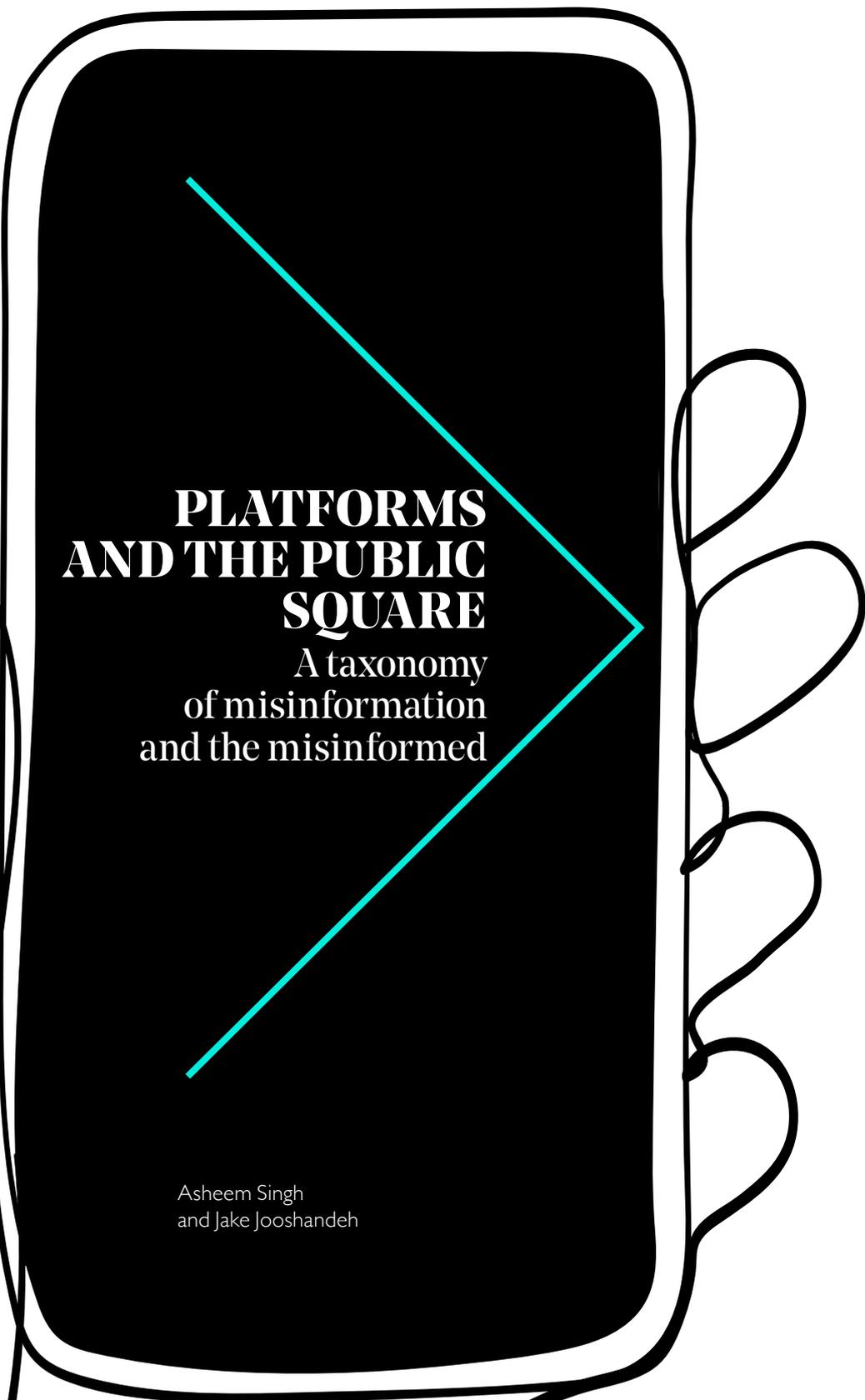


In partnership



+

A stylized graphic of a smartphone held in a hand. The phone is black with a white outline. A cyan diagonal line runs from the top-left corner to the bottom-right corner of the screen. The hand is drawn with simple black outlines.

PLATFORMS AND THE PUBLIC SQUARE

A taxonomy
of misinformation
and the misinformed

Asheem Singh
and Jake Jooshandeh

NOVEMBER 2021

Acknowledgements

The authors would like to thank the team at BT who made this work possible; the team at DCMS for their partnership and support; this project's advisory board, named in the appendix to this report, who contributed knowledge and support throughout; to Maeve Walsh and Angela Saini who were especially helpful in lending shape to the scope of our enquiry and our colleagues at the RSA, without whose support this work would also not be possible. In particular, we owe special thanks to Will Grimond and Eleanor Toner for their sterling work as researchers to this report, to Amanda Ibbett, and to Benny Souto for her design work.

All mistakes contained herein are our own.

Contents

	Page no.
i. About the RSA	4
ii. Executive summary	6
1. Moving against misinformation	18
2. A world responds to the challenge	23
3. Five tribes of information consumer	36
4. The psychological dimension: interviews on the mechanisms of misinformation	51
5. Combatting online harm: Recommendations	58

REALISING



CHANGE

We are the RSA. The royal society
for arts, manufactures and commerce.
We unite people and ideas to resolve
the challenges of our time.

The RSA (royal society for the encouragement of arts, manufactures and commerce) believes in a world where everyone is able to participate in creating a better future. Through our ideas, research and a 30,000 strong Fellowship we are a global community of proactive problem solvers, sharing powerful ideas, carrying out cutting-edge research and building networks and opportunities for people to collaborate, influence and demonstrate practical solutions to realise change.

We define our ambitions as:

Our vision

A world where everyone is able to participate in creating a better future.

Our purpose

Uniting people and ideas to resolve the challenges of our time.

We are

A global community of proactive problem solvers.

About our partner

BT Group is the UK's leading telecommunications and network provider and a leading provider of global communications services and solutions. Our principal activities in the UK include the provision of fixed voice, mobile, broadband and TV (including Sport) and a range of products and services over converged fixed and mobile networks to consumer, business and public sector customers. We believe the internet has been overwhelmingly positive and empowering, connecting people and information they would not have had access to before. However, we recognise that trust is under threat from a range of potential online harms which

includes disinformation and misinformation. BT continues to work to make the internet a safer place while respecting personal freedoms, by offering free technology tools, supporting online safety education and awareness, and working in partnership with organisations such as the RSA to investigate and propose remedies to tackle misinformation and disinformation.



SUMMARY

The scale of the challenge

Welcome to the misinformation age. For all the facts and analysis, we have at our fingertips, just a click or swipe away, it is misinformation which can spread quickest, catch most attention, and cause most damage, particularly in times of uncertainty.

- In the UK 49 percent of adults consume news via social media. Of those that do, 71 percent use Facebook and 46 percent use Twitter.¹ In 2016 in the US 62 percent of adults name social media as a news source.²
- On social media platforms like Facebook or Twitter, false stories have a greater reach and likelihood of virality than true stories. Falsehoods are 70 percent more likely to get retweeted than truths.³
- These falsehoods reach 1,500 people on average six times quicker than accurate news stories.⁴
- Since the start of 2020 the 'false content producers' industry online, defined as "sites that NewsGuard determined repeatedly published content that is probably false", has grown by 102 percent.⁵
- Outputs from fact-checking organisations saw an increase of 900 percent between January and March 2020.⁶
- Oxford University reported that 88 percent of the misinformation they analysed was found on social media but was particularly acute on privately encrypted chat channels such as WhatsApp.⁷
- Further evidence from Oxford University and Reuters contends that politicians, celebrities, and influencers create 20 percent of coronavirus and 5G conspiracy theory claims content but are responsible for 69 percent of social media engagement.⁸

The morass of misinformation has real consequences for society.

- False or mistimed rumours directly affected our generational fight against Coronavirus: a CNN report on a lockdown in Northern Lombardy, hours before the Italian government made an official announcement, is thought to have been a possible factor in a surge of cases in Southern Italy as residents fled south.⁹
- On 6 January 2021, there was an insurrection on the Capitol in Washington DC by a mix of Trump supporting QAnon and 'stop the steal' conspiracists. The rioters believed that mass electoral fraud had taken place and there would be a form of global reckoning for deep state actors.¹⁰

- 1 Ofcom (2021) News Consumption in the UK: 2021. Ofcom [PPT] Available at www.ofcom.org.uk/_data/assets/powerpoint/doc/0026/222479/news-consumption-in-the-uk-2021-report.pptx [Accessed 19 August 2021].
- 2 Gottfried, J and Shearer, E (2016) News Use Across Social Media Platforms 2016. [online] Washington D.C.: Pew Research Center. Available at: www.pewresearch.org/journalism/2016/05/26/news-use-across-social-media-platforms-2016/ [Accessed 3 September 2021].
- 3 MIT Sloan School (2018) Study: False news spreads faster than the truth [online] Available at: mitsloan.mit.edu/ideas-made-to-matter/study-false-news-spreads-faster-truth [Accessed 3 September 2021].
- 4 Ibid.
- 5 Kornbluh, K, Goldstein, A and Weiner, E, 2020. New Study By Digital New Deal Finds Engagement With Deceptive Outlets Higher On Facebook Today Than Run-Up To 2016 Election. [online] The German Marshall Fund of the United States. Available at: www.gmfus.org/news/new-study-digital-new-deal-finds-engagement-deceptive-outlets-higher-facebook-today-run-2016 [Accessed 22 October 2020].
- 6 Brennen, JS, Simon, FM, Howard, PN and Nielsen, RK (2020) Types, Sources, and Claims of COVID-19 Misinformation [online] Available at: reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation.
- 7 Ibid.
- 8 Ibid.
- 9 Cinelli, M et al. (2020). The covid-19 social media infodemic [online]. Available at: arxiv.org/pdf/2003.05004.pdf.
- 10 Porter, T (2021) QAnon supporters believed marching on the Capitol could trigger 'The Storm,' an event where they hope Trump's foes will be punished in mass executions. [online] Business Insider. Available at: www.businessinsider.com/qanon-trump-capitol-attack-belief-precursor-the-storm-2021-1?op=I&r=US&IR=T [Accessed 31 August 2021].

The scale and effect of misinformation has been steadily growing through the last decade. Governments, civil society, and platforms themselves have already been attempting to respond to the danger. This research considered the work done so far and offered a new way of thinking about how we respond: by thinking of misinformation, not as one intractable, homogenous social problem, but by understanding who the spreaders and blockers of misinformation are, and how we best work with them to mitigate or enhance their activities.

Box 1: Definitions

Misinformation: inadvertent sharing of false or inaccurate information. Misinformation can be true but incomplete or dated, thereby rendering it inaccurate or open to interpretation.

Disinformation: 'the deliberate creation and dissemination of false and or manipulated information that is intended to deceive and mislead audiences, either for the purposes of causing harm, or for political, personal or financial gain'.¹¹

Within this report we default to the word 'misinformation' to describe false information whether deliberate or not but will clarify if we specifically mean inadvertent sharing. We use the term 'disinformation' to refer to deliberate false content.

The response so far

Civil society

Civil society includes fact-checkers, universities, researchers, advocacy groups, citizen movements, and even arms of media outlets. These groups provide critical nuance, context, and balanced opposing views to information online. They also provide vital research and knowledge on the scope and scale of misinformation online and how best to combat it. Through our research we came across a range of examples which we feel are most useful to learn from:

- **Citizen movements:** civil society is both an independent voice, and a repository of knowledge and cultural understanding. **The Elves**, to take one example, are a one of a number of independent group of citizens in **Lithuania** who united in the battle against misinformation.¹²
- **Specially convened commissions:** the Oxford Internet Institute (OII) is looking at how misinformation, propaganda, and divisive political news spreads and the impact of this on the health of the UK's 'online information ecosystem'.¹³ The Royal Society is looking at how to create a 'positive information environment'. LSE's Truth Trust and Technology

11 DCMS Committee. (2019). Misinformation and 'fake news' [online]. Available at: publications.parliament.uk/pa/cm201719/cmselect/cmcomeds/11791/1179104.htm [Accessed 08 September 2021].

12 Soon Wan Ting, C and Goh Ze Song, S. (2017). What Lies Beneath the Truth: A literature review on fake news, false information and more. National University of Singapore Institute of Policy Studies, Working Paper 31, pp 18-38.

13 For instance, see: Oxford Internet Institute (no date) Programme on Democracy and Technology [online] Available at: www.oii.ox.ac.uk/research/programme-on-democracy-and-technology [Accessed 24 August 2021].

Commission (T3), which concluded in 2020, identifies an information crisis in the UK.¹⁴ The commission identifies ‘five giants evils’ of the information crisis affecting the UK public; confusion about what information to believe, cynicism regarding even trustworthy information, apathy as distrust erodes engagement, fragmentation of citizens into ‘truth publics’, and irresponsibility on the part organisations that aren’t transparent or accountable.

- **Fact-checkers:** alongside independent fact-checkers such as **Full Fact** or **Infotagion**, there have also been contributions from the traditional media to the fact-checking mission.¹⁵ The **BBC** has assembled a permanent team, **Reality Check**, with the aim of exposing false or misleading stories, alongside Channel 4’s **FactCheck**.¹⁶ Fact-checkers most obviously provide independent analysis on what can be stated as factually true, but beyond this they also provide critical context and nuance to the debate.

Regulation

Our original polling, with the public, taken before the pandemic, found that **71 percent said they want a ‘stronger independent regulator on the quality of news’**. All over the world, the legal and regulatory framework to tackle misinformation is evolving but fundamentally is still lacking. We find at least four major non-mutually exclusive categories of government control. This list is also non-exhaustive but meant to highlight some of the most common means:

- 1 Mandate misinformation content moderation and removal.** Countries around the world have begun to put into place laws that mandate the removal of content deemed misinformation and disinformation. Ranging from the Philippines, to India, to France, such laws pose obvious dilemmas on who decides what is false and who is being silenced because of it.
- 2 Regulating *illegal* content.** Other nations, in attempting to steer from the dilemmas mentioned in (1) have instead opted to mandate the removal of purely illegal content. For instance, racist content, illegal sexual imagery, terrorist content, or fraud. This is the direction taken by Germany and is broadly in line with the UK’s draft Online Safety Bill.
- 3 Demanding firmer self-regulation by online firms.** Increasingly firms are enacting their own policies to limit the spread and damage of misinformation. Yet governments are also getting firmer in pressuring platforms to do more, voluntarily, and to simply uphold their own terms of service.
- 4 Resilience building through media and information literacy.** This can be done through education and training, but also through transparency or tools. Building resilience through a population is of course a highly desirable and liberally minded route, it is, however, by no means perfect, slower to start, and longer to reap benefits.

¹⁴ London School of Economics (2020) Tackling the information crisis: a policy framework for media system resilience [PDF] London: LSE. Available at: www.lse.ac.uk/media-and-communications/assets/documents/research/T3-Report-Tackling-the-Information-Crisis-v6.pdf [Accessed 24 August 2021].

¹⁵ Infotagion (undated) Infotagion [online] Available at: infotagion.com/ [Accessed 08 September 2021] and see: Full Fact (undated) Full Fact [online] Available at: fullfact.org [Accessed: 09 September 2021].

¹⁶ [PDF] Available at: dlssu070pg2v9i.cloudfront.net/pex/pex_carnegie2021/2021/06/05091934/draft-OSB-CUKT-response-FINAL-1.pdf

Self-regulation

Platforms are increasingly implementing their own policies in an attempt to stem the tide of misinformation. Between 2016 to 2018, there were more than 125 announcements of self-regulation measures.¹⁷

Yet platforms are often highly inconsistent and untimely in their application of their own codes of conduct, due to little oversight or repercussions. Avaaz shows that Facebook's application of their own rules around fact-checked content, ie that they must receive a flag with links to the fact-checked article, was highly inconsistent with two-thirds of content which had been fact-checked not having a label. What's more it could take up to 22 days for a flag to appear.¹⁸

Platforms can also revert to a one-size-fits-all mentality. To our knowledge, platforms do not seek to offer tailored messaging or means of overcoming misinformation in the same way they may have with sophisticated targeting techniques of users for advertising. To better overcome misinformation, platforms should better understand their users attitudes and behaviours surrounding information gathering and misinformation.

The UK's draft Online Safety Bill in focus

The UK government's draft bill on online harms has advised the inauguration of an 'advisory committee on misinformation and disinformation' by Ofcom. This committee will provide advice on how regulated services should deal with misinformation and disinformation.

Civil society has widely acknowledged the Bill as a significant step forward but there are a number of concerns around the breadth of the Bill and the definitional ambiguities. Critically, the Bill is quiet and vague on how to systematically deal with and limit the threat abundant misinformation can have on society and democracy.

The Bill does provide scope for harms to the individual to have been 'indirectly' caused, as noted by Carnegie, or for misinformation to be in scope so long as it results in 'physical or psychological' harm to an individual.¹⁹ But how and when such a harm would be deemed to have taken place is much less clear than in other provisions within the Bill.

It is our view that the exclusion of 'collective' harms from the scope of the Bill is a critical failing which must be addressed if we are to hope to improve the accuracy and nature of our information ecosystem. Misinformation is a social problem as much as function of individual misapprehension or ignorance. To understand how we mitigate it, therefore, we must think in terms of group behaviour and activity.

