Weaponizing the Internet

Version 1.0 March 19, 2023

If you spot typos, or have feedback to share, please send to alexbr4cornwall@gmail.com Thanks!

In the early 1990s, I co-authored a document, "Demystifying the Internet". It was an exciting time; we had the first truly global medium, with all the possibilities it heralded. Surely, good would win over evil, and the internet would bring democracy to all corners of the world. Thirty years later, this essay describes how authoritarian regimes and bad actors are winning in the clash of global systems.

Bad actors - whether authoritarian regimes, alt-right ideologists, nationalists, religious fundamentalists, those who feel threatened by a move to a more inclusive global liberal world view, or others determined to wreak havoc - are free to use social media and related technologies to meet their destructive goals.

They use traditional marketing tactics and "black hat" techniques (fake news for example), while exploiting weaknesses evident in human behaviour, as well as the algorithms that social media platforms deploy that reward engagement. The anonymous nature of the digital environment also facilitates troll behaviour.

This essay has five parts. In part 1, we look at how traditional marketing tactics have been deployed in the context of an information war highlighting the Cambridge Analytica case. In part 2, we explore the implications of the new media landscape, the information bubbles it creates, and the popularity of memes for sharing content. Part 3 explores human behaviour, how the online world alters our innate instincts, and makes us more prone to believing and promoting conspiracy theories. Part 4 explores the social media landscape and the perverse impacts from the platforms' desires to keep users engaged. In Part 5, we summarize the current state, look at the future direction of technology, and examine four potential solutions to tackle these concerns.

Part 1: Marketing Tactics in an Information War

In the first instance, the tactics adopted by bad actors come straight out of a traditional marketing playbook – target your audiences and promote tailored content that yields action. Let's begin with a well-known case study, Cambridge Analytica (CA), and how this organization manipulated the Brexit referendum in the U.K. and the Trump election in 2016; both these outcomes were desired by Russia:

1. **Identify targets**. Cambridge Analytica essentially stole customer data from Facebook profiles. The data they acquired was based on personality profiling (introvert versus extrovert, for example), as well as psychographics (including consumers' attitudes and beliefs, inferred from Facebook behaviour). They combined this with geo-targeted data, demographic data, and data from other sources outside of Facebook (from Experian, from voter records, and more). CA and other digital marketers know that targeting is key to presenting the right content to the people that they want to convince, enrage or silence with their messaging. Adding to the mechanisms of targeting, social media platforms exacerbate this issue by helping content find its audience through their algorithms, powered by artificial intelligence/machine learning.
2. **Content creation.** Once targets are identified, content is created and targeted to custom recipients on a granular level. This content might well be **fake news** or any material that is manipulative but resonates with its targets; it might be selectively curated real news that

resonates with audiences, or exaggerated news or distorted facts that resonate. This content is designed to cause outrage and fear, leading to increased engagement with the content which is then amplified by social media algorithms. The content also creates confusion and alters its recipients' understanding of reality.

During political elections, there are essentially three goals for these tactics: to *persuade* current voters of a point of view, to *suppress or increase* voter turnout of specific target groups (for example, the youth vote), or to simply *wreak havoc* in society to undermine democracy.

Outside of elections, these same tactics are deployed on issues such as global warming and climate change denial, COVID 19 and government mandates for control, and the next existential crisis.

Trolling also occurs to target and demonize those who threaten the perceived status quo of the dominant culture (white, male, Anglo Saxon) and those in authority.

This is all designed to wreak havoc, cause chaos, doubt, confusion and exhaustion, which destabilises and reduces our certainty of reality. Gaslighting, a term that won 2022 word of the year from the Merriam-Webster dictionary, describes this behaviour. The goal is to essentially turn citizens on one another - all while reducing our capacity, opportunity and willingness to engage in meaningful debate.

In the Cambridge Analytica case, they were essentially manipulating human behaviour for a desired outcome (which is what all marketers do, actually), but they were using information warfare tactics, - psyops.

For Cambridge Analytica, and disinformation campaigns in general, this is simply **marketing 101**: identify your audience at a granular level (micro-targeting) and create content that will be appealing or threatening to your audience to nudge them to take action (or inaction). They are using persuasion to get their audience to trust them. The ultimate goal is to develop an audience they've shaped that then blindly consumes everything they serve later. All the while, this erects barriers to critical thinking. They are also using another standard marketing tactic – repetition. If you repeat something often enough, people start to believe it.

