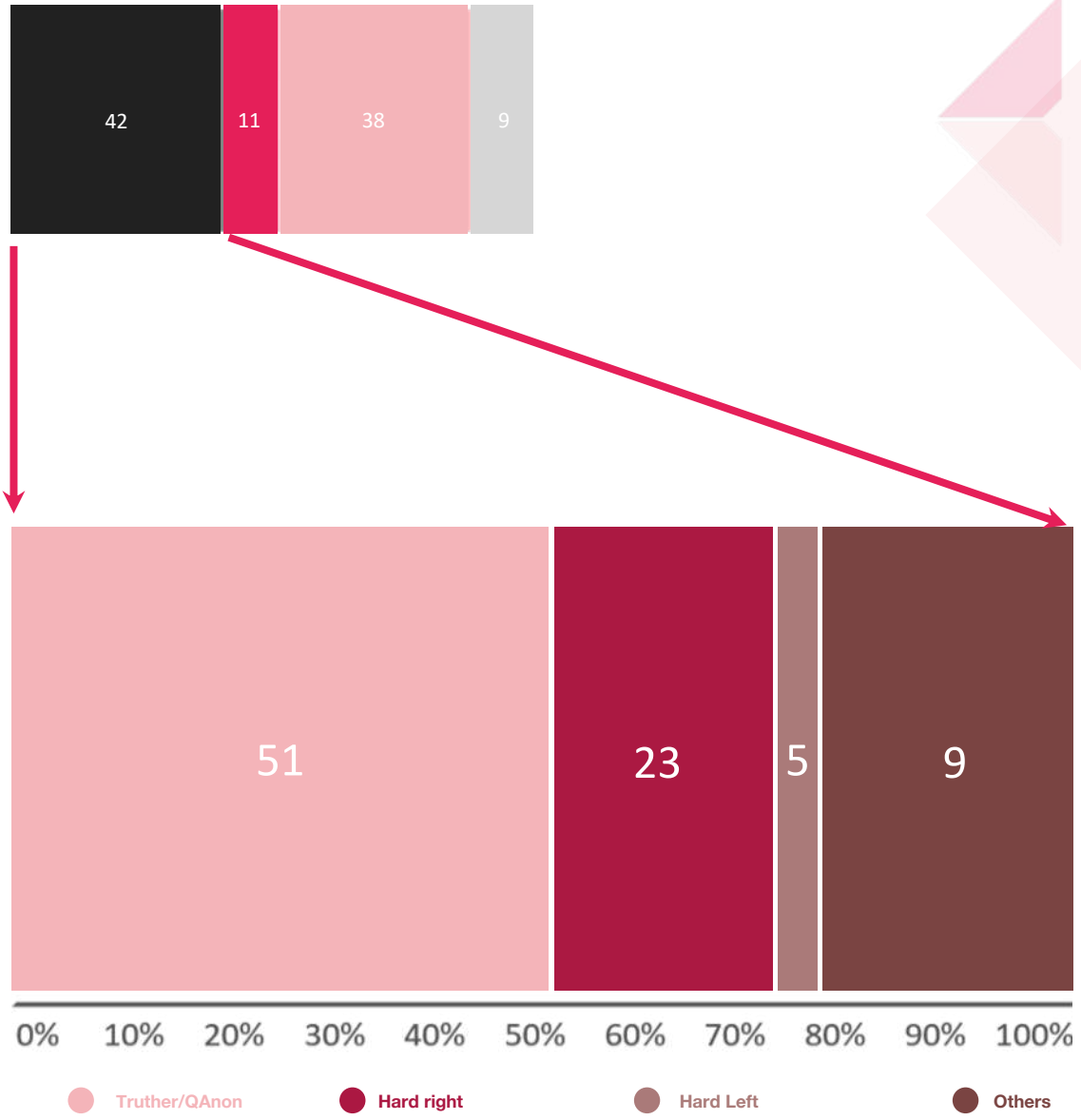
Fig 13. Comparing proportion of anonymous and attributed users ignoring 5G conspiracies