17 Taylor, E, Walsh, S and Bradshaw, S (2018). Industry responses to the malicious use of social media. Riga: NATO STRATCOM COE. Available at: stratcomcoe.org/pdfjs/?file=/cuploads/pfiles/web_nato_report_-_industry_responsense.pdf?zoom=page-fit

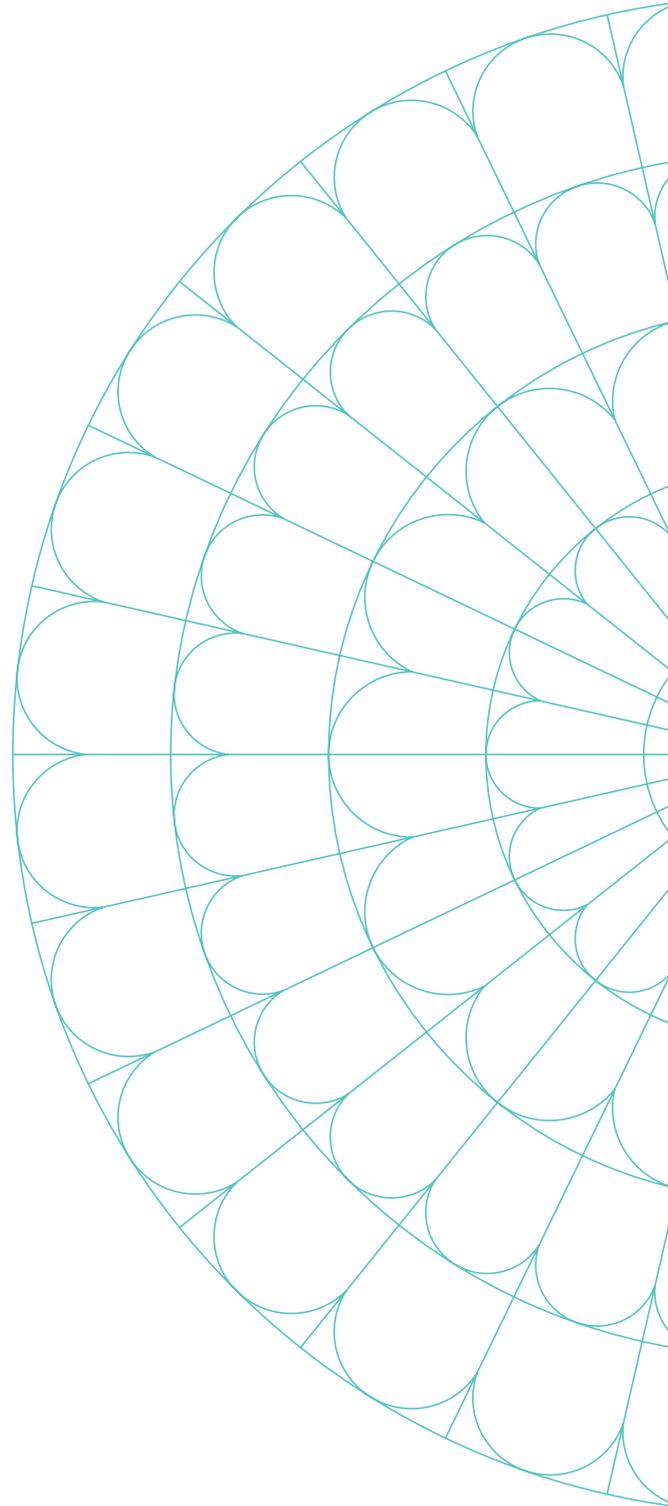
18 Avaaz (2020) How Facebook can Flatten the Curve of the Coronavirus Infodemic [PDF] Avaaz. Available at: avaazimages.avaaz.org/facebook_coronavirus_misinformation.pdf [Accessed 24 August 2021].

19 Carnegie UK (2021) The Draft Online Safety Bill: Carnegie UK Trust initial analysis. Carnegie UK p.16 [PDF] Available at: d1ssu070pg2v9i.cloudfront.net/pex/pex_carnegie2021/2021/06/05091934/draft-OSB-CUKT-response-FINAL-1.pdf

The five tribes of information consumer: a taxonomy

In January and February 2020 the RSA conducted polling with Kantar research. This resulted in five key profiles of the UK public, based on their behaviours and attitudes around information consumption, both online and offline, and in regard to both good quality information and misinformation. We wanted to understand who is susceptible to, consuming, and spreading misinformation and why.

Overall, we found it is not any one group. It's many groups of people. Many people from many different backgrounds, attitudes, political persuasions and mindsets. This affirmed to us the importance of producing and trialling tailored, or targeted, responses and solutions to misinformation. While advertising and selling online is often highly targeted, to our knowledge the same is not done for the spread of good and reliable information. But to be able to begin this journey, we must understand the behaviours and motivations of different groups. We therefore identified our 'five tribes of information consumer':





The Hyperactives

The Hyperactives are the superspreaders of online news and information. They like to share news and pay less attention to what they read and share to others and are therefore more likely to inadvertently share misinformation.

12 percent of sample



The Selectariat

The Selectariat are a small group of highly critical consumers of news and information. They are the only people within the sample to use established fact checkers.

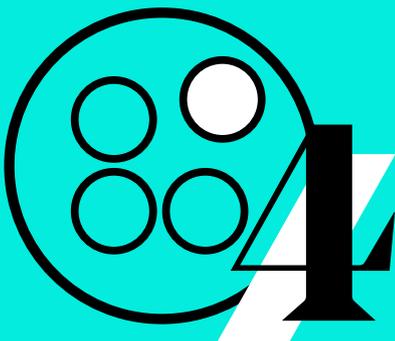
7 percent of sample



The Hoaxers

The Hoaxers are ardent believers in common modern conspiracy theories: 5G is bad for health, global warming is a hoax, and vaccines are harmful. To counter these theories we must first identify, understand, and then reach out to these groups

6 percent of sample



The Disenfranchised

The Disenfranchised feel out of step with the general population. They have a strong distrust of 'the media' and feel that no media outlets 'speak for them'. This persona highlight the inherent difficulty with reaching certain critical groups, who are disengaged from mainstream sources of information and resistant to edicts to change.

14 percent of sample



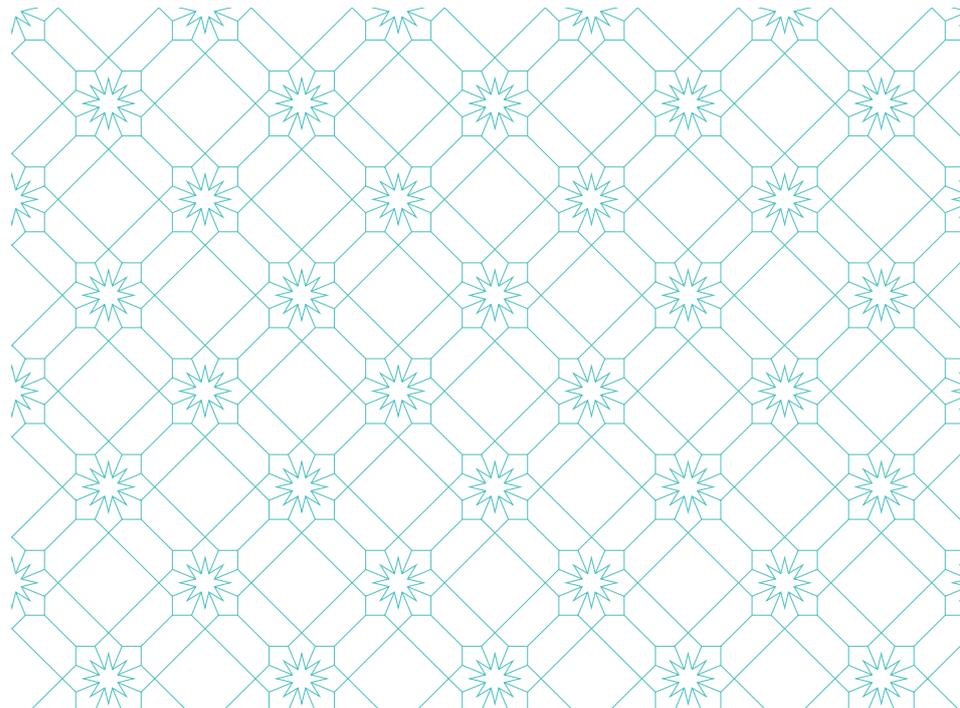
The Traditionalists

The traditionalists use online media the least. They heavily rely on offline news: TV, print news, and radio. This group, however, are still susceptible to believing disinformation and conspiracies, showing it is not a problem that can just be tackled online, offline solutions are needed too.

11 percent of sample

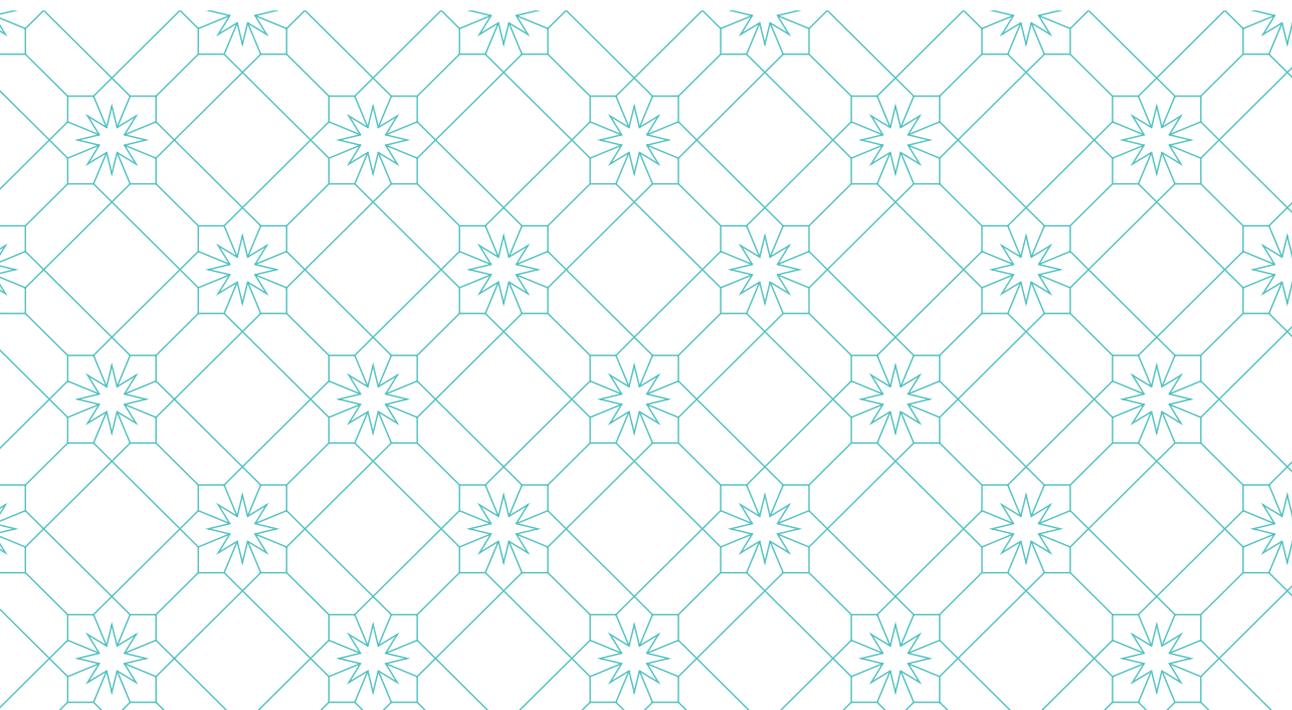
1. The Hyperactives are the superspreaders of information online, including misinformation. They are highly active on social media and take relatively little care with what they share. Forming 12 percent of our sample, they are also relatively young, with 36 percent being under the age of 35, double the rate for the sample as a whole. Compared to other groups, very few Hyperactives say they do not trust any media at all. Hyperactives were also less discerning of the fake headlines we showed them and were more likely to believe in common conspiracy theories. Overall, the Hyperactives are a group which, through their frenzied social media lives, inadvertently and occasionally purposefully share misinformation, thereby perpetuating its presence and reach.

2. The Selectariat are the most discerning group of the profiles we identified. They show greater concern for the spread and effect of misinformation and back this up with model behaviours. They were the most likely to cross-reference the news they saw (57 percent report doing this) compared to the sample as a whole (33 percent). They were also most likely to know and be able to name established fact-checkers. However, the Selectariat are only a small group, representing 7 percent of the whole sample, and just under half being aged 35-55. If we wish more people to be as discerning in the information they consume or share, then further understanding the attitudes and behaviours of the Selectariat is vital.



3. The Hoaxers are the modern conspiracists. They believe, or are inclined to believe, all of these common conspiracy theories: 5G is harmful, vaccines are harmful, and global warming is a hoax. This group is often younger and highly active on social media, and in particular Facebook – half of Hoaxers see news on social media daily, and of those that do 84 percent said they used Facebook. The Hoaxers might be what many have in mind when they consider who believes, and spreads, misinformation and disinformation. However, our research shows while they are important, they are not the only group prone to believing false information. It is therefore important to remember a view in which the only people who are vulnerable to misinformation are ‘others’, which almost never includes oneself or people we know, could result in complacency.

4. The Disenfranchised feel out of step with the general population. They are a group who say they feel that no mainstream media outlets represent them or their views. As such, they have a strong distaste for the media, the government, and politicians. In short, they do not like or feel supported by ‘the establishment’. They are the largest personality profile we found comprising 14 percent of the sample and are a broadly representative spread of the whole population in terms of age, gender, education, socio-economic group, and income. What unites them is distrust in authority a libertarian mindset – they are more likely to report feeling ‘governments have too much control over people’s everyday lives’. This also further highlights the difficulties with building single solutions to misinformation with an implicit or explicit type of user in mind.



5. The Traditionalists are those that rely most heavily on ‘traditional’ or offline news: TV, print news, and radio. They are broadly older than other groups, with two-thirds being 55+, compared to 45 percent of the sample, and slightly more likely to be men (57 percent).

The Traditionalists were more likely than most to say they trusted and felt represented by traditional media. This group is interesting because we find the Traditionalists are not immune to misinformation – scoring broadly the same as the sample average in their belief of our false headlines. However, because of their relatively analogue world, Traditionalists can fall out of view of researchers, technology companies, and potentially of regulators. When thinking about longer-term solutions to misinformation, such as information literacy programmes or improved regulation, we must consider how those who are less digital would use and access these services.

The five tribes we analysed are just the beginning. From our research we found these to be five of the most important groups, but they do not comprise the entire possible spectrum. Further research should be done to locate other profiles and refine them, in order to provide better services, solutions, and ultimately better information to everyone.

Combatting online harm: Recommendations

Understanding the nature of misinformation actors within our information ecosystem compels us to look at the problem of misinformation anew. Our study elicited four principal recommendations, but we hope the research and practice communities will be able to lean on the tribes in future studies.

For now, we argue for the following:

1 The draft Online Safety Bill should explicitly include societal harms caused by misinformation within its remit. It is clear that misinformation is not just an individual but a social problem, with social vectors of spread and social solutions. This must be recognised in legislation. In practical terms, we recommend this is done through a body separate from Ofcom which we term the Office for Public Harms. The Office would have the responsibility to investigate societal harms caused by misinformation and disinformation. It would then publish its findings publicly, as well as warn platforms of issues it finds and offer advice to Ofcom as to means of addressing the issues. **The Office should also act as a ‘misinformation ombudsman’** whereby it investigates cases of societal or individual harms brought to it, or would investigate where there has felt to have been overreach into the right to free expression. The Office would then state the expected redress. We believe the new body should be made up of a pluralist panel of stakeholders including citizens, Ofcom, platforms and wider industry, traditional media, civil society,

researchers, and other experts. Such a panel would have greater legitimacy than the current heavily platform-controlled information online.

2 A ‘polluter pays’ levy of social media firms to counter misinformation. The problem of misinformation is a social problem, but it has been magnified and intensified by the behaviour of social media companies. Thus, the solution must be funded by them.

The levy would be used to fund the workings of the Office for Public Harms (see recommendation 1), in a similar arrangement to the current funding model for the Advertising Standards Authority (ASA). The levy should also be used to fund media and information literacy drives in the UK, an innovation fund to research best practice in overcoming the harms caused by misinformation, and other further research. Other organisations, such as Glitch, have called for a tax on social media firms equivalent to 10 percent of the recent Digital Services Tax – which itself is set at 2 percent of revenue of particular internet firms.

3 Track and trace system for fake news. Over the course of our research, we uncovered multiple layers of social initiative that seek to counter the social and civic problem of misinformation spread. We believe some of these should form part of the code of conduct to be written by Ofcom (as stipulated within the Online Safety Bill). We have dubbed these a ‘track and trace system for fake news’. These include:

- i.** Corrections of false or misleading content online should be published within 24 hours of the content being released by the relevant fact-checking agency.
- ii.** Proportionate push notifications to users who have seen or interacted

with misinformation should be used wherever possible. By proportionate we mean that those who have engaged more directly should receive stronger messages than those who passively scrolled past.

- iii. Accounts and users who regularly share misinformation, including notable public figures, should be removed.
- iv. Comprehensive data on the reach and engagement with misinformation, and subsequent engagement with corrections, should be made publicly available for researchers and organisations to improve their messaging or services.

It is important to note however that such solutions should be trialled and iterated, based on the data of their effectiveness with different profiles of user. This was the key takeaway from our own profiling exercise; there is no single idealised misinformation consumer or spreader. Profiling users on their behaviours around information online will help us support and tend the public square.

4 A citizens' convention on misinformation: online rights, freedom, and the 'right to be informed'. To emphasise the collective nature of the misinformation problem, to guide the work of government and of the Office for Public Harms, a citizens' convention on misinformation should be called. When asked the right questions it is our belief that citizens can meaningfully engage and add to the debates between freedom of expression, harms caused by misinformation and the right to accurate information.