Cambridge Analytica is just one case study we will focus on, in this essay, to help describe the tactics deployed in this information battle. We will also be looking at conspiracy theories like Pizzagate and QAnon and why they are followed and how they spread, as well as the appalling Gamergate conspiracy. We will also do a deep dive into why we have become so vulnerable to disinformation and online warfare tactics.

Somewhat of a precursor to this current state of the internet, a TV example of this, was the emergence of Fox News in the late 1990s. Fox News tells its viewers repeatedly that Fox is the authentic alternative to the "fake news media" and the "liberal media", as they describe their competitors. Some regular viewers of Fox thus blindly follow the network's ideas and conspiracies under the false assumption that if they don't, they will be manipulated by mainstream media. Fox News's goal has never been to report the news accurately. It seeks to be the loudest among the conservative voices.

Of course, Fox News and its shenanigans are very much in the public domain in the United States. Social media manipulation is much less so, and therefore less trackable, more covert and more insidious; essentially, we cannot easily observe the damage wrought by social media disinformation and trolling.

It's our data, and their stories are false

There are three big issues with this information battle. Firstly, the data that Cambridge Analytica used to create audiences for targeting is our personal data; we do not provide *informed consent* for the use of this data. Provision of informed consent underpins privacy legislation like the GDPR (EU) and the CCPA (California).

We have essentially become the product, for sale to the advertisers. This is the dominant business model of the social media platforms; for example Meta, which owns Facebook, Instagram and WhatsApp, relies on advertising for a whopping 97% of its revenue.

Secondly, the content that appeals to the audience is often fake news content designed to trigger emotional reactions. This manufactured content is generally protected by freedom of speech laws; in most of the Western world, we are free to say anything, true or not, provided it's not hate speech. These laws were originally designed to protect speech in the public square from the influence and manipulation of big government.

Finally, trolling behaviour is an integral part of this information battle. Trolls further amplify falsified content with fake accounts, and trolls raid the comments of those who challenge the content with their counter-narrative, in order to silence them, or overwhelm their voice.

Part 2: The New Media Landscape

New media foment the narrative and support communities

Different media have varying levels of rules regarding freedom of speech, including online platforms, newspapers and TV stations. In the virtual world, we see this with more permissive platforms like Reddit and 4chan. These non-restrictive portals combine with platforms like Rumble and YouTube (and networks like its Alternative Influence Network (AIN)) to help incubate, foster, and develop storylines that often originate from sites like Breitbart News and InfoWars. These conspiracies then spread to other, more visible platforms like Facebook, Instagram and Twitter; this process is known as normiefication: the content becomes visible to "normies".

Platforms like Reddit and YouTube serve as a bridge between 4chan and social media channels like Facebook, Instagram and Twitter. Conspiracy theories like QAnon and Pizzagate started on 4chan, jumped to other platforms, and ultimately emerged on traditional social media channels and offline – that's normiefication.

The media landscape

Before examining more closely the detail of the media that propagate conspiracies and fake news, it is worth looking at the spectrum of media, across the political landscape and consider the differences between the media on the left of the political spectrum and the right. Here, we are focused on the United States.

The left, or democrats, tend to get their mainstream news from mainstream media like the New York Times, Washington Post and CNN. Further left there are sites like the Huffington Post and Daily Beast, but readers of those sites tend also to be exposed to, and read mainstream media. There is a "reality-check dynamic" at play, that tends to promote truth-seeking in their reporting. Bias here relates to stories that they run, rather than the truth in those stories.

The right of the political spectrum's centre of gravity aggregates around Fox News, Newsmax and Aon and alt-right sites like Breitbart News and InfoWars. These readers and viewers are effectively disconnected to the mainstream media, and much less likely to read the Wall Street Journal, the right's mainstream news source. Truth-seeking is not the driving motivation of these audiences, and the media's goals focus on ratings and providing content the audience is seeking to confirm their identity, rather than on truth-seeking. The media's goal is to appease and grow their audiences. Another aspect to this is that they will then favour political candidates that support their narrative, in terms of interviews and coverage that they provide those candidates. This is akin to a propaganda feedback loop. If one media doesn't toe the line, the audience moves to a media that does support their bias. A key audience on the far right of the political spectrum is Trump's base, a community that Trump has cleverly cultivated since he first ran for presidential office.