Additionally, we can see that in attributed conversations, the vast majority of people mention the term “5G” in a context that is unconnected to conspiracy theories. In anonymous conversations it seems 38% of users are not focused on 5G conspiracies in their interactions. When we look at attributed accounts, that number jumps to 67%



Anonymous Accounts Deep Dive

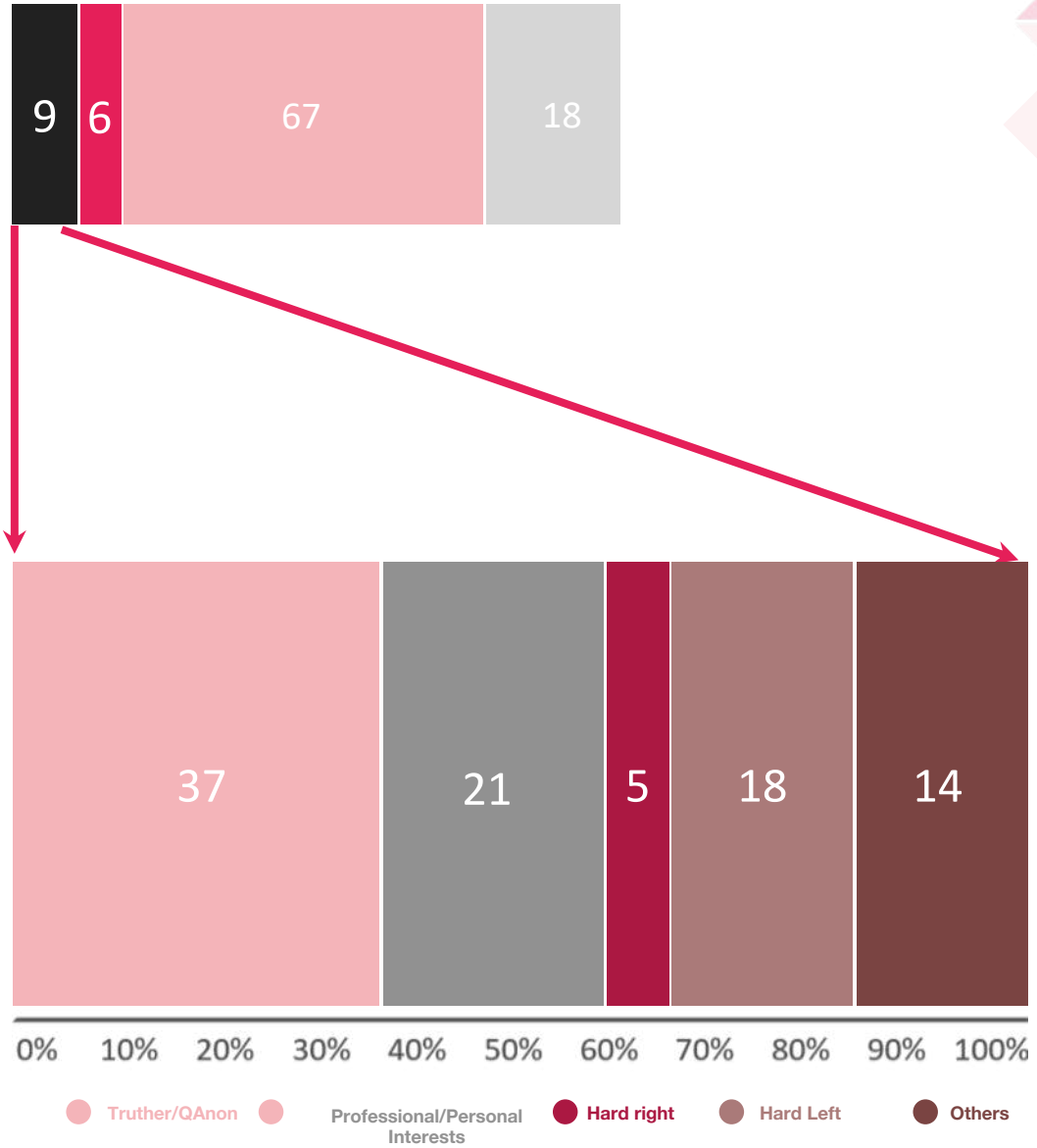
Fig 14. Examining anonymous 5G conspiracy promoting accounts by worldview