Conclusion and further work

The pandemic has been an unfortunate reminder of the power of bad faith actors, using poorly regulated but hyper connected social media, to become amplifiers and vectors of misinformation. Even relatively mainstream public figures, not just grey-and-black market 'shock-jocks', can inadvertently fall into the trap of sharing false or misleading hype stories.

Reform and tight regulation of online harms, including both harms to individuals and societal harms, is needed. We believe our recommendations would support this cause. Yet there is also something in the way that complex and technical information is understood and shared which is broken. This needs to be investigated and our educational and institutional practices should be reformed, with much greater emphasis and drive towards information and media literacy for all ages. These challenges sit behind the recommendations in this piece.

An Office for Public Harms or a citizens' convention may begin the process for a cleaner public square, but further work in this series should examine these issues in more detail. Fail to progress this work, to realise the social and institutional nature of misinformation spread, and the routes to a better information ecosystem are narrowed. This report is a taxonomy and a starting point of a civic response to the epochal challenge of misinformation. But the harm of misinformation is not merely personal but social and democratic.

MOVING AGAINST MISINFORMATION

We live in an age characterised by an abundance of online, social media-based information and of misinformation.

The scale of the misinformation challenge

We live in an age characterised by an abundance of online, social media-based information and of *misinformation*.

To give a sense of the scale of this challenge consider:

- Social media are increasingly cited as influential sources of news in polls – a 2016 study from Pew Research found that 62 percent of adults in the US name social media as a news source.²⁰ In the UK 49 percent of adults consume news via social media. Of those that do, 71 percent use Facebook and 46 percent use Twitter.²¹
- On social media platforms like Facebook, false stories have a greater reach and likelihood of virality than true stories. Falsehoods are 70 percent more likely to get retweeted than truths.²²
- These falsehoods reach 1,500 people on average six times quicker than accurate news stories.²³
- Since the start of 2020 the false content producers' industry online has grown by 102 percent.²⁴
- Outputs from fact-checking organisations saw an increase of 900 percent between January and March 2020.²⁵

Just as the printing press was the technology that drove the enlightenment era – bringing a new wave of human development and thought through sharing ideas, books, and pamphlets – so the internet is the technology driving the current age of information. And misinformation.

Crafting a civic and policy response rooted in practice and evidence is an epochal challenge. In 2019, we at the RSA were encouraged to turn our attention to this challenge; to help progress society's understanding of the way that mass movements and civil society might be enjoined in the fight against misinformation. Then in 2020, the Covid-19 pandemic proliferated. The research that forms the substance of this report became research not only into the idea of freedom of speech, or the activities of online trolls, but into the most potentially harmful misinformation pandemic of our time, and how we mobilise against it.

To explore these issues, we drew upon a mixed-methods research approach that combined a thorough literature review, a public opinion poll in partnership with Kantar of 2,074 UK adults, the views of experts within our advisory board (named in the appendix), through expert interviews, and through a full day roundtable event which took place in February 2020. This research is captured in the report that follows.

20 Gottfried, J. and Shearer, E. (2016) News Use Across Social Media Platforms 2016. Op cit.

21 Ofcom (2021) News Consumption in the UK: 2021. Op cit.

22 MIT Sloan School (2018) Study: False news spreads faster than the truth. Op cit.

23 Ibid.

24 Kornbluh, K, Goldstein, A and Weiner, E, 2020. New Study By Digital New Deal Finds Engagement With Deceptive Outlets Higher On Facebook Today Than Run-Up To 2016 Election. [online] Op cit.

25 Brennen, JS, Simon, FM, Howard, PN and Nielsen, RK. (2020) Types, Sources, and Claims of COVID-19 Misinformation. Op cit.

Covid-19

The World Health Organisation warned at the beginning of the Covid-19-pandemic that we were battling not just the virus but with the information ecosystem around it: both the pandemic and the 'infodemic' of pandemic 'trolls and conspiracy theories' and bunk cures.²⁶

Outputs from fact-checking organisations saw an increase of 900 percent between January and March 2020.²⁷ Most was spun or recontextualised factual information whereas less was completely fabricated, the former receiving far greater social media traction. Oxford University reported that 88 percent of the misinformation they analysed was found on social media but was particularly acute on privately encrypted chat channels such as WhatsApp.²⁸

Politicians, celebrities, and influencers were a key factor in the spreading of 5G and coronavirus claims. Evidence from Oxford University and Reuters contends that politicians, celebrities, and influencers create 20 percent of coronavirus and 5G conspiracy theory claims but are responsible for 69 percent of social media engagement.²⁹

According to the International Fact-Checking Network, coronavirus misinformation in particular came in five waves:³⁰

- 1** Myths about the origin of the virus.
- 2** Claims that falling down is a coronavirus symptom. The second wave was shaped by videos of people falling down in China, allegedly as a result of Covid-19.
- 3** False cures. The third wave dealt with fake remedies and unreliable vaccines. Fact-checkers have debunked stories claiming that household items such as bleach, very hot water, garlic soup or a special kind of tea are viral cures.

4 Supremacist claims. The fourth wave focused on claims that certain religions or races were immune to the virus.

5 Testing and lockdowns. As people were forced to stay at home, the fifth wave dealt with claims related to testing and quarantines. Some of these were deliberate falsehoods, whereas others were misunderstandings about the rules.

To this we would then add a sixth: vaccine misinformation. This takes on various guises, from links with Bill Gates or attempts to control the population, to reprisals of previous health scares around vaccines such as links to illness or fertility.³¹

False or mistimed rumours like these have already had direct harm our generational fight against the virus. To take one example, a CNN report on a lockdown in Northern Lombardy, hours before the Italian government made an official announcement, could have been a possible factor in a surge of cases in Southern Italy as residents fled south.³² Similarly, in London, prior to any official guidance from the UK government within the first lockdown, rumours of a lockdown and of army enforcement were circulating widely on social media, WhatsApp, and in the official press, spurring potentially infected individuals to leave the city. An exodus reoccurred at Christmas 2020, this time

26 BBC News (2020) WHO Warns Against Coronavirus 'Trolls' [online] 08 February. Available at: www.bbc.co.uk/news/world-51429400.

27 Brennan, JS, Simon, FM, Howard, PN and Nielsen, RK. (2020) Types, Sources, and Claims of COVID-19 Misinformation. Op cit.

28 Ibid.

29 Ibid.

30 Suárez, E (2020) How Fact-Checkers Are Fighting Coronavirus Misinformation Worldwide. Reuters [online] 31 March. Available at: reutersinstitute.politics.ox.ac.uk/risj-review/how-fact-checkers-are-fighting-coronavirus-misinformation-worldwide.

31 See: Goodman, J and Carmichael, F (2020) Coronavirus: Bill Gates 'microchip' conspiracy theory and other vaccine claims fact checked [online] BBC News. Available at: www.bbc.co.uk/news/52847648 [Accessed 31 August 2021] Also see: Schraer, R (2021) Covid vaccine: Fertility and miscarriage claims fact-checked [online] BBC News. Available at: www.bbc.co.uk/news/health-57552527 [Accessed 31 August 2021]

32 Cinelli, M et al. (2020). The covid-19 social media infodemic, Op cit.

due to last minute changes to guidance, prior to the second lockdown.³³

Conspiracy and false information in some senses then reached its zenith on 6 January 2021, when the world witnesses an insurrection on the Capitol in Washington DC by a mix of Trump supporting QAnon and 'stop the steal' conspiracists. The building was breached, and US politicians were forced to evacuate. The rioters believed that mass electoral fraud had taken place and there would be a form of global reckoning for deep state actors.³⁴ Such events should act as a warning to democracies around the world as to what could happen if conspiracies, misinformation, disinformation, and toxic content grow unchecked.

Toxic content

Today's information ecosystem is not only punctured by half-truths and lies but is also characterised by **toxic content**; content targeting human frailty and outrage. This sits within an ever-expanding question of wider **online harms**.



In the first quarter of 2018 alone, Facebook removed more than 2.5m pieces of hate speech and 'violent content' from their platform.³⁵ One survey by Ofcom suggested that one quarter of children aged 12–15 said they had seen abusive content online, and 20 percent said they had seen hate speech.³⁶

As the content we are recommended and see online is increasingly automated, the problem grows exponentially. Algorithms share and, in some cases, create content. This is then promoted using the tools of what Harvard Academic Shoshana Zuboff has referred to as 'surveillance capitalism' - data about individuals' habits and proclivities collected and analysed at scale and correlated with new types of toxic content. Meanwhile the accountability for the production and platforming of such content is said to be more diffuse and problematic. As well as an era of misinformation, we live increasingly in an age of plausible deniability about who is responsible for it.

Box 2: Definitions

Misinformation: inadvertent sharing of false or inaccurate information, which can be true but incomplete or dated, thereby rendering it inaccurate or open to interpretation.

Disinformation: 'the deliberate creation and dissemination of false and or manipulated information that is intended to deceive and mislead audiences, either for the purposes of causing harm, or for political, personal or financial gain'.³⁷

Within this report we default to the word misinformation to describe false information whether deliberate or not, but will clarify if we specifically mean inadvertent sharing. We use the term misinformation to refer to deliberate false content.

33 Hockaday, J, 2020. Last minute scramble to leave London as capital enters tier 4 lockdown. Metro, [online] Available at: metro.co.uk/2020/12/19/last-minute-scramble-to-leave-london-as-capital-faces-tier-4-lockdown-tomorrow-13778954 [Accessed 3 September 2021].

34 Porter, T (2021) QAnon supporters believed marching on the Capitol could trigger 'The Storm,' an event where they hope Trump's foes will be punished in mass executions. Op cit.

35 Rosen, G (2018) Facebook Publishes Enforcement Numbers for the First Time. [online] Facebook. Available at: about.fb.com/news/2018/05/enforcement-numbers [Accessed 3 September 2021].

36 Ofcom, 2020. Online Nation. Raising awareness of online harm. [online] Ofcom. Available at: www.ofcom.org.uk/data/assets/pdf_file/0027/196407/online-nation-2020-report.pdf [Accessed 22 October 2020].

37 DCMS Committee. (2019). Misinformation and 'fake news'. Op cit.

Tending the public square

What can we do about the misinformation deluge? What should we do?

We have been conditioned to think of online speech as a matter of rights and freedoms. Space precludes a detailed philosophical investigation of the alternatives. Suffice to say that this publication, while mindful of the importance of such critiques, takes a slightly different approach. We look at it from the perspective of the sanctity of the public square and our collective responsibility towards it.

One cannot inoculate oneself against lies, but we can identify actors and pathways that enable the wanton dissemination of falsehoods and tend the public square in such a way as to limit their impact on the whole community.

This mission can never, by definition, be the work of one organisation, individual or plan. That's why we sought in this research study to understand better the kind of institutional and system-wide factors that we might leverage in this direction. We sought to understand the way that people – individuals and groups – make use of and dispose of information. For our starting point is this: the challenge we face will require nothing less than a mass mobilisation of diverse actors in order to make progress. And to so mobilise, we need to imagine a regulatory shape and structure which would allow a diverse range of actors to work together.

We also need to understand in what ways different people interact with information and misinformation, what lures and triggers belief in false information, and what can be done to limit the spread and damage.

The insight and original research about what drives our misinformation ecosystem, and how to remedy it, contained herein, we hope, is useful in itself in the broader discussion around misinformation. We also present conclusions and recommendations we have drawn in the course of the limited time frame of this project, especially as they pertain to the current UK debate about online harms.

This paper is structured as follows:

In chapter 2, coming next, we paint the broad strokes of the current response to misinformation. We show how global civil society and national regulators have attempted to use the tools at their disposal to take on misinformation crossing national and international boundaries, associations, and groups.

In chapter 3, we share our own original polling, which yields five tribes of British information consumer. If you would like to learn who moves misinformation and why, you may skip here.

Interviews with experts follow in the next chapter, who helped shed further light on the psychological motivations undergirding each of the five tribes. There are many fascinating insights contained herein.

Finally, in the fifth chapter, we close with recommendations centred on those five tribes and especially as they pertain to current British legislation around online harms. We hope this section will also be useful for international readers interested in their own jurisdictional response to the misinformation challenge.



**TO THE
CHALLENGE**

**A WORLD
RESPONDS**

Facebook, Twitter and Instagram, streaming platforms such as YouTube, and chat applications such as WhatsApp and Telegram have amassed a huge global membership and significant power as major purveyors of information.

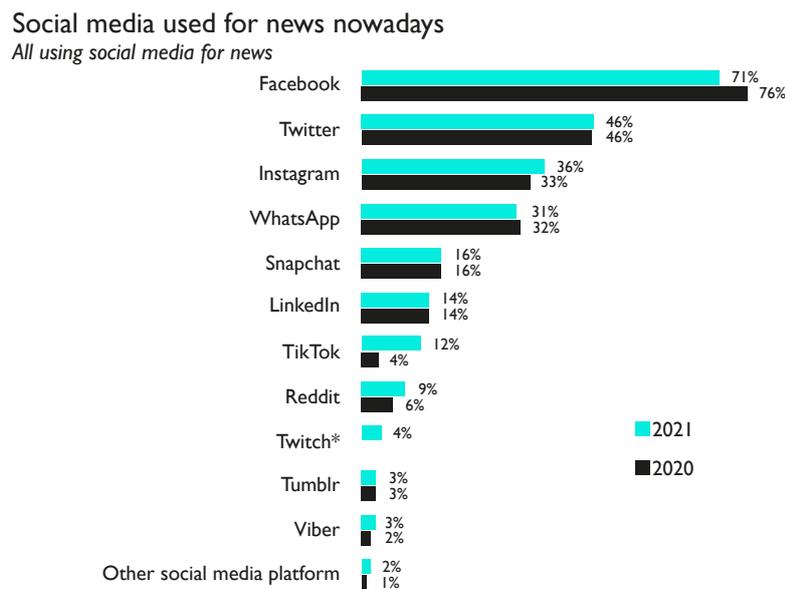
2 A world responds to the challenge

According to a survey conducted in 2021 by Jigsaw Research for Ofcom, 49 percent of UK adults consume news via social media. Of those that do, 71 percent use Facebook and 46 percent use Twitter.³⁸ As well as digital platforms hosting traditional media, these tools have widened the range of amateur producers of information to include citizens, amateur journalists, freelance writers, bloggers, and advocacy groups.

Per recent reports, Twitter averages 199 million monetisable daily users;³⁹ 2.76 billion people use Facebook, Instagram, or Messenger daily;⁴⁰ YouTube also holds an audience of 2 billion users monthly.⁴¹

In response, policymakers, NGOs, academics, and private institutions across the globe are working to understand the nature and scope of misinformation and its effect at a local, national, and global scale.

Figure 1: Ofcom and Jigsaw research on news consumption via social media⁴²



38 Ofcom (2021) News Consumption in the UK: 2021. Op cit.

39 Twitter (2021) Letter to Shareholders [online] Available at: s22.q4cdn.com/826641620/files/doc_financials/2021/q1/Q1'21-Shareholder-Letter.pdf [Accessed 12 August 2021].

40 Facebook (2021) FB Earnings Presentation Q2 2021 [online] Available at: s21.q4cdn.com/399680738/files/doc_financials/2021/q2/Q2-2021_Earnings-Presentation.pdf [Accessed 13 September 2021].

41 YouTube (2021) YouTube for press [online] available at: blog.youtube/press [Accessed 12 August 2021]

42 Ofcom (2021) News Consumption in the UK: 2021. Op cit.

The civil society response: messy, fractured, heroic, necessary

Civil society, be they citizen led initiatives or other non-governmental actors, have a critical role to play in limiting the spread and damage of misinformation. Civil society movements and organisations provide critical nuance, context, and balanced opposing views to information online and so should be considered and included within long-term solutions. What follows is the tip of the iceberg in terms of the global civil society response, covering five areas: media companies, universities, commissions, citizen movements and media literacy networks.

Media companies

There have also been contributions from the traditional media to the fact-checking mission. The **BBC** has assembled a permanent team, **Reality Check**, with the aim of exposing false or misleading stories, and sits alongside well-known independent fact-checkers such as **Full Fact**, or Channel 4's **FactCheck**. New in the British fact-checking landscape is **Infotagion**, a fact-checking site with the backing of Damian Collins MP, formed during the pandemic to fight the 'information contagion about Covid-19'.⁴⁴ In addition, **Le Monde** has developed a verification unit **Les Décodeurs** who use a decoding machine to speed up the fact-checking process by alerting readers to previously verified claims.⁴⁵



“The BBC can't edit the internet, but we won't stand aside either.”

James Harding,
BBC News Chief, January 2017⁴³

Traditional media remain a key component of the information ecosystem, as becomes clear when considering the Traditionalists in our profiling in chapter 3. Yet we also know from our profiling that fact-checking organisations remain underutilised and little known to the public, despite organisations such as **Full Fact** being used by Facebook to fact check claims online, as we have seen.

- 43 Jackson, J. (2017). BBC sets up team to debunk fake news. [Online] available at: www.theguardian.com/media/2017/jan/12/bbc-sets-up-team-to-debunk-fake-news [Accessed 06 October 2021].
- 44 Infotagion (no date) About - Infotagion. [online] Available at: infotagion.com/about [Accessed 24 August 2021].
- 45 Mantzarlis, A. (2016). Le Monde wants to build a B.S.-detector Poynter [online] 22 November. Available at: www.poynter.org/fact-checking/2016/le-monde-wants-to-build-a-b-s-detector-2.