Another critical aspect of the approach of the cluster of media on the right is to demonize mainstream media like CNN, the New York Times and the Washington Post, labelling them as fake news. Thus when the mainstream media challenge the right's narrative, their audience can easily discount this counter-narrative as nothing more than lies, coming from fake news to protect their control over the narrative.

This **erosion in trust** of mainstream media (MSM) institutions generally and in democracy as a whole, has been a key strategy for those propagating fake news. They label MSM as "liberal media" and "fake news", a term made popular by Donald Trump during the 2016 U.S. elections. This is a great example of the framing effect, discussed later, where Trump was able to label traditional media as fake media, before that label could be applied to, well, fake news media.

New media is more accessible

Traditional media often uses subscription paywalls as part of their business model. Real, fact-based journalism, which is a foundation of a successful democracy, comes at a steep cost. Subscriptions often complement advertising revenue. The downside is news content is less accessible to the general public, limiting its potential to go viral. (Note: The Guardian is an outlier here, as their content is freely available. The Guardian did break the Cambridge Analytica story, along with the New York Times and Channel 4 in the U.K.).

Newer media like Breitbart (Steve Bannon), Aon, InfoWars (Alex Jones) and even Fox News are more freely available. Fox News does require cable subscriptions, but remains broadly accessible compared to other traditional print media news outlets like the New York Times and the Washington Post. Fox News is also available across the Fox Network. Their content is designed for virality, engaging content that is infused with emotion. And now anyone can create their own media enterprise, whether it's on YouTube (Alternative Influence Network), a Podcast (Joe Rogan) or blogging platforms like WordPress, Tumblr, and LiveJournal. The goal of these media is to get their content out there, making it *as accessible as possible*. Although their business model includes advertising revenue, their key goal is to get others to share their point of view.

Let's examine, in more detail, some of the new media organizations, business models and key protagonists. These examples do show how the media has been used to support Trump.

Breitbart News was founded in 2005 by Andrew Breitbart as a news aggregator for conservative news. It has evolved since to become more of a tabloid news site, combining original content that includes extremist views, xenophobic, sexist and hate speech with a conspiratorial trollish style - which has become a medium for the alt-right movement. It is basically at war with Hollywood, the left and the Republican establishment. Steve Bannon served as executive director from 2012 to

2016, returning again after serving in the Trump administration from 2017 to 2018. Under Bannon, Breitbart News appeared to be a significant Trump booster. Breitbart News focused on issues like climate change denial, COVID disinformation and Gamergate while pushing anti-Clinton and anti-Obama conspiracy theories.

The business model for Breitbart News relies heavily on advertising. A campaign from Sleeping Giants (a social media activist group that pressures advertisers to remove advertisements from conservative media) targeted advertisers on Breitbart News to attempt to persuade them to stop supporting the alt-right site. This campaign was very effective, significantly reducing the number of advertisers and the quality of advertisers, which harmed Breitbart News significantly. As an aside, sites like Breitbart News and InfoWars rely on programmatic advertising, where companies bid for spots on large advertising exchanges to place their advertisements in front of specific people, rather than on specific websites. This is how companies can claim that they are unaware of where their advertisements are being displayed, and how Sleeping Giants was able to impact this situation by alerting brands directly, and shaming them into blocking Breitbart News as an advertising host. Breitbart News is also supported by benefactors like the Mercer family who also supported Trump and bankrolled Cambridge Analytica.

Steve Bannon, who has been instrumental in the growth of Breitbart News as a medium for the alt-right, he is a big player in the far-right movement in the United States, and has also had impact in the world at large, especially in Europe. Bannon started out with a traditional career which saw him serve in the Navy and then move to Investment Banking.