- **Truther/QAnon =** Most prominent theme of 42 of 82 accounts (51%) and second most prominent theme of 4 of 82 accounts (5%)
- **Hard Right =** Most prominent theme of 19 of 81 accounts (23%)
- **Hard Left =** 4 of 82 accounts (5%)

Attributed Accounts Deep Dive

Fig 15. Examining attributed 5G conspiracy promoting accounts by worldview



- **Truther/QAnon =** Most prominent theme of 7 out of 19 accounts (37%)
- **Personal/local outreach =** 32 of 149 accounts (21%)
- **Hard Right =** 1 of 22 accounts (5%)
- **Hard Left =** 4 of 22 accounts (18%)

Examining Anonymity

Pushing Conspiracy

We used social listening software Murmurate to examine the behaviour of the anonymous accounts mentioning 5G as a whole and compare it to the behaviour of fully attributed accounts. The results showed a marked difference in the volume of activity and the kind of content that was being promoted

Number of Tweets published by all the accounts in each group on a daily basis from March 1 to April 18, 2020:

Anonymous

March 1: 597 Tweets
April 18: 2,327 Tweets

Attributed

March 1: 87 Tweets
April 18: 449 Tweets

Most common hashtags used by the accounts in each group from March 1 to April 18, 2020 (apart from #Coronavirus and spam related):

Anonymous

Attributed

Anonymous accounts tweeted and retweeted much more more often than attributed accounts over the 6-week period examined in this study. Anonymous accounts' tweeting focused far more intensely on QAnon-related content. Attributed accounts did also mention QAnon, but to a lesser degree while also turning their attention to issues popular with more left wing sentiment in the UK, such as conflict between Israel and Palestinians.

Examining Anonymity

Conspiracy Content

Accounts Retweeted Most Often

Anonymous Accounts:

Attributed Accounts:

Anonymous and attributed accounts engaging in conversations about Coronavirus and 5G were retweeting different types of accounts. Anonymous accounts were promoting (by retweeting) the content of accounts that were 100% anonymous - where all four identity markers were missing. Attributed accounts were retweeting accounts that were less anonymous. A closer look at the accounts being retweeted shows that a higher degree of anonymity correlates to more extreme content. @WarPlanPurple and @DalesDweller4 depicted above both promote extreme conspiracy theories - such as anti-vaxing and QAnon tropes. The accounts favoured by attributed accounts, such as @inevitable_ET and @David Furness, focus on demonstrating support for right wing politicians and policies, such as Brexit.

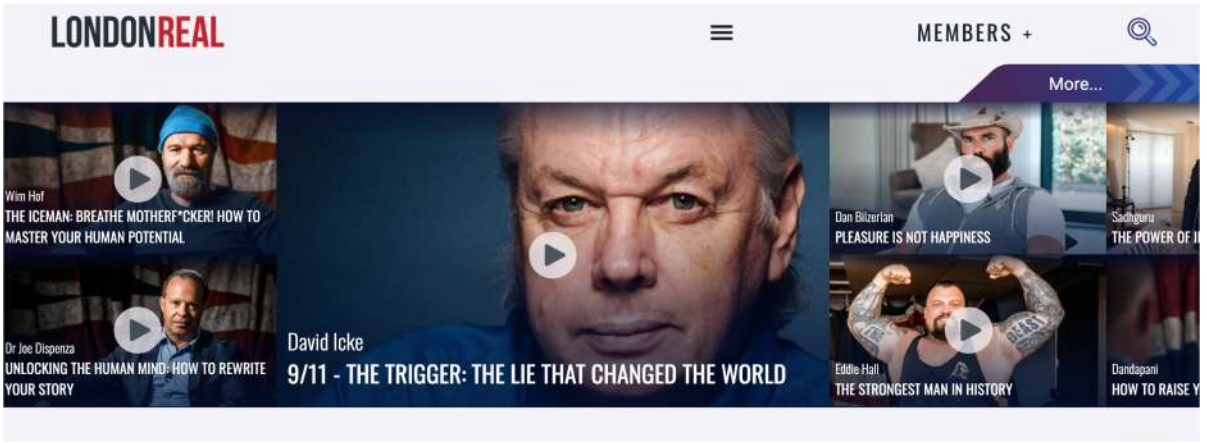
Examining Anonymity

Content

Content Shared and Promoted Most Often

Anonymous Accounts:

Attributed Accounts:



Anonymous accounts tended to promote content designed to appear like professional news articles and video packages, with the focus on "high conspiracy" - such as accusations that Bill Gates is part of a global plot to control minds, and airliners spray mind-control drugs as they fly overhead. UK users of attributed accounts however are heavily drawn to content on an online channel known as London Real - which runs accounts on Twitter, Facebook and YouTube, alongside other major social media platforms. London Real TV is characterised by its high production values and charismatic anchor, Brian Rose. The outlet has an editorial focus on self improvement through insights on business, health and education. The conspiracy-theory content is usually delivered in the form of interviews with prominent conspiracists like David Icke. Viewers are urged to hear the arguments of Icke and similar figures and then judge for themselves. London Real TV's popularity with attributed accounts suggests they are attracted by high production values and interviews with known (albeit extremely disreputable) individuals. Anonymous accounts were much more willing to share content from a wider range of unattributed sources.

Observations

A key observation from our research is that anonymous accounts appear disproportionately in emotionally charged conversations, are far more active in terms of how often they comment and retweet and are much more likely to be promoting extreme conspiracies. Not only were more anonymous accounts present in discussions like the role of 5G technology in causing illness, but those accounts shared links and retweeted others far more often than accounts that had more transparent identities. Their increased activity means individuals present in those conversations would have a higher exposure to the points of view they were advocating, which would result in the impression those views were more prevalent than they were in reality.

When we looked deeper into the specific activity of anonymous accounts in conversations that mentioned 5G technology in the context of the pandemic, we found that they were more than four times as likely to be promoting conspiracy theories than accounts where the user's identity was clearly visible. As we peeled back another layer and looked at the nature of the 5G conspiracies, we found that QAnon - a particularly dark corner of the conspiracy theory pantheon - made up more than half of the 5G conspiracies being being promoted.

The QAnon movement was linked to an armed attack on a pizza restaurant in Washington DC in 2016 after the movement's followers claimed online that it was the headquarter of a child sex abuse ring run by Hillary Clinton. In 2019, the FBI in an internal memo identified QAnon as likely to spawn further violent acts in the future.

Overall, our research showed that while greater transparency did not remove conspiracy theories from the conversation, it did result in less focus on conspiracy-promoting content and less extreme conspiracies. However, it also became clear that when identities are more transparent, the conversations focused a lot more on professional and personal interests, with users seeking to discuss their ideas and activities and make connections with like-minded individuals. Anonymous accounts tended to prioritise prolysterising their views over engaging others.

We also observed that very few accounts - three among 222 attributed accounts and 10 among 194 anonymous accounts - were identified as "inauthentic" by bot detection software. This came as a surprise considering that the research team came across a number of accounts acting in ways that suggested they were not merely accounts of regular private individuals as claimed. Several accounts, for example, focused on constantly attacking Nigerian politicians. One claimed to be a UK account but seemed to post cut and paste messages praising Saudi Arabia's rulers on a regular basis. Another posted frequently about pets, occasionally interspersed with messages supportive of QAnon. All of these accounts alongside a significant number of others were engaging in what looked like inauthentic behaviour, but it appears they were being operated by real people with concealed identities. This last point, has implications for the way social media platforms understand "inauthentic behaviour".

Findings

1. Anonymous Twitter users are more likely to engage with, and promote, 5G Coronavirus conspiracy theories than users displaying their real name

Our research found a clear relationship between anonymity and the spread of the 5G conspiracy theories on Twitter. Anonymous users were over-represented in the UK Twitter conversation about Coronavirus and 5G, both in terms of the number of anonymous accounts present and in terms of the volume of tweets which they produced. Within that conversation, anonymous users were far more likely to be actively promoting conspiracy theories, as opposed to engaging with them critically or out of curiosity, and were more likely to be circulating links to conspiracist articles and websites. Amongst all UK Twitter users promoting variants of the 5G conspiracy conspiracy, anonymous users were more likely to be promoting the most extreme and violent variants.