Universities

Verificado 2018 was a similar initiative created as a collaboration between 60 universities and media and civil society organisations to provide political fact-checking in **Mexico**, where politics and journalism can be notoriously corrupt.⁴⁶ Although short-lived, the initiative did garner some success in conveying legitimacy to the public and was widespread in its reach; in a four-month period the website received over five million views.⁴⁷ The longer-term impacts are still to be observed but the intention was that participating journalists would continue to work collaboratively and instil verification in their everyday practices long after the group dispersed.⁴⁸

Specialty convened commissions

The following institutions are also a critical part of the counter-misinformation eco-system:

- The Oxford Internet Institute is looking at how misinformation, propaganda, and divisive political news spreads and the impact of this on the health of the UK's 'online information ecosystem'.⁴⁹ Through empirical tracking of the reach and spread of this new form of online media on some of the UK's biggest social media platforms, Twitter, Facebook, Reddit and YouTube, the OII hopes to better understand the importance of producers in divisive content creation, how stories and narratives are spread, a comparison of different misinformation dissemination strategies at a national level, and the impact of exposure to misinformation, propaganda, and divisive political news on the consumer.

- The Royal Society is looking at how to create a 'positive information environment', thinking about how digitalism is affecting peoples' interactions with information, how tech can be used to create or find misinformation, and what other environmental controls tech can influence.⁵⁰ The Royal Society will look at technological trends that could influence who sees what information and the impact this has on consumers, communities, and society.
- LSE's Truth Trust and Technology Commission, which concluded in 2020, identifies an information crisis in the UK.⁵¹ The commission identifies 'five giant evils' of the information crisis affecting the UK public; confusion about what information to believe, cynicism regarding even trustworthy information, apathy as distrust erodes engagement, fragmentation of citizens into 'truth publics', and irresponsibility on the part of organisations that aren't transparent or accountable. As part of their research programme LSE have also put out policy briefs on a variety of subjects from platform responsibility to media and information literacy in the school curriculum.

46 Hughes, S, and Márquez-Ramírez, M. (2017). Examining the practices that Mexican journalists employ to reduce risk in a context of violence. *International Journal of Communication*, 11(23), pp499–521.

47 Martínez-Carelló, N and Tamul, D. (2019). (Re)constructing Professional Journalistic Practice in Mexico: Verificado's Marketing of Legitimacy, Collaboration, and Pop Culture in Fact-Checking the 2018 Elections. *International Journal of Communication* 13, pp2596-2619.

48 Trewinnard, T. (2018). #Verificado2018: Mexican media unites to fight mis and misinformation ahead of historic elections Medium [online]. Available at: medium.com/popupnews/verificado2018-mexican-media-unites-to-fight-mis-andmisinformation-ahead-of-historic-elections-6ff34c72cbb7.

49 For instance, see: Oxford Internet Institute (no date) Programme on Democracy and Technology. Op cit.

50 Royal Society (no date) Digital technology and information [online] Available at: royalsociety.org/topics-policy/projects/digital-technology-and-information [Accessed: 24 August 2021].

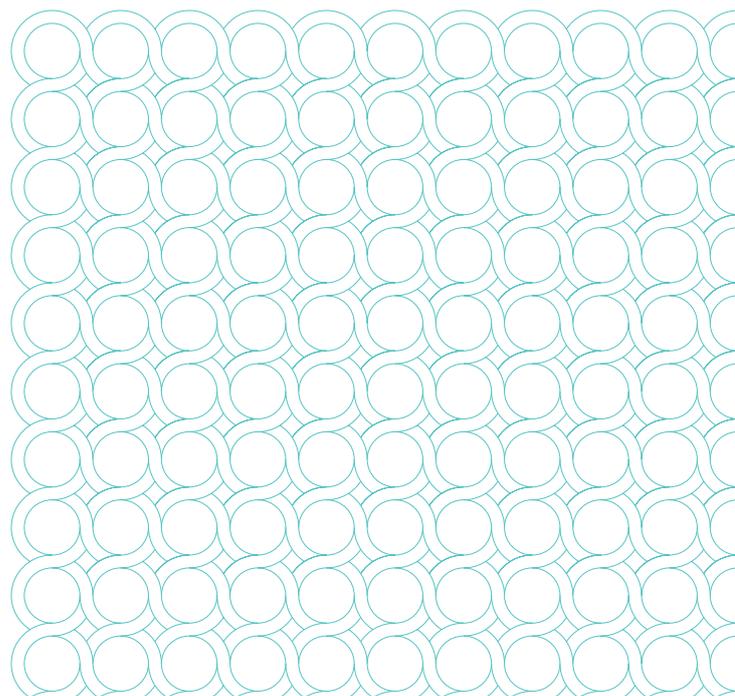
51 London School of Economics (2020) Tackling the information crisis: a policy framework for media system resilience. Op cit.

Citizen movements

Civil society is both an independent voice, and a repository of knowledge and cultural understanding. **The Elves**, to take one example, are a one of a number of independent group of citizens in **Lithuania** who united in the battle against misinformation.⁵² They coordinate via Skype and Facebook, with the aim to bolster the authenticity of social media through denouncing fake accounts and exposing falsities. This approach has been highly successful, receiving recognition from the Lithuanian government and NATO. In addition to combatting scepticism towards the authorities, it has encouraged an environment of greater social trust.⁵³ Crucially, this is operating in a Baltic state with close proximity to Russia, which may be fostering an inherently stronger resistance to misinformation.

Media literacy networks

BBC Academy, the training arm of the broadcaster, provides reporting, education and training on misinformation and fake news for current and potential future employees of the BBC, and wider broadcast professionals in UK media sector.⁵⁴ It covers media literacy on areas such as microtargeting and filter bubbles, as well as tools for journalists on verifying content. Other organisations such as **First Draft News** also have well-regarded training programmes to combat misinformation.⁵⁵



- 52 Soon Wan Ting, C and Goh Ze Song, S. (2017). What Lies Beneath the Truth: A literature review on fake news, false information and more. Op cit.
- 53 Au-Yong, R. (2017). Call to empower citizens to combat fake news [online] 20 June. Available at: www.straitstimes.com/singapore/call-to-empower-citizens-to-combat-fake-news.
- 54 A public facing site for training against misinformation can be found here: BBC (no date) Beyond Fake News. [online] Available at: www.bbc.co.uk/beyondfakenews [Accessed 24 August 2021].
- 55 First Draft (no date) Training [online] Available at: firstdraftnews.org/training [Accessed 24 August 2021].

The regulatory response: often too light touch, often not agile enough

The depth and diversity of the civil society response is both a function of the scale of the problem and the difficulty that traditional regulators have dealing with misinformation as a social challenge. RSA polling research found that lawmakers remain unsure of how to tackle the burgeoning power of social media giants, and the interrelated social and technological issues associated with our new (mis)information ecosystem. Only 15 percent of British MPs reported that they believed that parliament has the capacity to respond. And only 29 percent believe they themselves know enough about the issues at play. This may well have been elevated due to the awareness-raising and agency emphasising impact of the introduction of the draft Online Safety Bill but it shows a relatively recent under-developed capacity within parliament. We also know that the British public want robust action on misinformation. Our pre-pandemic polling with the public found that **71 percent said they want a ‘stronger independent regulator on the quality of news’**.

All over the world, the legal and regulatory framework to tackle misinformation is evolving but fundamentally is still lacking. Despite varying experiments and attempts at limiting the spread and impact of misinformation the problem continues.

Attempts to control social media content, including but not limited to, misinformation and disinformation, fall into a number of broad categories. Below we list but four of these ways. This is non-exhaustive and not mutually exclusive, but an attempt to pull apart the most common governmental and non-governmental approaches taken thus far.⁵⁶

1 Mandate misinformation content moderation and removal. Several countries around the world have begun to put into place laws that mandate the removal of content deemed misinformation and disinformation. Ranging from the Philippines, to India, to France (though in France it is only meant to be used around elections). Such laws pose obvious dilemmas on who decides what is false and who is being silenced because of it. Of course, platforms themselves do remove misinformation, and receive high levels of public pressure to do so, but the key difference here is that removal is mandated by law.

⁵⁶ For a longer list of the types of regulation and control of online spaces, see: Full Fact (2018) Tackling misinformation in an open society [PDF] Full Fact: London. p11. Available at: fullfact.org/media/uploads/full_fact_tackling_misinformation_in_an_open_society.pdf [accessed 25 August 2021].

2 Regulating illegal content.

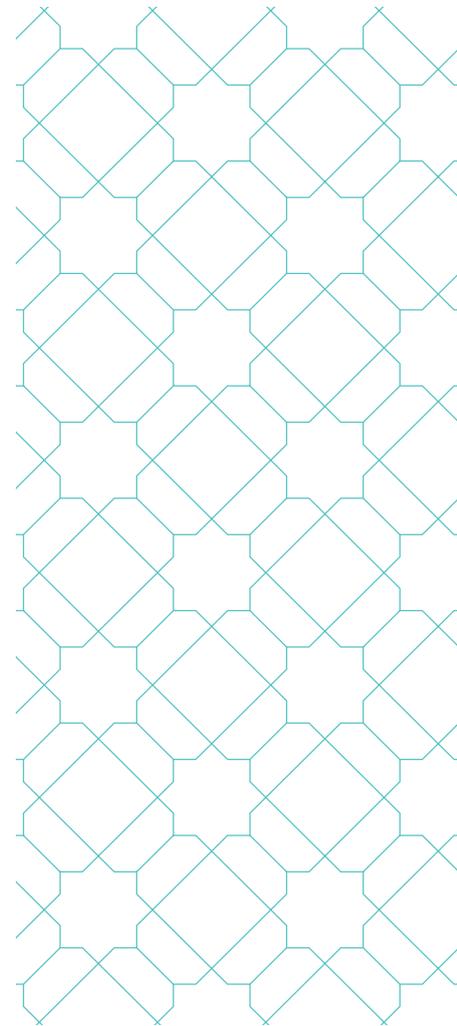
Other nations, in attempting to steer from the dilemmas mentioned in (1) but still wanting to make the internet broadly safer have instead opted to mandate the removal of more 'obviously' illegal content. For instance, racist content, illegal sexual imagery, terrorist content, or fraud. This is the direction taken by Germany and is broadly in line with the UK's draft Online Safety Bill. Ignoring difficulties in assessing when content becomes illegal, this approach is politically less contentious but can mean misleading or 'democratically' damaging content can still proliferate which raises the broader question about what we consider to be harmful.

3 Self-regulation by online firms.

Increasingly firms enacting their own policies limit the spread and damage of misinformation. Yet governments are also getting firmer in pressuring platforms to do more, voluntarily, and to simply uphold their own terms of service. Damian Collins MP, in an interview with Channel 4's Jon Snow on the Online Safety Bill, described platforms not upholding their terms and conditions as misleading and efforts to overcome this as a form of 'basic consumer protection regulation'.⁵⁷ This can be done alongside other means, and features as an element of the Online Safety Bill.

4 Resilience building through media and information literacy.

This can be done through education and training, but also through transparency or tools. Building resilience through a population is of course a highly desirable and liberally minded route, it is, however, by no means perfect and slower to start, and longer to reap benefits. However, this remains a critical piece of armour in shielding against misinformation and disinformation.



57 Collins, D (2021) Interview on Channel 4 News interviewed by Jon Snow. [Online] Channel 4, 29 July 2021. Available at: www.channel4.com/news/super-committee-to-examine-governments-online-safety-bill [Accessed 25 August 2021].

Mandating content moderation and removal of misinformation and disinformation

France passed a bill in 2018 consenting to the immediate removal of content that is deemed to constitute misinformation, with violations involving removal of content, the shutting down of sites, or penal and financial sanction.⁵⁸ The scope of the legislation is limited to content deemed a 'manipulation of information' within the three months preceding an election.⁵⁹ The legislation includes provisions for increased transparency on election advertising online and allows for political candidates, political groups, public authorities and individuals to sue for the removal of content, which must then be authorised by a judge within 48 hours.

While clearly a strong attempt at limiting misinformation, it does little to remedy the amplification of content (social content has the potential to reach many millions within 48 hours) and does little to inform and educate retrospectively – ie informing those who have already viewed such content that it has been removed and why.

Regulating illegal content

Misinformation can be inherently difficult to regulate because, while it is misleading and on aggregate can create systemic or societal harms, it is often not illegal to claim, for instance, that vaccines are harmful. Often dubbed 'legal but harmful' content, some countries have instead opted to regulate more obviously illegal content such as hate speech, terrorist related content, or racist content. Of course, difficulties with consistently determining what is hate speech without stifling free speech are still present, yet nevertheless some nations have opted to pursue this avenue.⁶⁰

Germany introduced the Network Enforcement Act in 2018 which requires social media platforms to take down illegal, racist or slanderous content within 24 hours. A failure to comply can result in a fine of up to €50m. The main criticisms of this measure relate to the impact on free speech and expression, platform self-censorship, and the transferal of responsibility in making complex legal decisions away from public authorities.⁶¹ However, the far-reaching nature of the Act must be considered alongside the historical context of widespread propaganda and political manipulation in Germany's past.

58 DCMS Committee. (2019). Misinformation and 'fake news'. Op cit.

59 Funke, D and Flamini, D. (2020). A guide to anti-misinformation actions around the world Poynter [online]. Available at: www.poynter.org/ifcn/anti-misinformation-actions/#france.

60 For a summary of the tensions inherent in regulating 'legal but harmful' content, as well as illegal content such as hate speech, see: House of Lords (2021) Free for all? Freedom of expression in the digital age London: House of Lords p36. Available at: committees.parliament.uk/publications/6878/documents/72529/default

61 Tworek, H and Leerssen, P. (2019). An Analysis of Germany's NetzDG Law [online]. Available at: www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf.

UK in focus: draft Online Safety Bill

The UK government's response has been caught within the scope of its Online Safety Bill. The goal is to create a statutory 'duty of care' for platforms to improve the safety of their users, which will be regulated by Ofcom.⁶² The scope of this Bill is on a full range of online harms, including harm to children, illegal sexual, and terrorism related content. The duty of care applies to all platforms for user-generated and shared content online.

The Bill puts great emphasis on the control and removal of illegal content, for instance through mandatory 'illegal content risk assessments' and on the protection of services or content which are likely to be accessed by children. The Bill is also specific in that harms done to individuals are prioritised whereas **societal** harms, may not be covered, unless specifically stated by the Secretary of State.⁶³

The Bill does give powers to Ofcom to review 'harmful but legal' content, but what constitutes 'harmful content to adults' is not clearly defined. It also requires harm to be done to individuals; societal or groups harms would therefore be out of scope. It is left to risk assessments by Ofcom, and by the biggest service providers to identify this content. This would include much of the misinformation and disinformation we see and refer to in this report. As we are concerned with protecting the health of the public square this approach raises issues about whether common goods on which, for example, public health and democracy depend can be captured effectively in this framework.

The only obvious and direct action taken in the draft Bill is the formation of an advisory committee on misinformation and disinformation by Ofcom. This committee will provide advice on how regulated services should deal with misinformation and disinformation.

The Bill does provide scope for harms to the individual to have been 'indirectly' caused, as noted by Carnegie, or for misinformation or disinformation to be in scope so long as it results in 'physical or psychological' harm to an individual.⁶⁴ But how and when such a harm would be deemed to have taken place is much less clear than in other provisions within the Bill.

Civil society has widely acknowledged the Bill as a significant step forward but there are a number of concerns around the breadth of the Paper and the definitional ambiguities. Critically, the Bill is quiet and vague on how to systematically deal with and limit the threat abundant misinformation can have on society and democracy. It is our view that the exclusion of collective harms from the scope of the Bill is a critical failing which must be addressed if we are to hope to improve the accuracy and nature of our information ecosystem.

62 Department for Culture Media and Sport (2021) Draft Online Safety Bill. HM Government: London [PDF] Available at: assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safety_Bill_Bookmarked.pdf

63 Carnegie UK (2021) The Draft Online Safety Bill: Carnegie UK Trust initial analysis. Carnegie UK p.16 [PDF] Available at: d1ssu070pg2v9i.cloudfront.net/pex/pex_carnegie2021/2021/06/05091934/draft-OSB-CUKT-response-FINAL-1.pdf

64 Ibid. Also see: House of Commons (2021) Digital, Culture, Media and Sport. Committee Oral evidence: The Work of the Department for Digital, Culture, Media and Sport, HC 44. Op cit.

Building resilience through media and information literacy: an untapped resource

Policies that aim to build resilience against misinformation are emerging across the globe. Some intend to convene debate and develop research to better understand this evolving field, whilst others focus on education and building skills in information literacy and critical thinking. Their ambition, however, is congruous; to aid citizens in navigating the labyrinth of online information.

European countries, and Nordic countries in particular, have been at the forefront of resilience-led interventions and have been commended for their focus on tackling misinformation. **Sweden**, for example, targeted its efforts on education, both through a school curriculum focused on media literacy, and a handbook for public-sector employees on countering misinformation operations.⁶⁵ Investment in media and information literacy is also a focus in **Finland**, where it is seen as a civic competence, important to every citizen from an early age.⁶⁶

Improving societal resilience is paramount for the future. Although the presence of misinformation is almost unavoidable, the harmful impact is not. In the global online environment, an ability to display critical judgement about multiple information sources is pivotal.⁶⁷



“I understand that regulation may hurt our business, but I think it’s necessary. Getting these issues right is more important than our interests.”

Mark Zuckerberg, CEO Facebook, April 2019⁶⁸



65 Cederberg, G. (2018) Catching Swedish Phish: How Sweden is Protecting its 2018 Elections. Harvard: Belfer Center for Science and International Affairs.

66 National Audiovisual Institute. (2020). Finnish Media Education: Promoting Media and Information Literacy in Finland [online]. Available at: kavi.fi/sites/default/files/documents/mil_in_finland.pdf.

67 Information Literacy Group. (2018). CILIP Definition of Information Literacy [online]. Available at: infolit.org.uk/ILdefinitionCILIP2018.pdf.