Later, Bannon convinced Goldman Sachs to invest in Internet Gaming Entertainment, which had a business model of exploiting the online game, 'World of Warcraft'. They paid Chinese players of the game for the in-world items and gold they would earn playing the game, which they could then sell for profit on to other players, who could avoid the labour of earning the swag. Through this experience, Bannon saw first-hand the online gaming subculture and how he could exploit it for his own gains for the alt-right movement. Subsequently, Breitbart News's coverage of Gamergate helped develop a connection between 4channers and Breitbart News. This would then expose 4channers to Trump politics. Breitbart News also covered the "Great Meme War", the efforts of 4channers to meme Trump to the Presidency, further mobilising 4channers to the Trump camp and Breitbart News. Bannon understood how to manipulate the 4chan audience for Breitbart News, and Breitbart News also posted about 4chan. But Breitbart News would maintain an explicitly light-touch relationship with 4chan.

Bannon is very much an anti-establishment figure, who has ideologically supported populist nationalist movements with a strict anti-globalist view. The fact that a strict nationalistic view also attracts white supremacists and antisemites made things challenging for Bannon, despite his claims to be neither.

Bannon was also instrumental in the development of Cambridge Analytica and the influence that Cambridge Analytica had on Brexit and the 2016 U. S. elections. By the time those elections came around, Bannon had taken a leave of absence from Breitbart News and worked directly for Trump as chief strategist and served in the White House for several months before returning to Breitbart News. Bannon was key to helping shape Trump's narrative and many of his slogans, including "lock her up", "build the wall" and "drain the swamp", playing on the Trump supporters' cultivated emotions of anger, anxiety and fear.

There is evidence that Breitbart News shared content from the Russian propaganda sites RT News and Sputnik. There is further evidence that those Russian sites also helped amplify content from Breitbart News that served their needs of pushing confusion, chaos and the erosion of trust in democratic institutions. While there is yet no evidence of collusion between the Russians and Breitbart News, it certainly appears to be an opportunistic situation for both parties.

Bannon's relationship with Trump is interesting. It is clear that Trump realized how important Bannon was both for his elevation to Presidential candidate and his progress to the White House. Trump hired Bannon leading up to his election win and made him chief strategist on his team. But Bannon was not shy in sharing his opinions, which sometimes challenged the Trump inner circle; as a result, Bannon was fired, and returned to Breitbart News for a short period of time, from which he was also then let go.

InfoWars was formed in 1999 by Alex Jones, who had previously worked in radio broadcasting and needed his own platform to share his more extreme views. Jones was effectively a new media pioneer. The site, and its companion social media platforms – his YouTube channel had 2.4 million followers before it was banned - focused on conspiracy theories, much of which were false flag events. A false flag is the idea that an event that took place did not happen as described or did not happen at all; the claim is that the event's fiction is designed to impact broader policy goals of those in authority. Basically, these false flag events are engineered by the government to increase their control. For example, Jones claimed that the Sandy Hook School shooting did not occur, but instead was a false flag created by the Government to increase gun control. 9/11, the Oklahoma Bombing and the Moon Landing are other alleged false flag events that were widely promoted on InfoWars. On the other hand, Pizzagate was a conspiracy theory which started on 4chan and was heavily promoted on InfoWars, because, like QAnon, the conspiracy theory supports the overall narrative of an evil global elite that Jones wanted to expose and dismantle.

Some of these conspiracy theories did result in legal action, that then resulted in retractions, apologies and legal damages.

In 2017, InfoWars talk shows and other content was reaching 10 million visitors per month, a greater reach, at the time, than traditional media like the Economist. The business model for InfoWars includes revenue from advertisements placed on its site and on some of its social media channels; in 2018 advertisers were asked to boycott the media properties, harming this revenue source. A major portion of InfoWar's revenue, however, comes from product sales from the site. These products include diet supplements and other survivalist-type products - products that you'd need for an impending Armageddon. The site also marketed supplements that helped protect people from COVID 19, which resulted in issues with the FTC. At its peak, InfoWars earned more than $50 million in revenue per year.