Whilst this study focused on 5G and Coronavirus, we suspect that anonymous accounts are likely to play a similarly significant role in the circulation of other conspiracy theories, and potentially also in amplifying more extreme positions in discussion of other emotive topics. We did not identify anything in the data to suggest that the patterns we observed regarding anonymous and attributed accounts are exceptional to the 5G coronavirus conspiracy theory.

2. Automation/non-automation is an insufficient test of "authenticity" for the purposes of assessing the risk of harmful behaviour

The accounts we examined appeared to be, overwhelmingly, accounts that were being run by real people. The "anonymous" accounts were not for the most part "bots" - they appear to be real people hiding their real names, and apparently much more likely to circulate conspiracy theories from this position of concealment.

The very limited number of accounts identified by bot detection software during this study demonstrates the severe limitations of a focus on automation as the main indicator of whether or not an account is likely to engage in harmful behaviour. It is possible that this reflects progress by Twitter in detecting and removing harmful "bots", but if so Twitter's success in this limited area has not prevented the spread of dangerous Coronavirus conspiracy theories on their platform.

Recommendations

On the basis of our key findings, we make the following recommendations:

1. Automation/lack of automation is not on its own a sufficiently reliable indicator of potential for harmful behaviour and needs to be supplemented by other factors. It may be that the low levels of bot/automated activity which we identified within the Coronavirus-5G conversation reflects a diminution in the prevalence of such accounts thanks to succesful intervention from Twitter. However, this has not on its own led to a diminution of harmful activity. Other interventions are clearly required.

1. Social media platforms, and policy makers, should consider concealed identity, through anonymity or pseudonymity, as a risk factor for harmful behaviour such as spreading dangerous disinformation. Approaches to managing anonymity need to be developed which mitigate such risks.

1. Some actors, such as rights activists living in authoritarian states, or whistle-blowers vulnerable to losing their jobs, rely on anonymity to avoid reprisals. We therefore recommend that approached are developed which seek to prevent the misuse of anonymity, whilst retaining it as an option for those who need to protect their identity for entirely legitimate reasons

1. Platforms' current approach to identity verification should be expanded in order to offer users more choices and greater insight into the potential motivation of actors seeking to engage them. All users should be offered options of verifying their identity or confirming they are using a pseudonym. Platforms could build on their current verification offers (such as Twitter's blue tick and Facebook's verification of political advertisers) and make these available to all users.

1. It should be easily visible to all users whether or not another user has chosen to be anonymous, has confirmed that they are using a pseudonym, or has chosen to use their real name - and where a user is claiming to use their real name, whether or not that name has been verified

1. All users should have the choice as to whether they want to be contactable by accounts which are anonymous or have an unverified identity status.

1. Given their great ability to manipulate and influence a conversation, accounts that promote emotive partisan material at scale, or purport to be a "news outlet", could be subjected to more stringent identity requirements - as is currently the case for example for accounts on Facebook which wish to pay for political adverts. This requires further exploration.

Acknowledgements

This research paper was commissioned and funded by Clean Up the Internet.

Valent Projects, a digital communications company focused on social impact, designed and managed the project. The quantitative data work was carried out by digital research agency Sulis Insights.

Software used in this project included Sulis Insight's F(r)actionTM AI tool and Quaking Aspen's Mumurate, a proprietary social listening tool. The team also used Facebook's Crowd Tangle to see how content was being shared on social media.

SULIS INSIGHTS

MURMURATE

About the Author

Amil Khan is founder and director of Valent Projects, a digital communications agency for social impact. Amil has spent 20 years working on the frontlines of information, politics and conflict. He started his career as a foreign correspondent and investigative journalist for Reuters news agency and the BBC, reporting from war zones in the Middle East and Africa. He then worked for the UK government as a senior strategic communications expert with a special focus on conflict. He studied the rise of Disinformation as a weapon of war and helped devise strategies for countering it. Since Valent's founding in late 2019, Amil, a former Chatham House fellow, has led teams undertaking complex research projects and implementing digital communication strategies for those seeking to create positive change.