68 Feldman, B. (2019). Facebook Sets Aside \$3 Billion for FTC Fine, But Wall Street Isn't Worried. [Online] Intelligencer. Available at: nymag.com/intelligencer/2019/04/facebook-sets-aside-usd3-billion-for-ftc-fine.html [Accessed 06 October 2021].

Self-regulation: fool's errand?

Both in recognition of democratic and societal harm and due to external policy pressures, digital platforms have begun to self-regulate, ie enact their own policies and codes of practice around content on their sites. Between 2016 to 2018, there were more than 125 announcements of self-regulation measures.⁶⁹ Since the pandemic, platforms have gone even further in their activist approach (see box 2 below). These are, in all probability, a step in the right direction, but there are a number of existing tensions which explain why this approach is not working.

First, there are inconsistencies both within and across platforms in what constitutes misinformation, and how it is dealt with. Some of this is inevitable, if we do not want a single arbiter of truth then inconsistencies about what constitutes misinformation, and at what level it is deemed removable, is highly difficult to avoid. For instance, major platforms can be inconsistent in what constitutes climate misinformation and when this content should be removed. Facebook, despite stating that climate misinformation is a serious problem on their site, reportedly does not mention it on their advertising or community standards.⁷⁰ While only in June 2021 it was reported that Twitter would begin to more actively promote credible climate content.⁷¹

However, what is not inevitable are inconsistencies in process. We saw earlier in research by Avaaz that even when only considering Facebook, the platform often fails to flag fact-checked content in a timely or consistent manner.⁷² Practices across platforms are equally inconsistent (see below). We have also seen internationally examples where governments, Germany for instance, have mandated that content be removed within 24 hours if it constitutes hate speech. In our recommendations we detail a 'middle road' whereby Ofcom, as the central regulatory of online harms can put in place minimum standards for codes of practice when dealing with misinformation, while a dedicated misinformation regulatory authority, independent of Ofcom, can investigate individual and societal harms by misinformation, act as a form of misinformation ombudsman to offer a secondary means of redress and protection for citizens and society from misinformation, and act as a secondary and independent voice for what should be considered harmful misinformation (see chapter 5).

69 Taylor, E, Walsh, S and Bradshaw, S. (2018). Industry responses to the malicious use of social media. Op cit.

70 Avaaz (2021) Facebook's Climate of Deception: How Viral Misinformation Fuels the Climate Emergency [online] Avaaz. Available at: www.secure.avaaz.org/campaign/en/facebook_climate_misinformation [Accessed 31 August 2021].

71 Guynn, J (2021) Twitter will promote credible information with new climate change topic after criticism over misinformation [online] USA Today. Available at: www.eu.usatoday.com/story/tech/2021/06/03/twitter-pressure-climate-change-global-warming-topic/7512518002 [Accessed 31 August 2021].

72 Avaaz (2020) How Facebook can Flatten the Curve of the Coronavirus Infodemic. Op cit.

Means of self-regulation?

Facebook has launched a whole host of initiatives to augment the battle against misinformation. This includes enlisting the help of third-party fact-checkers and content moderators, more clearly labelling political advertisements and their funders, and revising their algorithm to demote articles that display misinformation warning signs.⁷³ The successes of such interventions are inherently difficult to measure, with the workings behind the algorithms and code enveloped in trade secrecy, as well as the outcomes of different measures not shared publicly. Facebook does share information on the volume of harmful content on its site, but critically does not share data on what happens to fact-checked, or otherwise moderated content and subsequent virality.⁷⁴ But even with this information it would be hard to know the qualitative impact on democracy and society.

Mark Zuckerberg has appeared to be somewhat hostile to cooperation thus far, refusing to give evidence before the UK-led international grand committee who were investigating the role of misinformation in elections.⁷⁵ This tension between government and platforms is a key area for remedy in the strive for a shared approach.

Twitter initially took a similar approach to Facebook but has since banned political advertising as of November 2019. This poses a challenge for other tech companies and sets a moral precedent. Twitter CEO Jack Dorsey has been forthcoming in his approach publicly apologising and making concrete changes. The platform has a widely recognised verification badge which is primarily to indicate celebrity but also signal to users that content is from a credible source, but this measure

is dependent on the existing level of media trust at a local or country-wide level. Verification badges may be less effective in Greece for example, where the media market is characterised by very low levels of trust and there is extreme fragmentation in the online news market.

The context in which the platform is operating is vital in determining the effectiveness of the intervention. WhatsApp, for example, introduced a feature which shows when a message has been forwarded and limits the number of forwards per group. This may seem negligible in a UK context, but for countries such as India, WhatsApp's largest market at 400 million users, this is a pivotal move. Reuters found that WhatsApp was misused in at least three ways in India for political campaigning.⁷⁶ The intervention does, however, still leave the responsibility of identifying the authenticity of a message to the user. This highlights the importance of a multifaceted approach to taking on the misinformation challenge, involving hard regulation but also partnership and a focus on information literacy and critical thinking skills.

73 Taylor, E, Walsh, S and Bradshaw, S. (2018). Industry responses to the malicious use of social media. Op cit.

74 See here for latest Facebook report on moderated content: Facebook (2021) Community Standards Enforcement Report, Second Quarter 2021 [online] Available at: www.about.fb.com/news/2021/08/community-standards-enforcement-report-q2-2021 [Accessed 25 August 2021].

75 Kates, G (2018) 3 more countries join "international grand committee" targeting Facebook's Mark Zuckerberg [online] CBS News. Available at: www.cbsnews.com/news/3-more-parliaments-join-international-grand-committee-targeting-facebook-mark-zuckerberg [Accessed 31 August 2021].

76 Vengattil, M et al (2019) In India election, a \$14 software tool helps overcome WhatsApp controls Reuters [online] Available at: www.reuters.com/article/india-election-socialmedia-whatsapp/in-india-election-a-14-software-tool-helps-overcome-whatsapp-controls-idUSKCN1SLOPZ.

Box 3: Changes in self-regulation in the time of Covid-19

Facebook early in the pandemic began more aggressively notifying users when they have interacted with proven misinformation on their platform, however they do not debunk the story directly, instead directing them to a generic WHO MythBusters website.⁷⁷

Within the pandemic **Twitter** has taken further action. First, it has been trialling a 'read before you retweet' feature to encourage better understanding and critical thinking, and recently announced trialling a 'report misinformation' function.⁷⁸

Facebook-owned **WhatsApp**, a much more difficult app to manage in terms of information veracity due to its end-to-end encryption, has taken two actions to fight against the infodemic. It has put further limits on the ability to forward content or messages, limiting the forwarding to one contact at a time. However, this is easily circumvented by manually attaching the file to messages or copy and pasting the message.⁷⁹

Instagram link to government health sites, such as the NHS or WHO site when you search #coronavirus, as well as removing hashtags linked to misinformation.

TikTok have been directing searches for coronavirus to the WHO website. And Snapchat created AR filters alongside the WHO and linked users to WHO official advice.

Platform-led resilience building

Twitter, **Facebook**, and **Google** have all created their own version of a plug-in to aid users escape their 'filter bubble' - FlipFeed, Escape Your Bubble, PoliEcho. They allow users, in various ways, to replace their own feed with that of a different political worldview or highlight the extent to which their feed is subject to political algorithmic bias.⁸⁰ Again, however, the impetus is on the individual to take action.

Other such approaches have focused on the development of research. Twitter, for example, provides funding for the Atlantic Council's Digital Forensic Research Lab.⁸¹

In the information literacy space, Facebook and Mozilla have launched the News Integrity Initiative which is a collaborative project designed to improve the general public's media literacy.⁸² These interventions form an important part of the bigger picture, but there are legitimate questions over who should be responsible for the educational side of the response and criticism that it is easier to support education programmes than mothball business models driven by misinformation and toxic content.

These questions are particularly pertinent precisely because there is, at least within the UK, a serious lack of government initiative towards media and information literacy, or to address misinformation online.

However, we must also consider *who* the resilience building efforts are for. *Who* do we find needs resilience-building the most, and equally *who* is already doing well? To help answer this, below we consider the five tribes of information consumer:

- 77 Newton, C. (2020). Coronavirus Misinformation Is Putting Facebook To The Test The Verge [online] 17 April. Available at: www.theverge.com/interface/2020/4/17/21223742/coronavirus-misinformation-facebook-who-news-feed-message-avaaz-report. [Accessed 6 October 2021].
- 78 BBC News. (2021). Twitter tests 'misleading' post report button for first time. [online] Available at: www.bbc.co.uk/news/technology-58258377 [Accessed 6 October 2021].
- 79 Hadavas, C. (2020). Whatsapp Limits Message Forwarding To Curb The Spread Of Coronavirus Rumors Slate Magazine [online] 08 April. Available at: slate.com/technology/2020/04/whatsapp-message-forwarding-disinformation-coronavirus.html?via=recirc_recent [Accessed 6 October 2021].
- 80 MIT Media Lab. (2020). Promoting deeper learning and understanding in human networks [online]. Available at: www.media.mit.edu/projects/flipfeed/overview.
- 81 Gadde, V and Roth, Y. (2018). Enabling further research of information operations on Twitter Blog [online] 17 October. Available at: blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html.
- 82 Mullin, B. (2017). Can trust in the news be repaired? Facebook, Craig Newmark, Mozilla and others are spending \$14 million to try Poynter [online] 03 April. Available at: www.poynter.org/tech-tools/2017/can-trust-in-the-news-be-repaired-facebook-craig-newmark-mozilla-and-others-are-spending-14-million-to-try.

**In 2020 the RSA
polled the public to
better understand
how individuals are
navigating the changing
information ecosystem,
including how much
misinformation they
think they see and how
they respond.**

Five tribes of misinformation consumer

3 Five tribes of information consumer

Social media platforms enable advertisers to target particular groups. Historically this has included groups interested in conspiracy theories, such as anti-vaccine content, and even those with anti-Semitic beliefs.⁸³ Platforms accumulate detailed profiles of who is targeted and why, but often closely guard this commercial information.

It was our hypothesis that such a process could be done for public good, enhancing the quality of information online. To do this we first wanted to understand how information and misinformation is consumed today, by conducting our own qualitative survey. Through analysis of our polling responses, conducted alongside Kantar Research, The RSA elicited five tribes of (mis)information consumer.

The five tribes are:

- **The Hyperactives:** spread news, real and fake, at pace, without paying huge attention to what they are sharing.
- **The Selectariat:** are the most discerning group of news consumers, and spend more time actively reading a range of sources and fact-checking what they read. This group represent a relatively minor proportion of the population.
- **The Hoaxers:** are those that believe in common conspiracy theories and are more likely to believe content received from friends or family than the rest of the sample.
- **The Traditionalists:** encompass those that get the majority of their news from TV and print media, rather than online.
- **The Disenfranchised:** are characterised by their disdain for authority figures: the media, the government, politicians, and scientists.

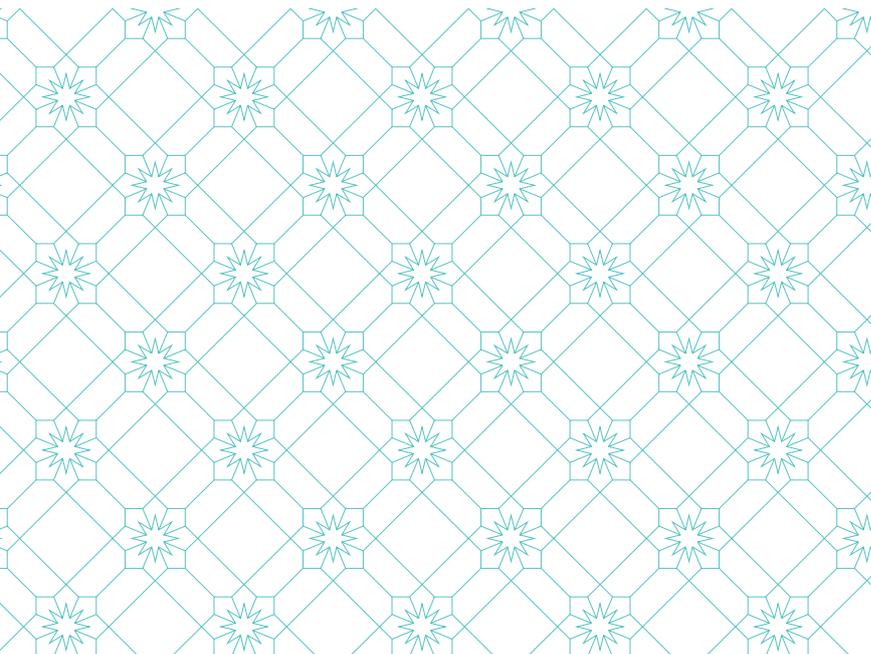
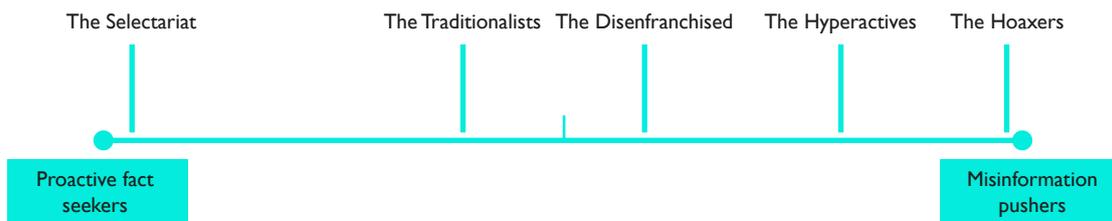
83 See here: Wong, J (2019) Revealed: Facebook Enables Ads To Target Users Interested In 'Vaccine Controversies'. The Guardian [online] 15 February. Available at: www.theguardian.com/technology/2019/feb/15/facebook-anti-vaccination-advertising-targeting-controversy [Accessed 28 April 2020]. Also see: Sankin, A (2020) Want To Find A Misinformed Public? Facebook's Already Done It. The Markup [online]. Available at: www.themarkup.org/coronavirus/2020/04/23/want-to-find-a-misinformed-public-facebooks-already-done-it [Accessed 28 April 2020]. And see: Angwin, J, Varner, M and Tobin, A (2017) Facebook Enabled Advertisers To Reach 'Jew Haters' ProPublica [online] 14 September. Available at: www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters [Accessed 28 April 2020].

Five tribes of information consumer

The groups were constructed by observing patterns of behaviour and attitudes among survey participants, through responses to key research questions. These groups are not mutually exclusive and are not meant to represent the entire range of possible groups. Instead show the five groups felt to be the most influential, those often forgotten in the debate, or those in need of extra research and analysis.

In figure 2 (below) we also place these groups on a scale from 'proactive fact seekers' to 'misinformation pushers'. As can be seen The Selectariat are the group with the most positive information behaviours. The Hoaxers have the most negative in that they regularly seek and purposefully push misinformation, which differs from The Hyperactives who are instead more likely to be, though not exclusively, inadvertent or accidental pushers of misinformation. The Traditionalists and The Disenfranchised are instead relatively neutral on this scale.

Figure 2: The five tribes on scale of information behaviours



Methodology

In 2020 Kantar polled a representative sample of 2,008 people on the topic of misinformation. The questionnaire involved a range of questions on information consumption habits, views on the media environment and regulation. The questionnaire included a section where the public were presented with a selection of real and false headlines and asked to state whether they believed them to be true or not, as well as a more general questionnaire to determine personality traits in the respondents, developed by Kantar.

The questionnaire consisted of four stages:

- 1** Personality profiler: questions regarding participant's habits, outlook, personality traits and views on politics and society.
- 2** Fake news screen: participants were invited to judge the validity and interest they showed on a section of fake news questions, and were queried on whether they believed certain conspiracy theories.
- 3** Behavioural questions: we also asked how the public consumes news, how often, whether this was online or offline, and how for how long.
- 4** Attitudinal questions: how did our participants feel about the news environment and attempts at regulation and/or preventing false information.

Researchers recreated false headlines used in the fake news screen from stories on Full Fact's website at the time of designing the poll in December 2019. The respondents were then asked two sets of questions for each headline. First, they were asked to select if they had either previously seen the headline, would be interested to read more, or are uninterested in the news story. Second, they were asked whether or how much they felt the headline to be true or not: 'I am sure this is true'; 'I think this

story is probably true'; 'I suspect this may not be true'; 'I do not believe this story at all'.

Example of false news stories include:

- Cambridge university banning the wearing of poppies.⁸⁴
- A quote from leaked US and UK trade documents regarding the sale of NHS assets.⁸⁵
- The BBC donating money to the Conservative Party.⁸⁶
- Jeremy Corbyn attending an IRA funeral.⁸⁷

For the true headlines the research team selected a series of true news stories from various reputable news platforms that were specifically not 'headline' news, in order for them to be not too well known. The same questions as above were asked for each respondent.

To analyse the findings, we used Pearson correlation coefficients to identify trends in the data, along with a consideration of specific research questions, such as belief in conspiracy theories, or use of fact-checking platforms. Our overarching findings on attitudes to news, (Key findings) uses data weighted for age and gender. We then singled out certain groups relevant to our analysis. For this stage of the analysis (the Information Ecosystem: Key Players) we used unweighted data, due to a desire to compare smaller, individual subgroups within the data.

84 Krishna, R. (2019). Cambridge University Has Not Banned Poppies Full Fact [online] 17 October. Available at: fullfact.org/online/cambridge-poppies-ban [Accessed 24 June 2020].

85 Full Fact. (2019). This Quote About Selling NHS Assets Does Not Appear In The Leaked UK-US Trade Documents [online]. Available at: fullfact.org/election-2019/nhs-docs-fake-quote [Accessed 24 June 2020].