Alex Jones, who created and runs InfoWars, is a prominent conspiracy theorist. And many of the conspiracy theories he promotes are rooted in the New World Order conspiracy theory. Roughly translated, this conspiracy theory states that there is a global elite that is running the world and using mechanisms like economic crises, sophisticated surveillance and insider terror attacks, to control the world's populations. It is a conspiracy theory that basically rails against recent trends towards globalism, while claiming global control and deep state control of the global elite. Jones was influenced by his early childhood experiences of small town police corruption, and subsequently the book, "None Dare Call it Conspiracy" which details the New World Order conspiracy theory. This book was written by Gary Allen, the PR person for the John Birch Society. Many of the conspiracy theories that Jones propagates on InfoWars are essentially derivative of the New World Order

conspiracy theory. They are delivered by Jones with a tone of deep paranoia, using anxiety and fear as a motivator to engage his audiences.

In Trump, Jones found a candidate who resonated with some of his world views, or at least was going to fundamentally challenge the status quo in Washington D.C. Jones's relationship with Trump has not always been positive; Trump is not as extreme or paranoid as Jones, so some of his policies were not aligned with Jones. But despite their rocky relationship, Jones supports Trump's conspiracy theory that the 2020 U.S. Presidential election was a fraud, and helped fund and participate in rallies in Washington D.C. before the Capitol attack.

While it's quite clear as to why Jones would support the anti-establishment candidate Trump, it's perhaps more alarming to realize that Trump supported the conspiratorial Jones through his election campaign and after becoming president. Roger Stone, who worked on the Trump election campaign, was also a frequent guest on InfoWars, and served as a go-between between the two. InfoWars received a White House press pass, even if it was only a temporary one. Trump was interviewed by Jones on InfoWars, and Trump would repeat conspiratorial assertions from InfoWars during his election rallies, including the alleged poor health assertions of Hillary Clinton. Basically Jones would stoke narratives that were favourable towards Trump, and Jones had an audience that supported Trump.

During the 2016 Trump election campaign, there was evidence to suggest that Russian trolls amplified news from InfoWars and Breitbart News that was favourable to Trump. This became part of a wider FBI investigation to determine whether this activity was conducted with the cooperation of InfoWars and Breitbart News. Spreading pro-Trump and anti-Clinton news stories was a significant InfoWars effort. This tactic also enabled them to shift the focus of the news cycle. When Trump did something that was going to create negative coverage (mocking a disabled journalist, for example), they could spread content that would demonize his rival (highlighting content from Clinton's hacked emails). InfoWars has also republished hundreds of articles from RT News over several years without permission. Alex Jones has also appeared on RT News shows.

As of this writing, InfoWars and Jones's media empire is in serious jeopardy, due to litigation that has resulted from the harm caused by his propagation of conspiracy theories, particularly the false flag assertions regarding the Sandy Hook School shooting.

When you contrast Bannon and Jones, it is clear that Bannon was a strategist for the alt-right movement and was part of the inner-circle of influence for Trump while Jones was an outsider whose influence was less direct. But both had significant influence on the information battles and the spread of disinformation surrounding the 2016 U.S. election and subsequent years, and their impact was enhanced by the interests of the Russian government.

RT News and **Sputnik** are Russian state-sponsored media that target viewers and listeners outside of Russia. They serve as a source of propaganda for the Russian leadership. The goals of both media are to sow doubt and confusion in the western world, undermining the cohesion of western democracies, while also defending the foreign policy of the Russian state. Basically, their twin goals are to improve the image of Russia to the outside world while sowing discord throughout the democratic world.

For the latter goal, populist leaders and nationalism around the world better suit Russia, hence their support for Brexit and the Trump administration in the United States. They create content that they hope gets traction and is easily spread, and use disinformation (for example, COVID 19 disinformation) and promote conspiracy theories; not just their own conspiracy theories, but

conspiracy theories that suit their destructive aims, propagated in the west by western media like InfoWars and Breitbart News. Steve Bannon has appeared on RT News on many occasions, by his own admission. Alex Jones has also appeared on RT News, and his conspiracy theories are great content for RT News to exploit, whether it is the 9/11 Truthers conspiracy, the New World Order conspiracy theory, Pizzagate or QAnon.

The **Internet Research Agency** (IRA), which is based in Russia, is allegedly controlled by Yevgeny Prigozhin, a Russian oligarch who also owns the Wagner Group. IRA is an often cited example of a troll farm that also wreaks havoc in these information wars. Troll farms are another type of media organization that is unique to the internet. They produce fake news and fake profiles on social media in order to disseminate fake news content and erode trust in institutions and media. The fake profiles also amplify content from outside sources (like InfoWars and Breitbart News) that align with their disinformation goals. Studies have shown that troll farms exist worldwide, and are often sponsored by their governments.