Appendix 1: Definitions

Different definitions exist for several of the terms used extensively throughout this report. Below, we outline the definitions we used for the purpose of conducting this study:

Anonymity:

Anonymity exists in different forms and to different degrees. It can be both a subjective experience ("feeling anonymous") and a relationship ("anonymous to X"). In this report we have taken anonymity to be the absence of a clear identity to most other users of the platform. It would be reasonable to assume that such users may also "feel" anonymous, but we have not specifically looked for this. We have not considered within this study the extent to which users may be identifiable by parties other than ordinary users of the platform, e.g. law enforcement, or the platform itself.

Pseudonymity:

Pseudonymity refers to the use of a false name, giving a user a disguised identity. For the purposes of this report we define as "pseudonymous" a user who is using a name which gives the impression of being real, but is not their actual name, for example a real person called John Doe tweeting under the name of "John Smith". Given the lack of verification of identity for the vast majority of Twitter users not covered by the "blue tick", it is often very hard to tell if a user is pseudonymous or using their actual name.

Automated/Bot Accounts:

Although automated accounts are not inevitably problematic (for example a level of automation may be used by brands for legitimate advertising), social media platforms usually draw the line at collections of automated accounts acting in coordination. Such techniques, sometimes also called "bot networks", are often referred to as "coordinated inauthentic behaviour", which social media companies deem to be in violation of their terms and conditions and may lead to account removal. We avoid referring to these accounts as "inauthentic" because there are many other forms of inauthenticity (e.g. inauthenticity of identity) which are not necessarily associated with automation. "Automated/Bot accounts" is a more precise term.

Misinformation and Disinformation

In this report, we lean on the framework for defining false information in the public domain developed by First Draft News. The information integrity non-profit sees Misinformation and Disinformation as part of a wider "information disorder":

Misinformation: First Draft describes Misinformation as false information that is shared by individuals who do not realise it is false. Their actions, First Draft adds, are often driven by psycho-social factors related to the desire to demonstrate their identity online

Disinformation: Disinformation is false information designed to cause harm. First Draft notes three motivating forces; financial gain, political gain, or the desire to cause disruption for its own sake

Appendix 1: Definitions

Conspiracy Theory:

In his book *A Culture of Conspiracy* Michael Barkun defines a conspiracy theory as the belief that a group of individuals are acting in secrecy to bring about some sort of malevolent outcome. In the recent report *Trust No One: Understanding the Drivers of Conspiracy Theory Belief*, the anti-discrimination charity Hope Not Hate makes the point that conspiracy theories spread in times of crisis and are difficult to disprove because proponents take evidence to the contrary as attempts to silence them and therefore "proof" of their accuracy. This gives them, what Harvard scholar Cass Sunstein calls, a "self-sealing quality".

Truther:

"Truthers" originally referred to conspiracy theorists who believe the US government was behind the 9/11 attacks rather than the Al Qaeda networks of Osama bin Laden. However, we have noticed that those who believe the facts around other historic events - such as the moon landings or assassination of John F Kennedy - were also covered up by the US government also refer to themselves as "Truthers". We noticed that conspiracy theorists who focus on one event - such as the 9/11 attacks - see conspiracy theorists who focus on a different event as fellow travellers and are quick to accept that the episodes of interest to them are part of a wider narrative. As such we have grouped "Truthers" together under one label, and linked them to QAnon believers due to their common belief and general Hard Right leanings.

QAnon:

Travis View, a researcher who has studied the QAnon phenomenon, describes it as a movement built around the conspiracy theory that the world is run by a cabal of satan-worshipping paedophiles who control world politics and media. The movement - which believes in shadowy evil forces, a prophet-like figure known as "Q", and a messiah (often Donald Trump) who will usher in utopia - has been described as cult-like. QAnon positions tend to track closely to the American Hard Right and its resemblance to fascism has been noted. The movement has been linked to violent activities, most famously the 2016 attack on a Washington DC pizza restaurant by a heavily armed man who said he had intended to save imprisoned children from a sex ring run by Hillary Clinton. The incident is popularly referred to as "Pizzagate".