86 Phillips, T. (2019). The BBC Did Not "Donate" Money To The Conservative Party Full Fact [online] 30 November. Available at: fullfact.org/online/bbc-not-donate-conservative-party [Accessed 24 June 2020].

87 Krishna, R. (2019). This Is Not An Image Of Jeremy Corbyn At An IRA Funeral Full Fact [online] 08 November. Available at: fullfact.org/online/jeremy-corbyn-ira-funeral [Accessed 24 June 2020].



The Hyperactives

The Hyperactives are the superspreaders of online news and information. They like to share news and pay less attention to what they read and share to others and are therefore more likely to inadvertently share misinformation.

12 percentage of sample



The Selectariat

The Selectariat are a small group of highly critical consumers of news and information. They are the only people within the sample to use established fact checkers.

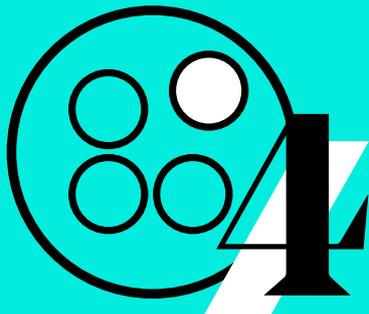
7 percentage of sample



The Hoaxers

The Hoaxers are ardent believers in common modern conspiracy theories: 5G is bad for health, global warming is a hoax, and vaccines are harmful. To counter these theories we must first identify, understand, and then reach out to these groups

6 percentage of sample



The Disenfranchised

The Disenfranchised feel out of step with the general population. They have a strong distrust of 'the media' and feel that no media outlets 'speak for them'. This persona highlight the inherent difficulty with reaching certain critical groups, who are disengaged from mainstream sources of information and resistant to edicts to change.

14 percentage of sample



The Traditionalists

The traditionalists use online media the least. They heavily rely on offline news: TV, print news, and radio. This group, however, are still susceptible to believing disinformation and conspiracies, showing it is not a problem that can just be tackled online, offline solutions are needed too.

11 percentage of sample



The Hyperactives

The Hyperactives are the super spreaders of online news and information. This group are those who said they liked to share news on social media the most.

They also pay less attention to what they read and share, which makes them more likely to inadvertently share misinformation, later discovering it was false.

There is also some evidence to show they are more likely to believe misinformation than the sample as a whole.

Small but significant group: they form **12%** of our sample

More than twice as likely to have said they share news they later found out was false: **27%** compared to **11%** of the whole sample



They are often younger: **36%** are **under 35**, double that of the sample



Very few said they **do not trust** any media

More likely to believe fake headlines shown: **39%** versus **28%** of the sample



More likely to believe conspiracy theories: **41%** versus **28%** of the sample



The Hyperactives

Key variable: most likely to regularly share content on social media.

Misinformation must be shared to be seen and believed. This can happen through bots or fake accounts, but commonly is spread by ordinary people who may be, as we all can, less scrupulous or misinformed at times. But while we can all be guilty of inadvertently sharing misleading information or failing to fully read and understand what we share, our research indicates there is a small but important group who are more likely to do this than others: The Hyperactives.

The Hyperactives, form around 12 percent of our sample, are characterised primarily by their very high frequency of sharing news through social media and chat apps. They are also far more likely to have shared a story they later discovered was false, at 27 percent, compared to 11 percent of the whole sample.

They are more likely to believe the fake stories and conspiracy theories that we presented to them, believing 39 percent of false headlines and 41 percent of the conspiracy theories versus 28 percent of the sample for both.⁸⁸ This figure perhaps also points toward a greater awareness about fake news – this group, while actively sharing false information, have also detected untruths in online content.

They are generally younger than the rest of this sample, with largest age group in this segment 24-35,⁸⁹ and are more likely to use Instagram (56 percent, versus 29 percent across the sample.)

Finally, those who share online news frequently are also likely to trust news sources – very few within this group said that they trusted no media at all.

Hyperactives are of particular interest because, while we are all vulnerable to believing misinformation in one form or another, there is a clear group of social media users who are the most **active** and are therefore more likely to be the sharers of misinformation.

Tackling misinformation online will without question require a detailed analysis and understanding of The Hyperactives, who are the biggest creator and spreader of content – both true and false – we found.

Key insight: we think of the Hyperactives as ‘trigger happy’: folks who are unknowingly spreading fake news online. But could their propensity to act be deployed in aid of taking on the misinformation challenge?

⁸⁸ It should be noted also that these self-reported statistics are likely to be an underestimate, given that we know there is little-to-no feedback loop for those who interact with misinformation.

⁸⁹ 36 percent of Hyperactives are under 35, double that of the sample as a whole.



The Selectariat

The Selectariat are highly critical consumers of news and information. They are selected by their use of established fact checkers, a very rare behaviour among the British people.

This group already have the behaviours that are commonly stated as 'desirable' by many commentators, yet not many people actually do it.

We must find ways to expand this behaviour among the general public.

Very small group: they form just **7%** of our sample.

47% are aged between 35-55, compared to **35%** of the sample



Show a greater concern than the sample for the effect of disinformation on themselves (**36%** against **27%**); on their friends and family (**62%** versus **35%**); and on the quality of democracy (**83%** versus **59%**)

On themselves:



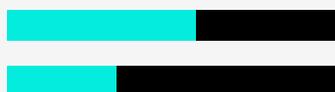
On friends and family:



On quality of democracy:



Far more likely to cross reference news with other sources: **57%** versus **33%** of the whole sample



The Selectariat

Key variable: regular use of established fact-checking platforms, such as Full Fact, FactCheck, or Infotagion.

A very small group of those we questioned, around 7 percent, had used, and could name, such fact-checking platforms. This group, **the Selectariat**, is predominantly middle aged,⁹⁰ and unsurprisingly were more concerned about fake news than the rest of our sample, including the impact it is having on the wellbeing of themselves and their friends and family, as well as the broader impacts on society.⁹¹ They are very active on Twitter (58 percent, versus 29 percent of the whole sample), and prefer the Guardian to tabloid newspapers.⁹²

The Selectariat are also far more likely to cross-reference news articles: 57 percent report checking where a story has been reported elsewhere to check its validity, compared to just 33 percent for the entire sample. They are less likely to be inclined to believe stories from our fake news screen than the rest of the sample.⁹³

They are keen on sharing online, but not significantly more than the sample as a whole: just 11 percent of the Selectariat share as frequently those in our Hyperactive category. This group are also far more likely to care about related issues, such as freedom of the press and speech.⁹⁴ They are relatively clued up on tech, and care about issues such as data privacy.

Solutions using the Selectariat could focus on building the capabilities of this group into one at the vanguard of challenging false information online.

At present, this group highlights the limitations of how fact-checking is currently used as an approach to tackling misinformation, precisely because it is sparingly and inconsistently applied, to the point where very few people use such services.

Key insight: existing solutions (fact-checking platforms) are only used by a small number of news consumers; more should be done to platform those solutions but also those consumers who are civic-minded enough to seek them out.

90 Almost half (47 percent) are between 35-55. For the whole sample, this group is around a third (35 percent).

91 Impact on friends and family, 62 percent vs 35 percent; concern for own wellbeing 36 percent vs 27 percent; concern for quality of democracy, 83 percent vs 59 percent.

92 48 percent, versus 16 percent for the sample as a whole.

93 They believe 25 percent of false headlines, versus 28 percent for the whole sample.

94 Concern for freedom of speech 56 percent vs 47 percent; freedom of the press 40 percent vs 23 percent; data privacy 47 percent vs 43 percent.



The Hoaxers

The Hoaxers are the ardent believers in common modern conspiracy theories: 5G is bad for our health, global warming is a hoax, and vaccines are harmful.

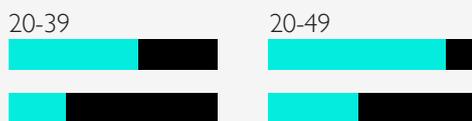
In recent years, the prevalence of these conspiracies has grown among the general public. To counter this phenomenon we must first identify, understand, and then reach out to these groups.

Very small but important group: they are **6%** of our sample.

Highly active on social media: 48% view news on social media daily



More likely to be younger: 62% are between 20-39, and 85% are 20-49 versus 27% and 43% of the sample respectively. More likely to be male



Were **more likely** than the whole sample to believe **fake headlines** presented to them

They are particularly active on **Facebook**: **84%** use it

Somewhat libertarian, 35% strongly agree that government interferes too much in the daily lives of people, versus 10% of sample



The Hoaxers

Key variable: high likelihood to believe common conspiracy theories.

Science scepticism has always been around, and our research is another in a litany of research showing some members of the public were already highly prone to believing in falsehoods regarding, for instance, hoaxes around 5G long before their supposed 'connection' to coronavirus proliferated on social media.

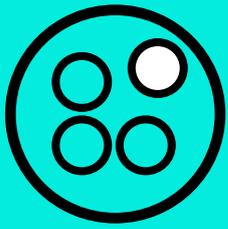
The rise of the anti-vaccination movement, for instance, is well documented; 30 percent of our sample are either inclined to believe, or believe, that the harmful effects of vaccines are hidden from the public. And 18 percent feel the same way about global warming, while 26 percent feel that the introduction of a 5G wireless network will have disastrous impacts on health and the climate.

A small group we call **the Hoaxers** (around 6 percent of the sample) believe or are inclined to believe in all of these anti-scientific conspiracy theories (5G is harmful, global warming is a hoax, and vaccines are harmful). They have been a subject of particular focus recently, with the latest hoaxes regarding vaccines and 5G in a string of anti-scientific conspiracies spreading rapidly on social media.

Our data shows that this group are young, highly active on social media, 48 percent say they see news on social media daily, and they are particularly active on Facebook (84 percent). This group tends to be far more trusting of news shared by their immediate contacts – 60 percent say they are more trusting than distrusting of news shared by their friends on social media, compared to just 20 percent for the sample as a whole.

This group also, perhaps, has certain libertarian inclinations: 35 percent strongly agree that the government interferes too much in our everyday lives, compared to 10 percent of the overall sample. Unsurprisingly, this group score far lower when asked to distinguish between real and false headlines, believing almost half of the stories they were presented with, and they are unlikely to have used fact-checking platforms. This demonstrates how the spread of peer-to-peer misinformation can accelerated on social media.

Key insight: the Hoaxers are highly entrenched and are a challenge to be reckoned with rather than a group to be enjoined in solutions.



The Disenfranchised

The Disenfranchised feel out of step with the general population. They have been selected by their strong distrust of 'the media' and they feel that no media outlets 'speak for them'

The Disenfranchised highlight the inherent difficulty with reaching certain critical groups, who are disengaged from mainstream sources of information and resistant to edicts to change.

From this group we must remember to be highly careful in language and method of communication.

They are the largest subgroup at **14%** of the sample

Less likely to **consume news online**, preferring print media

Demographically **similar** to the population at large

Strong dislike for 'the establishment' eg. government, politicians, the media, scientists.

Scored **better** on the **fake headlines** presented to them

Very **unlikely** to have used a fact checker (just **4%** had used)

No discernible standout characteristics across common **demographic factors**: SEG, age, gender, education, or income.

The Disenfranchised

Key variable: strong distrust for the media; feel that no media outlets speak for them.

A pervasive characteristic of discourse around media in the 21st century, and fake news, is a growing lack of trust in our media ecosystem. A significant section of the population is not engaged with our current media. Around 14 percent of our sample feel that there are no news outlets that represent their views, and that they don't trust the media.

The Disenfranchised have a strong disdain for almost all sections of the establishment they were asked about; government, politicians, media, and scientists.⁹⁵

They are far less likely to consume news online than the rest of our sample, preferring to stick to print media. They

are unlikely to have used a fact-checking platform, at just 4 percent. Nevertheless, when tested with our fake news headlines, this group read, and believe, slightly less fake news stories than groups with other consumption habits.⁹⁶

In general, this group doesn't subscribe to political issues we might call 'identity politics', caring less about LGBT and racial discrimination, for instance, than the general population.

This group highlights the difficulties with building single solutions for our information ecosystem, or with building solutions with an implicit or explicit type of user in mind. There is a group which feels highly alienated from all mainstream (and even what many would call reliable) sources of information.

How can we meaningfully and constructively engage with the roughly one seventh of the population who actively distrusts the sources that are pushed as trustworthy?

Key insight: the Disenfranchised is a significant and diverse group of people who feel out of touch with mainstream culture and news. Yet while they are found in many corners of society, they are united in their distrust for authority figures.

⁹⁵ More distrust than trust the government at 66 percent vs 38 percent; more distrust than trust politicians at 80 percent vs 55 percent; more distrust than trust scientists at 30 percent vs 13 percent.

⁹⁶ They believe 23 percent of false headlines presented to them, versus 28 percent for the whole sample.



The Traditionalists

The Traditionalists use online media the least. They heavily rely on offline news: TV, print news, and radio.

This group, however, are still susceptible to believing disinformation and conspiracies, showing it is not a problem that can just be tackled online, offline solutions are needed too.

Techno-solutionism will have limited effect for this group, 'analogue' approaches will be needed too

Traditionalists are
11%
of our sample

Generally trusting of mainstream media. **34%** say they agree with much of mainstream media outputs, compared to **25%** of the whole sample



This group are older than the average. **66%** are over 55, compared to **56%** of sample, and somewhat more likely to be male



Scored the **same as the whole sample** on their ability to correctly **identify fake headlines**

They are more interested in politics and local heritage, and are more likely to read a local newspaper (**30%** versus **20%** of sample)



The Traditionalists

Key variable: scored high on offline news consumption (TV, print, radio); scored low on online news consumption.

Researchers and commentators can often find themselves working on the implicit assumption that conspiracies and misinformation are a problem exclusively for those who regularly use and consume news from the internet. While, of course, the misinformation we refer to is almost always on social media, it would be a mistake to **only** target the remedies at those who regularly use the internet, because a sizeable proportion of the population will be less engaged with technological solutions. Instead 'analogue' solutions such as information literacy programmes will also be needed.

Our analysis shows those who consume the most media offline, and the least on the web, are no better than the population at large when it comes to believing in fake news and conspiracy theories. We therefore believe that this group are currently under looked and under-served when thinking about misinformation.

The Traditionalists consume most of their news through TV, radio, and print news, rather than online and social media news. They are more likely to be male and 55+.⁹⁷ Unsurprisingly, they are generally trusting of mainstream media across the board – 34 percent report that they agree with much of the output of mainstream outlets, compared to 25 percent for the sample at large. This group score almost identically to the sample as a whole in our fake news screen, demonstrating that vulnerability to, and belief in, fake news is not limited to those who solely consume news online.⁹⁸

Compared with the rest of our sample, this group is more interested in topics like politics, finance, and national heritage,⁹⁹ and are interested in reading local news; 30 percent read a local newspaper, compared to 20 percent of the sample in general. Traditionalists are more likely to discuss news with family or friends compared with the rest of the sample.¹⁰⁰

This group highlights that misinformation is not just an online problem. False information can also be spread by those who spend little time online, and those who consume a limited range of news sources which they trust may be less critical when exposed to new information. Remedies should therefore also be targeted through analogue means such as the provision of information and media literacy to both adults and young people.

Key insight: offline news consumers are vulnerable to misinformation too.

97 57 percent male, and 66 percent over 55, versus 45 percent for the whole sample.

98 This group believe 27 percent of false headlines and 28 percent of conspiracy theories, compared to 28 percent and 28 percent for the sample as a whole.

99 Politics, 64 percent vs 44 percent; finance, 56 percent vs 34 percent; national heritage, 44 percent vs 33 percent.

100 53 percent versus 32 percent.

THE PSYCHOLOGICAL DIMENSION:

INTERVIEWS ON THE MECHANISMS OF MISINFORMATION

The psychological foundations of what we believe and why are complex. We therefore interviewed psychologists and technologists to understand how individuals are driven by, perceive, and are affected by, the misinformation ecosystem. We split these into two kinds of online behaviour: individual and community or group.

Interviews on the mechanisms of misinformation

Individual psychological pathways

Confirmation bias

Confirmation bias is the process whereby individuals are more likely to read and believe information that agrees with their already-held world view or opinion. In modern social media, filter bubbles and echo chambers create environments that allow the confirmation bias to flourish, where we are primarily exposed to, and expose ourselves to, content we are already likely to believe. However, in contradiction to popular belief academic evidence gives relatively little weight to the idea of '*algorithmically* created filter bubbles'¹⁰¹, and actually states that social media use increases our range of news sources, primarily by simply increasing the amount of news consumed by the average user.

For instance, In the UK, conspirators claim there are links between Coronavirus and 5G. The exact details of the claim vary, from 5G harming public health and making people more susceptible to the virus, to the non-existence of Coronavirus and declining public health due to a side effect of 5G. Yet, as Full Fact write, myths on 5G, and myths on mobile signals generally, have existed long before Coronavirus but rumours have found new ears due to the panic and confusion the virus has caused.¹⁰² 5G links to Coronavirus therefore are, for some, a form of confirmation bias, whereby new external events are linked to their pre-existing beliefs about mobile signals.

101 Fletcher, R. (2020). The Truth Behind Filter Bubbles: Bursting Some Myths. Reuters [online]. Available at: reutersinstitute.politics.ox.ac.uk/risj-review/truth-behind-filter-bubbles-bursting-some-myths [Accessed 11 May 2020].

102 Rahman, G. (2020). Here's Where Those 5G And Coronavirus Conspiracy Theories Came From. Full Fact [online] 09 April. Available at: fullfact.org/online/5g-and-coronavirus-conspiracy-theories-came.

Box 4: Expert view

Tali Sharot is a neuroscientist at UCL, and an experimental psychologist known for her research on the neural bases for emotion, decision making, and optimism bias.