IRA set up several fake Facebook groups, for example, with names like "United Muslims in America", "Stop Islamization of Texas", "Secured Borders" and "Back the Badge". They used these groups to create and foment distrust and hatred amongst and between groups in the U.S. They were even able to organize rallies through their Facebook groups: "Being Patriotic" organized the "Miners for Trump" rally, and "Blue Lives Matter" was organized by "Heart of Texas".

Those who work at these organizations have tasks related to producing original content for the fake groups and profiles. They will also have goals in terms of sharing, commenting and retweeting content from other fake accounts as well as content from other sources that align. This might include content from sites like Breitbart News and InfoWars, commenting on articles related to conspiracy theories or pro-Trump and anti-Clinton stories. Content themes that might target those in the United States, for example, might be designed to reduce voter turnout of the African American community, using a blacktivist fake Facebook account.

Turning Point Action, affiliated with Turning Point USA, is a group that supports younger conservatives in the United States. It has also been accused of running a troll farm operation by paying young people in Arizona to post content on their social media without disclosing their affiliation with Turning Point Action. The content was designed to sow the seed of doubt regarding the 2020 U.S. election and the COVID 19 vaccine program.

Finally, we should acknowledge the number of **fake news sites** that have emerged, which, like Breitbart News, exploit the mechanisms of programmatic advertising for their business models. A hotspot for this opportunistic activity is in Veles in North Macedonia. These sites are easy to set up; anyone can buy an American-news sounding domain from Go Daddy and then set up a WordPress site. They scour the web for interesting stories, to either copy or edit and publish, using ever-exaggerated headlines to encourage more clicks. They seed their articles on Facebook groups that align with the sentiment of the stories. They use several fake Facebook profiles to help spread the stories on these Facebook groups. They also set up their own Facebook pages for their sites, using their fake user profiles to help build a large following.

And why do these fake news sites focus on Trump-related content? With a little experimentation, it was clear that stories that favoured the alt-right and conservative movement in the United States (Trump) were far more likely to go viral than other stories. The stories they could copy and edit already had outrageous headlines and sensational claims. This created far more opportunity to make

advertising revenue. Because of this, they focused their attention on the 2016 U.S. election and on content that supported Trump while demonizing his opposition.

There has been some speculation that IRA and others are behind the organization of these fake news sites in North Macedonia, so far the speculation is unconfirmed. But the owners of these sites already have plenty of financial incentives to engage in such activity thanks to the rewards from online advertising revenue.

These types of sites are engaging in this activity not to spread an ideology, but purely to make money from Google Ads advertisers who might not even be aware that their advertisements are supporting their sites. Their intentions are purely capitalistic.

Facebook and Twitter are becoming more adept in identifying these fake news sites, to attempt to shut down their operations. But these sites can move their social media efforts to more permissive platforms like Parler, which already has a receptive audience that has a significant bias for alt-right content.

Controlling and manipulating disinformation in foreign nations, as Russia does, is just one side of their disinformation effort. They also need to control the narrative within their own borders. They do this with a combination of laws and an increasingly isolated internet.

Roskomnadzor is the internet and media regulator in Russia. The organization manages a blocklist of website URLs that cannot be accessed in Russia. Intermediaries in Russia are liable for the content on their sites, unlike in the United States. The Russian state not only censors what Russians are able to view in Russia, but they also have a sophisticated surveillance system, and can instil fear into those who might want to share "misinformation" about the Russian regime. This has an increased chilling effect on Freedom of Speech. Friends and neighbours are encouraged to turn in those who propagate fake news about the Ukrainian conflict for example. Journalists are also jailed for spreading false information that counters the Russian narrative.

They justify the war in Ukraine as a "special military operation" to de-nazify Ukraine and avoid the genocide of those of Russian heritage in Ukraine. Basically, they frame Ukraine as a fascist regime that must be brought to account. They use the history of the great war (Nazi invasion of Russia in World War II) to justify their actions – despite the fact that in this case, Russia invaded Ukraine. They also passed a law making it a criminal offense to refer to the conflict as an invasion, but rather a special military operation. They also rail against the dangers of the hedonistic values of the west, targeting communities like the LBGTQ communities and justifying that they are at war with the west and western values – which are akin to pure Satanism. This is all part of the Kremlin's alternative reality.