One area of work looks at the psychological effect of motivation on information processing. Her research has found that new information is more likely to be integrated if it suggests a positive outcome for the recipient. Imaging studies show that the brain tracks negative information less than positive information as it is thought that people feel negative information applies less to them.

A new area of Tali's work that builds on the phenomenon of confirmation bias – the tendency to accept information that supports one's prior beliefs – looks at the effect of **other people's certainty** in their beliefs on information processing and the likelihood of believing the information. In short, a speaker who agrees with your world view, and is confident in their views, can further harden your beliefs. Whereas a confident speaker you disagree with matters less.

There is a way around this. If we are trying to change someone's fundamental beliefs, says Tali, finding common ground, starting from a point of agreement and building from there can encourage a change of perspective.

From a psychological perspective, if **misinformation goes against what a person believes, or wants to believe, it is unlikely to alter their viewpoint. Whereas misinformation aligned with pre-existing views can entrench their beliefs or lead it to further extremes.** Yet it may also be possible to convince someone of false or misleading information if it contains elements of truth or mutually agreed facts.

A further challenge that is unique to social media is the act of liking or sharing content which, by the mere physical act of 'liking' a user is psychologically committing to a piece of information, further cementing beliefs, and convincing people of their position.

Recent studies suggest that people can be triggered to think more critically about the information they process but they must have reason to do so. Tali suggests that **critical reasoning must be motivated.** For instance, users could be rewarded, or punished or shamed, towards interacting with better information

First mover advantage

A further psychological pathway relates to how humans update their beliefs upon hearing new information. This is a highly complex process we are still only beginning to grasp. Put simply, there is a 'first mover' advantage effect which means that in a drought of information, explanations, or facts that we first hear can become anchored in our beliefs and, even if wrong or later discredited, it can be hard to shift.

For example, within the pandemic a common source of misinformation online was reports of 'home remedies' that can cure the virus. A commonly recurring, and comparatively benign, story was drinking freshly boiled garlic water as a Chinese medicine overnight cure.¹⁰³ More concerning has been recommendations to drink chlorine dioxide, which is industrial bleach.¹⁰⁴

In any case, a drought of information as to what the virus was and how best to deal with it gave space for misinformation to settle, and it can be hard to help people update these beliefs.

Of course, people do update their beliefs all the time, new information can give us new understanding of the world, but how to best enable humans to believe credible information, en masse, is an art not a science.

To understand how humans update beliefs better, we spoke to Stephan Lewandowsky in December 2019, an academic psychologist of Bristol University with a particular expertise in this area.

He described to us that people often struggle to update their beliefs based on new information unless there is a new alternate reality and coherent story associated with it. Simply asking people to forget what they had previously heard, or simply stating it is false, without offering an alternate truth, can cause individuals to still rely on the false information, even though they know it is false.

That said, per Stephan, it is also the case that you are more likely to believe a correction that fits with a pre-existing worldview. If a retraction is opposed to an individual's worldview, however, this may cause a 'backfire effect' whereby belief in the misinformation becomes further entrenched.¹⁰⁵ First mover effects tend to incentivise speed of upload over veracity, an unfortunate hallmark of the web 2.0 media age.

The first mover advantage reminds us that it is critical that citizens are kept informed, and that the logic behind changes to public information are made transparent. For instance, the UK government has not updated the list of main symptoms of Covid-19, despite studies by ZOE indicating they had changed, and other national bodies, such as the Centers for Disease Control and Prevention (CDC), updating their symptoms list.¹⁰⁶

103 Spencer, S. (2020). Fake Coronavirus Cures, Part 2: Garlic Isn't A 'Cure' FactCheck [online] 11 February. Available at: www.factcheck.org/2020/02/fake-coronavirus-cures-part-2-garlic-isnt-a-cure.

104 Spencer, S. (2020). Fake Coronavirus Cures, Part 1: MMS Is Industrial Bleach FactCheck [online] 11 February. Available at: www.factcheck.org/2020/02/fake-coronavirus-cures-part-1-mms-is-industrial-bleach.

105 Lewandowsky, S, Ecker, UK, Seifert, CM, Schwarz, N and Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest*, 13(3), pp106-131.

106 See: ZOE (2021) What are the new top 5 COVID symptoms? [online] Available at: covid.joinzoe.com/post/new-top-5-covid-symptoms [Accessed 24 August 2021] Also see: Centers for Disease Control and Prevention (2021) Coronavirus Disease 2019 (COVID-19) – Symptoms [online] Available at: www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html [Accessed 24 August 2021].

Box 5: Expert view

Stephan Lewandowsky is a cognitive scientist based at the University of Bristol, where he is the chair of cognitive psychology at the School of Psychological Science.

An overarching theme of Stephan's work was that studies on basic cognition showed that **people struggle to update their memories or beliefs about an event**, even upon finding out that an element of the memory or event was untrue.

Stephan gave the example from his study, in which participants were told incomplete information about a fire, which appeared to point to negligence as the cause (for instance hearing on a police radio that a wiring cabinet is full of flammable liquid), a few moments later this is stated as false – there was nothing in the wiring cabinet. In the mind of individuals, oil fire can still appear to be a primary reason for the fire, even though they know and state there was nothing in the cabinet. That is until another narrative for why the fire started is given – for instance petrol-soaked rags were found elsewhere.

In effect, instead of simply stating something as 'false' it is more effective to give an alternate compelling narrative to update beliefs.

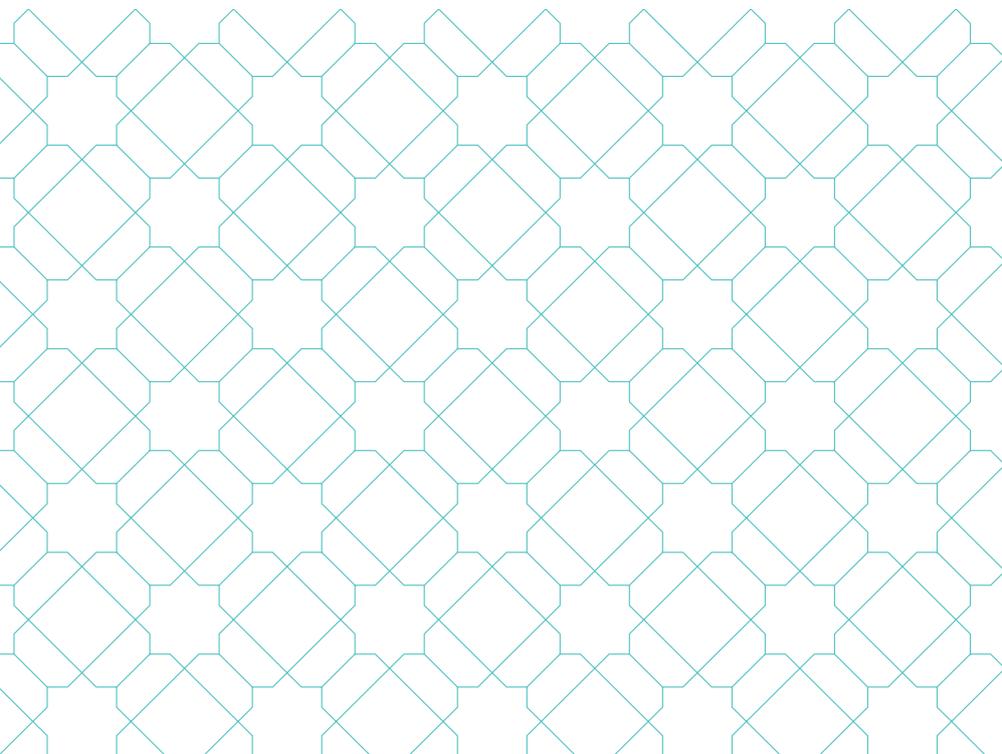
Reimagining this study on political views adds another layer of complexity as people are more emotionally committed to their political position. Thinking in terms of fact-checking, which is often purported as a solution to misinformation and disinformation, **it is not enough to falsify someone's political stance, a viable alternative must also be presented.** It is also noteworthy in more recent studies in the USA with Donald Trump and Bernie Sanders, participants were able to update their beliefs based on specific statements about each candidate, but it did not lead to a shift in attitudes towards either candidate.

We later recommend that users who have interacted or engaged with known misinformation should be **retrospectively** warned of this. What's more, we propose a regulatory body which would allow for best knowledge and practice, such as ensuring a compelling alternative narrative, are shared across the industry.

The illusory truth effect

Other vestiges of individual psychological pathways to beliefs are repeatedly seen on platforms. The 'illusory truth effect' can be more commonly summarised as 'repeating a falsehood until it becomes true', ie if a person hears the same (false) facts often, or often hear doubt poured on known truths, they are more likely to believe the falsehood or think it has some validity. Curated social media content can link people to content (misleading or false) that correspond to pre-existing beliefs, thus entrenching viewpoints that may be misinformed.

Equally, the sheer volume of content online, a product of the billions of users on social media platforms, and functions such as retweeting means that content will inevitably be repeated many times on an individual feed. Yet we also know that through content moderation, fact-checking, and control of algorithms, particular pieces of content will be viewed less often. Avaaz shows that this content and algorithm moderation can be slow and inconsistent, even on the singular platform Facebook.¹⁰⁷ Comparisons across platforms are likely worse. In recommendation 1 we detail a proposed independent body who would set cross-industry minimum standards on dealing with flagged content, taking inspiration from global examples we will see in chapter 3.



¹⁰⁷ Avaaz (2020) How Facebook can Flatten the Curve of the Coronavirus Infodemic. Op cit.

Group behaviour

In addition to misinformation being entrenched by individual weaknesses, group behaviours based around modern social media ecosystems can further entrench misinformation. Various phenomena are a direct result of the capacity of platforms not only to reach individuals but bring groups together into walled gardens of information and common cause. Again, there are myriad pathways. **Echo chambers** involve the purposeful, or inadvertent, creation of online spaces where individuals of similar viewpoints only hear each other speak. This, therefore, creates communities that harden each other's views and see their own views as having greater legitimacy than they may have in wider society. Growing right wing militias in the USA are an example of how closed communities online, even on the biggest platforms like Facebook, can seed growing extremism.¹⁰⁸

A well-known problem is the proliferation of conspiracy theory groups online. The past year and a half, as well as the several years preceding the pandemic, has seen a resurgence in anti-vaccination movements, 5G conspiracists, flat earthers, and more. These groups pose a variety of challenges. For instance, there is already a highly interventionist approach by internet companies on anti-vaccination movements, eg in promoting trustworthy content and pushing down disreputable sources, yet the problem grows. Also, there have been claims that YouTube and its recommendation algorithms was a direct cause of the growth in the flat earther movement.¹⁰⁹

A common theme that emerges from the above – the false sense of certitude and confidence induced by fake news, buttressed by technology, informed by groups of similarly misinformed, is of collective mania, breakdown of social trust, supplemented by revelations of genuine malpractice that augur declining trust in authority and wider social trust – whether deserved or otherwise. False equivalence between causes and groups reigns, thus allowing unreliable or extremist information a seedbed in which to flourish.

This, then, is the difficult context for those who seek to respond to misinformation crises at a governmental level, and in the field. Their efforts are the subject of the next chapter.

108 Beckett, L (2020) Michigan terror plot: why rightwing extremists are thriving on Facebook. The Guardian [online] Available at: www.theguardian.com/technology/2020/oct/09/facebook-rightwing-extremists-michigan-plot-militia-boogaloo [Accessed 25 August 2021].

109 Newton, C. (2020). How Youtube Helps Flat-Earthers Organize The Verge [online] 20 February. Available at: www.theverge.com/interface/2019/2/20/18232524/youtube-flat-earth-recommendation-algorithm-conspiracy [Accessed 12 May 2020].

For 260 years, the RSA's approach to progress and social impact has depended on aggregating information, reasoned debate, convening perspectives and resolving through conversation, and empathetic social design.

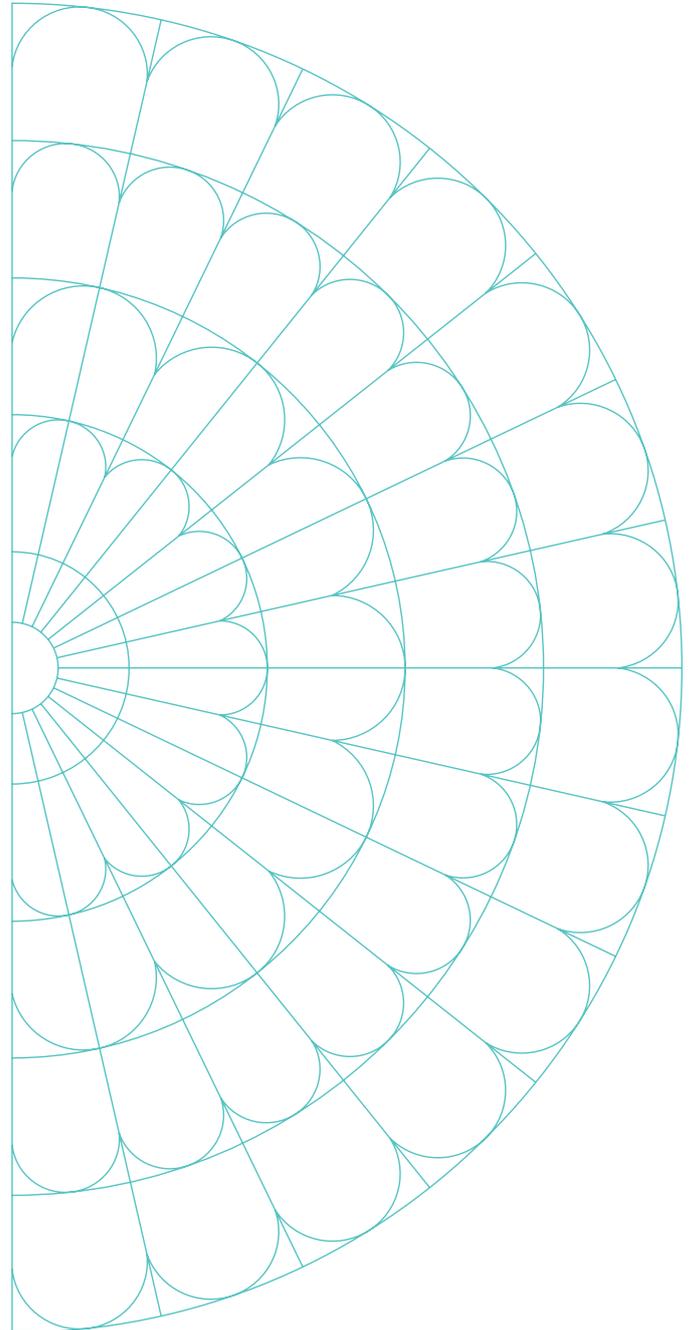
Misinformation increasingly threatens all these when the parameters of articulated truths themselves are at hazard.

Combating Online Harassment: Recommendations

As we saw in chapter 2, governments from around the world have sought various ways and means of combatting misinformation and disinformation. The UK government's response has been the Online Safety Bill. Broadly we think this Bill is a step in the right direction and will do an effective job at limiting the spread and impact of illegal harms online. However, we are concerned that it does not do enough to tackle other online harms and, in particular, misinformation and disinformation. As we explained earlier, the Bill does not attempt to cover societal or democratic harms. It also says nothing towards instances where there are systematic and organised attempts to mislead the public, even when this can lead to harms to *individuals* such as through public health impacts, to national security, or to democracy.

We also saw the inconsistencies within the current self-regulation regime. Where platforms are not held to account for their practices or guided on what constitutes best practice within a country. The results are untimely and inadequate responses which fail to neither prevent nor heal the damage from misinformation.

We expand upon these ideas over the remainder of this report.



Summary of recommendations

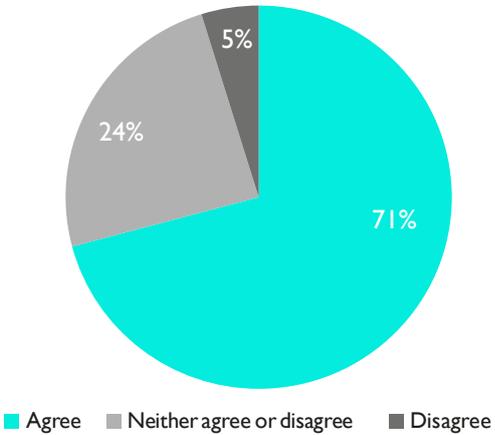
- 1. The draft Online Safety Bill should explicitly include societal harms caused by misinformation within its remit. We recommend this is done through a pluralist body separate from Ofcom – the Office for Public Harms.** The Office would have the responsibility to investigate societal harms caused by misinformation and disinformation. It would then publish its findings publicly, as well as warn platforms of issues it finds and offer advice to Ofcom as to means of addressing the issues. The Office should also act as a misinformation ombudsman whereby it would investigate cases of societal or individual harms brought to it, or would investigate where there has felt to have been overreach into the right to free expression. The Office would then state the expected redress. We believe the new body should be made up of a pluralist panel of stakeholders including citizens, Ofcom, platforms and wider industry, traditional media, civil society, researchers, and other experts. Such a panel would have greater legitimacy than the current heavily platform-controlled information online.
- 2. A ‘polluter pays’ levy of social media firms to counter misinformation.** The levy would be used to fund the working of the Office for Public Harms (see recommendation 1), in a similar arrangement to the current funding model for the Advertising Standards Authority. The levy should also be used to fund media and information literacy drives in the UK, an innovation fund to research best practice in overcoming the harms caused by misinformation, and other further research. Other organisations, such as Glitch, have called for a tax on social media firms equivalent to 10 percent of the recent Digital Services Tax – which itself is set at 2 percent of revenue of particular internet firms.
- 3. Track and Trace system for fake news.** Over the course of our research, we uncovered multiple means of countering misinformation. We believe some of these should form part of the codes of conduct to be written by Ofcom (as stipulated within the Online Safety Bill). We have dubbed these a ‘Track and Trace system for fake news’. These include:
 - i. Corrections of false or misleading content online should be published within 24 hours of the content being released by the relevant fact-checking agency.
 - ii. Proportionate push notifications to users who have seen or interacted with misinformation should be used wherever possible. By proportionate we mean that those who have engaged more directly should receive stronger messages than those who passively scrolled past.
 - iii. Accounts and users who regularly share misinformation, including notable public figures, should be removed.
 - iv. Comprehensive data on the reach and engagement with misinformation, and subsequent engagement with corrections, should be made publicly available for researchers and organisations to improve their messaging or services.
- 4. A citizens’ convention on misinformation: online rights, freedom, and the right to be informed.** To guide the work of government and of the Office for Public Harms, a citizens’ convention on misinformation should be called. When asked the right questions it is our belief that citizens can meaningfully engage and add to the debates between freedom of expression, harms caused by misinformation and the right to accurate information

Recommendation 1:

The draft Online Safety Bill should explicitly include societal harms caused by misinformation within its remit. We recommend this is done through a pluralist body separate from Ofcom

RSA polling from before the pandemic shows that 71 percent of the public stated that they want a 'stronger independent regulator *on the quality of news*'.¹¹⁰ This statistic should of course be taken with much caution, not least because a pandemic has occurred since, but also because the public will have highly varying views on what a 'strong independent regulator' means. Nevertheless, this speaks to general anxiety about the quality of our information eco-system and our insufficient mechanisms of redress for rightly concerned citizens.

Figure 3: RSA polling of the public in January 2020. Question: 'To what extent do you agree or disagree with the following statement ... We need a stronger independent regulator on the quality of news'



¹¹⁰ Emphasis has been added.

Combatting online harm: Recommendations

The current proposal for combatting this anxiety within the draft Online Safety Bill is the creation of an advisory committee on misinformation and disinformation, which would report to Ofcom, the regulator of online safety. The committee has the responsibility of advising Ofcom on how regulated services should deal with misinformation and disinformation. We note however that alongside the overarching layer of online harms architecture there are still significant gaps. Little thought has been given to how regulation can improve the confidence of the public in our information ecosystem, or veracity of information online, cope with **collective** harms, or offer means of redress for those who feel they have either been unfairly moderated or have come to harm due to misinformation.

This is hardly surprising. There is plenty to be concerned about the idea of government bureaucracies having sign-off on swathes of our social interactions and how we express ourselves. It has also been noted by Carnegie UK and others that the current Online Safety Bill gives a great deal of power to the Secretary of State.¹¹¹

Yet, we have also seen throughout this research the harms that can come to individuals and to the collective from misinformation and disinformation. It is our view that more robust action on misinformation must be done, while of course maintaining the right to free expression as enshrined in the Bill.

We therefore suggest that Ofcom and its subsidiaries has explicit remit to assess and address collective harms, as well as harms to individuals.

By collective harms we mean harm to important societal goals such as trust in medicine and science, public health, trust in the electoral system, or social trust and cohesion.

But we also believe that the governance of individual or societal harms caused by misinformation and disinformation should not be left just to platforms, but should be part of a shared regulatory framework.

We therefore suggest the creation of an independent body, the Office for Public Harms (OPH). This body would be made up of a pluralist panel including citizens, Ofcom, platforms and wider industry, traditional media, civil society, researchers, and other experts. It would supersede the advisory committee as proposed in the Online Safety Bill. We believe an independent and pluralist body is needed because it would offer the fairest means of enshrining freedom of expression, while limiting the amplification of known misinformation and disinformation. It is also a more appropriate means of improving the quality of information online than is currently suggested in the Bill because it offers a multi-stakeholder backstop to an otherwise primarily platform-controlled online information ecosystem.

¹¹¹ Carnegie UK (2021) The Draft Online Safety Bill: Carnegie UK Trust initial analysis. Carnegie UK. Op cit.

Our proposal builds on the strength of institutions such as fact-checking organisations and learns from notable other successful institutions such as the Advertising Standards Authority (see below). We also wish to build upon the ideas of Carnegie UK who previously proposed a system of ‘interlocking regulation’ which utilises a ‘subject expert’ regulator to sit alongside a central regulator - Ofcom.¹¹²

Within an interlocking regulation regime, the government’s online harms regulator, Ofcom, will work with specialist, subject expert, third party regulators, who are experts within the area of the online harm in question. For instance, if thinking about online fraud, bodies such as Trading Standards services already have specialist regulatory expertise. In this instance the specialist third party would be the newly formed Office for Public Harms. Ofcom’s role would be to investigate systemic issues, for instance, a failure to prevent harm through systemic aspects, such as platform design, while the specialist regulator (OPH) will have the powers to investigate and remediate individual cases of harm referred to it. The Office for Public Harms can also refer evidence on the nature and scale of the harm, and the systemic factors contributing to it, to Ofcom to take forward with the platforms.

We propose the **Office for Public Harms** would have responsibilities to:

- Investigate and analyse societal harms caused by misinformation. It would do this through transparency reports, information requests, and through harms being submitted to it by the public or by organisations. The Office would then publish its findings publicly, inform platforms of issues it finds, and advise Ofcom on potential changes to the procedural and systemic factors within Ofcom’s remit.
- Act as a misinformation ombudsman. Deal with user complaints and redress when there are individual or societal cases of harm caused by misinformation or disinformation, or, conversely, where content has been felt to be unfairly removed. This process should, in principle, not seek to remove content but only alter its algorithmic reach. It would also only occur when the platform’s own redress procedures were felt to be unsatisfactory.

¹¹² Perrin, W and Woods, L. 2020. Online Harms - Interlocking Regulation. [online] Carnegie UK Trust. Available at: www.carnegieuktrust.org.uk/blog/online-harms-interlocking-regulation [Accessed 30 September 2020].

Combatting online harm: Recommendations

The Office for Public Harms would also:

- Be formed of a pluralist panel of stakeholders, including citizens, platforms and wider industry, media outlets, representatives of Ofcom, but also civil society institutions, including fact-checkers, charities, and researchers. These groups should work together to enhance the legitimacy and decision making of **the Office**. Below we also propose a citizens' jury to support this work and to provide further critical legitimacy.
- **The Office** should particularly act as a voice and supporter of the work of independent fact-checkers, such as Full Fact, FactCheck, or Infotagion. Our research shows that they remain underutilised, and little known by the public, despite the critical work they do.
- **The Office** should conduct independent research on how providers could better prevent potential individual and societal harms, for instance by learning from global best practice, and by investigating messaging and communications to maximise public trust.
- Finally, **the Office** should advise Ofcom and relevant government departments, such as the Department for Education, on their respective media and information literacy programmes in relation to misinformation and disinformation.

Within this system of interlocking regulation, **Ofcom** would have the responsibility to:

- Investigate and remediate systemic issues within the sector. Such as the use and design of algorithms in promoting or dealing with misinformation. This would be for both individual and collective harms.
- Set cross-industry minimum codes of practice in dealing with misinformation and disinformation, as is within the Bill already. See recommendation three for ideas on what could be included within the codes of practice.

The Office for Public Harms could be funded and managed under similar arrangements to the Advertising Standards Authority, who also set minimum standards for content veracity. With the ASA, advertisers (in this case, online services providers) pay a levy in towards the set up and running of the independent institution (the OPH).

Recommendation 2:

A ‘polluter pays’ levy to combat misinformation, disinformation, and other online harms

Misinformation has a quantifiable cost. It has a direct effect on the health of our democracy and in light of the most recent coronavirus-related infodemic, the health of our citizens too.

We believe that taxes on online platforms above a certain, sizeable user base must proliferate as these platforms play an ever greater role in our lives. With platforms often arranging their affairs to avoid taxes, there are real questions about the strength of our international institutions and multilateral arrangements.¹¹³ Nevertheless, this should not quell our confidence in articulating an appropriate structure for the taxation of platforms, along with additional necessary enforcement agencies.

We consider a significant *online harms levy* to be a proportionate response to the negative externalities caused by misinformation and other harms online. This proposal would fit into a polluter pays principle.

Glitch, the organisation dedicated to action against online abuse, have also made such calls, arguing that just 10 percent of the recent digital services tax (which is set at 2 percent of revenue of certain internet firms) could be used to empower civil society, educate the public, and enforce legislation on online hate.¹¹⁴

We believe such a levy would specifically address the following:

- **Funding the Office for Public Harms:** ensuring there is effective remediation for users and that the public square is protected from malign influence.
- **Media and information literacy:** the personality archetypes we identified through UK public polling help indicate that all citizens would benefit from modern media and information literacy. Media discernment could be improved across all groups. Following in the footsteps of other states, such as Sweden and Finland, we propose the levy is used to improve digital literacy proficiency nationally.
- **Innovation fund:** to act an incubator for civic and tech ideas that combat misinformation, matched by UK Research and Innovation (UKRI).
- **Further research:** the tech sector is constantly evolving, as is the content and form of misinformation. We must remain ever vigilant.

The principle here must be simple: platforms and government have a duty, not just to make money and regulate respectively, but to empower civic action. They should empower digital literacy and agency in an era when such agency is elusive, largely as a result of their collusion. One specific example of business and government empowering civil society is in the realm of fact-checking.

113 Sweney, M (2019) Facebook paid just £28m tax after record £1.6bn revenues in UK. [online] the Guardian. Available at: www.theguardian.com/technology/2019/oct/11/facebook-paid-just-28m-on-record-16bn-earnings-in-the-uk [Accessed 31 August 2021].

114 Fixtheglitch.org. n.d. Tech Tax Campaign. [online] Available at: www.fixtheglitch.org/tech-tax-campaign [Accessed 4 November 2020].

Recommendation 3:

A Track and Trace system for fake news

Over the course of our research, we uncovered a number of specific codes of practice that were identified as ideas for social media platforms to implement. In the course of our prototyping sessions, we brought these together as a single design idea: the implementation of a **Track and Trace system** for misinformation. Ultimately these would be the decision of Ofcom, under their powers to set codes of practice within the Bill, but our suggestions are:

- Social media firms should be compelled to publish corrections within a maximum of 24 hours of them being released by the relevant fact-checking agency. Though in most cases it should be significantly less.
- Proportionate corrections should be pushed via notifications to all users who interacted or viewed with the original false or misleading content. By proportionate we mean that the type of notification, and advice or information it gives, should be determined by the nature of the interaction with the post. Those who actively sharing misinformation should not receive the same correction as those who idly scroll past it. However, even those who merely scroll past have a right to know that users on their timeline have shared, or continue to share, misinformation.
- Accounts or users, including those of public figures, that regularly share or post content that requires takedowns should be removed from the platform and without exception 'quarantined'.
- Comprehensive data on the reach and engagement with misinformation, and subsequent corrections by fact-checkers, should be made public for researchers to better understand the state of misinformation and the best ways of using fact-checking (and other means) to counter it.

Recommendation 4:

A citizens' convention of misinformation: online rights, freedom, and the right to be informed

Misinformation is a threat to our social and democratic institutions; to governments, to the work of civic organisations; to any who believe that social betterment and progress emerges from shared ideas, and rational debate and ultimately to our individual and collective wellbeing.

Yet, when considering remedies, including those we describe above, there are clearly trade-offs: between free speech and free thought, against the responsibility inherent within citizenship not to cause harm; between the right to be informed and privacy of individuals, even highly public figures; or on the occasionally shifting sands between heresy and orthodoxy, as with the early and evolving medical advice surrounding coronavirus.

For example, free speech is fundamental to a diverse public square, but freedom of speech can also be weaponised. Online platforms can be used by bad faith actors broadcasting fringe and extremist views, leaning of the sacrosanctity of 'freedom of speech' to defend and maintain their position. The landmark case of *Schenck v United States* is often quoted, where the defendant distributed flyers claiming that compulsory enlistment of draft-age men violated the First Amendment. Justice Oliver Wendell Holmes stated: "The most stringent protection of free speech would not protect a man in falsely shouting fire in a theatre and causing a panic"¹¹⁵ Today we are barraged with false claims of fire.

Recent attacks on 5G masts and engineers were a direct consequence of misinformation about coronavirus.¹¹⁶

Who is to be held accountable when misinformation has real world consequences for the health and safety of citizens, democracy, and society? If leafleting is likened to shouting in a theatre, with the advent of internet, digital platforms can act as a megaphone for bad faith actors on a global stage.

Often the public can clearly see the complications and fault-lines, consistently alluding to them when asked the right questions. Demos polling and focus groups showed that the public have an intuitive understanding of the moral dilemmas, though that view is amenable to discussion.¹¹⁷

There is then, energy for conversation about the kind of rights and freedoms we expect online, their scope and their remit. Might we rescind the right to freedom of speech in favour of 'the right to be informed'? A series of citizens' deliberations should surface that debate and bring new perspectives into the debate about the governance and stewardship of our shared online spaces. It would also offer crucial guidance and legitimacy for Ofcom and the Office for Public Harms.

115 Larson, Carlton FW. (2015) "Shouting 'Fire' in a Theatre": The Life and Times of Constitutional Law's Most Enduring Analogy. *William & Mary Bill of Rights Journal* 24:1 pp181-212.

116 Hamilton, I. (2020). Vandals Set 50 Cellphone Masts In The UK On Fire Because Of A Conspiracy Theory Linking The Coronavirus With 5G Business Insider [online] 06 May. Available at: www.businessinsider.com/attacks-cellphone-towers-coronavirus-5g-conspiracy-2020-4?r=US&IR=T.

117 Demos (2020) Online Harms: A snapshot of public opinion [PDF] London: Demos. Available here: demos.co.uk/wp-content/uploads/2020/10/Online-Harms-A-Snapshot-of-Public-Opinion-1.pdf [Accessed: 24 Aug 2021].

CONCLUSION

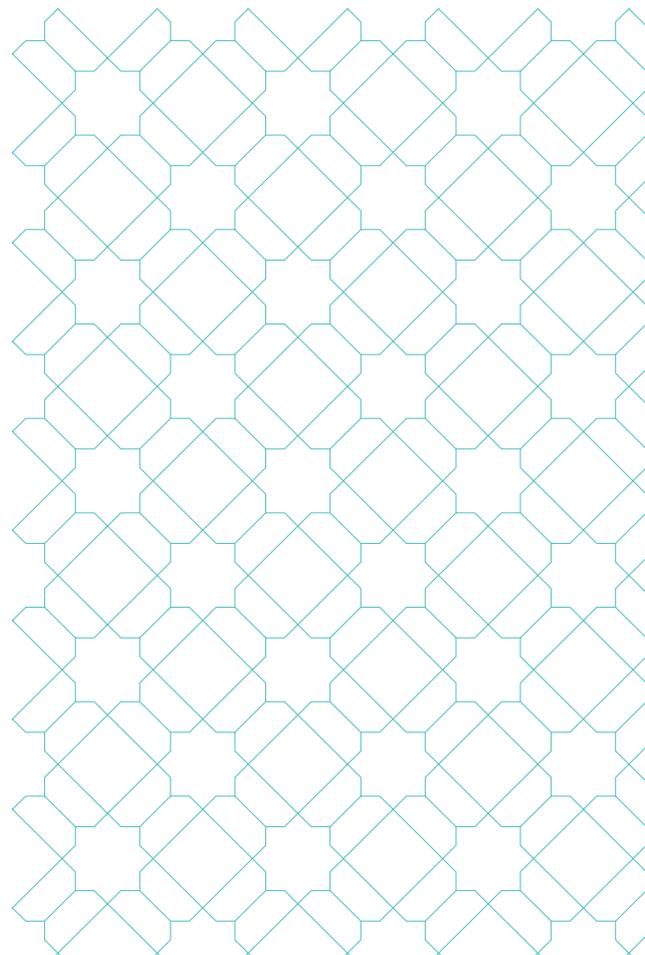
6 Conclusion

The pandemic has been an unfortunate reminder of the power of bad faith actors, using poorly regulated but hyper connected social media, to become amplifiers and vectors of misinformation. Even relatively mainstream public figures, not just grey-and-black market 'shock-jocks', can inadvertently fall into the trap of sharing false or misleading hype stories.

Reform and tight regulation of online harms, including both harms to individuals and societal harms, is needed. We believe our recommendations would support this cause. Yet there is also something in the way that complex and technical information is understood and shared which is broken. This needs to be investigated and our educational and institutional practices should be reformed, with much greater emphasis and drive towards information and media literacy for all ages. This challenge sits behind the recommendations in this piece. We believe much can be learned across platforms by creating a central body, the Office for Public Harms, which acts as both an independent watchdog and a means of sharing best practice.

An Office for Public Harms or a citizens' convention may begin the process for a cleaner public square, but further work in this series should examine these issues in more detail. Fail to progress this work,

to realise the social and institutional nature of misinformation spread, and the routes to a better information ecosystem are narrowed. This report is a taxonomy and a starting point of a civic response to the epochal challenge of misinformation. But the harm of misinformation is not merely personal but social and democratic.



The RSA (royal society for the encouragement of arts, manufactures and commerce) believes in a world where everyone is able to participate in creating a better future. Through our ideas, research and a 30,000 strong Fellowship we are a global community of proactive problem solvers. Uniting people and ideas to resolve the challenges of our time.



8 John Adam Street
London WC2N 6EZ
+44 (0)20 7930 5115

Registered as a charity
in England and Wales
no. 212424

Copyright © RSA 2021

www.thersa.org

ISBN 978-1-911532-72-9