It is noteworthy that Putin has an approval rating within Russia that typically hovers at around 80%, and actually increased at the beginning of the war in Ukraine. This appears to reflect the success of the information control that the Kremlin has in Russia, and also perhaps the chilling effect that comes from the fear of not supporting the leader.

Russia is effectively able to maintain strict control of information within its borders, all the while taking advantage of the liberal freedoms outside its borders to advance its foreign policy agenda.

Meme magic, organic spread

Memes, usually an image with text combined in a humorous way, enable the spread of information in easily digestible chunks of content. Richard Dawkins, an evolutionary biologist, conceived of

memes as the equivalent of a biological gene, its cultural parallel. In his work, The Selfish Gene (1976) Dawkins suggested memes proliferate as a process of Darwinian selection; they spread, they replicate, they mutate and evolve. It turns out, internet memes spread through social engineering – marketers' deliberate efforts to optimize their shareability to gain virality. Content is encoded as a meme, which is easily imitated and spread. Memes convey culturally relevant ideas relating to existing cultures and subcultures.

Memes are designed to be spread organically, which makes them a unique medium, when you compare them with media that are used by traditional brands, like infographics.

For memes to be shared, they require a passionate group of followers to willingly engage with the content and distribute to their social networks. That's what you have in the ideological landscape, a passionate army of volunteers - real and fake - who are willing to share content that supports their values. This makes memes uniquely suited for the spread of disinformation; they are designed to elicit an emotive response, unlike the meme's counterpart, the infographic, which is designed to appeal to the cognitive side of the brain.

Because memes are often humorous and absurd, they are broadly shared, helping the meme go viral, thus rapidly spreading potentially false narratives. Many memes are political in nature; key actors, with a vested interest in changing society, participate in online communities, anonymously, or under a pseudonym, on Reddit, 4chan, Twitter, in Facebook groups, or private messaging apps like WhatsApp, Telegram, Signal and Parler, to spread their content.

Memes will capture the nugget of an idea, and act as a nudge and a reminder to the recipient. We've all seen them in our social media newsfeeds. They appear harmless, but the volume of memes is designed to keep certain ideas top of mind and gently push the key messages of an organization, political party or ideology. Clever memes help the rampant spread of an idea, news or a conspiracy theory, because the message is simple enough to be understood quickly, and new information can be added to the meme over time.

Memes are used by all sides of an issue. For example, memes have been used to great effect in the Ukraine war by those supporting Ukraine. NAFO, a loose-knit collective of Ukraine supporters use memes to target propaganda from the Kremlin, basically trash talking the Russian narrative. By employing satirical humour in war and ironic trash talking, also known on the internet as sh*tposting, they attempt to silence Russian propaganda. In another example, when the bridge to Crimea was partially blown up, memes quickly flooded social media, celebrating the event and mocking Putin, it also happened to be Putin's birthday.

Part 3: Human Behaviour makes us prone to disinformation and trolling

We want community

As humans, **we want to belong**. We are generally spending more time online, so our communities to which we belong are increasingly online communities; this trend was exacerbated by the COVID 19 pandemic.

Many online communities are based on our own interests and passions like gardening and pet care. Some are focused on shared values, beliefs and ideology. This second type of online groups can develop their own culture, and group think. Content that is shared within these groups helps bond

the community. In-group jargon develops (for example on 4chan, lulz, cuck, cuckservative, kek and normies are jargon created by the in-group), along with in-group jokes. There is often an in-group narrative versus the out-groups, which creates an "us versus them" mentality; those within the group develop a sense of superiority over others who are not part of the established culture. An attack on the ideas of the group is an attack on each member's own identity. Newbies are mocked and hierarchies are established.

An example of this culture are the boards on 4chan, which is an image board whose posters are entirely anonymous (to the readers). The content that is posted is ephemeral; it exists only temporarily. Each board maintains a limited number of active threads. After a pre-defined number of posts, a thread can no longer be updated and therefore soon disappears.

4chan/pol/ is one of the most popular boards on 4chan, and the conversations you witness on 4chan/pol/ can be very nasty. Unlike more traditional social media platforms like Facebook, Twitter and YouTube, hate speech is allowed and transgressive behaviour is normalised. The content essentially includes a heavy dose of misogynistic, antisemitic and racially charged rhetoric. At the heart of 4chan culture is the norm of deception and trolling, and the use of irony.

4chan is known as the internet's "meme factory", although there are plenty of memes that are also created on Reddit, including the subreddit r/MemeEconomy. Because content is ephemeral on 4chan, it is essential to convey ideas quickly and with impact; this creates incentives for the creation of absurd and clever memes that stick. And the anonymous nature of the community allows for rapid experimentation without consequence - if a meme that's created is not popular or immediately memorable, it is soon forgotten; you can create several memes to see which ones stick. In order for a meme to work on 4chan, it has to be easy to understand, immensely funny or shocking, and memorable.

While memes are a key part of expression of 4chan/pol, memes are part of a wider narrative that is infused with a heavy dose of irony. It's hard to distinguish between serious and ironic discourse on 4chan/pol/, and irony is used as an excuse to avoid responsibility for one's words or the harm they might cause; essentially a poster can claim they did something "4 the lulz", a catch phrase that expresses doing something for the pure fun of it, to get a rise out of "normies", the out-group.

The /pol/ designation (of the 4chan/pol/ board) stands for "political correctness", which is the antithesis of what the board represents, which is ironic; it essentially counters our offline world's need for political correctness. 4chan/pol/ is a home to the alt-right movement, although not exclusively that movement, nor was the board originally established for that movement. But it is the birthplace of conspiracy theories like QAnon; a theory that broadly (and somewhat absurdly) promotes the idea that a small elite controls the world, a theory that finds an audience amongst those who abhor authority and seek a narrative to counter that authority.

4chan/pol/ supported Trump during his 2016 election bid; Trump was the ideal anti-establishment candidate who shared some of the ideals of the 4chan community, captured in the Trump slogan, "drain the swamp". And some of Trump's actions only increased their affinity for the unorthodox candidate. In their efforts to support Trump, 4channers would perform raids of other social media platforms and comment sections, discouraging pro-Clinton sentiment and pushing anti-Clinton messaging including her health concerns, a conspiracy theory related to her health.

When Trump was elected, news stories proclaiming, "We actually elected a meme president" followed. While the efforts from the 4chan community likely had little overall impact on the election, they certainly provided vocal support with their clever use of memes that spread virally. Weeks prior

to the election, the Clinton email leak provided terrific source material for meme creation; it also proved to be the birth of the #Pizzagate conspiracy theory, further fuelling the anti-establishment narrative. Clinton's description, that half of Trump's supporters belonged in a basket of deplorables, only provided further ammunition for a community which was not short of narratives to meme. The r/The_Donald subreddit would serve as a bridge between 4chan and the mainstream social web. Keys to the success of their meme campaign, "The Great Meme War", was the use of humour, ambiguity and irony – which disguised the true intent and meaning of the content.

Was the 4chan community pro-Trump on an ideological basis, or did they simply prefer to stump for a candidate who was a threat to the current establishment? Whichever it was, they certainly attracted real white supremacists to their cause. Their appeal to the alt-right is quite obvious, as they were triggering their enemies: including liberals, established republicans, gays, and minorities. Were all those who participated in the meme campaigns doing it for ideological reasons, or were they doing it for "the lulz?" – doing it for the pure enjoyment of it?

During the campaign, memes often included the cartoon character "Pepe the Frog". Trump retweeted a meme of himself as Pepe the Frog titled "You can't Stump a Trump". Donald Trump Jr. shared a meme titled The Deplorables, which included a caricature of Pepe the Frog, along with Trump, Trump Jr., Alex Jones and several other conservatives. Pepe the Frog, a cartoon character from the 2005 comic, Boy's Club, has been essentially appropriated as a symbol for the alt-right movement, just like the Swastika was appropriated by the Nazi regime. The Swastika had been a popular symbol through the ages, especially in the Indian religions of Hinduism, Buddhism and Jainism. In 2016, Pepe the Frog was declared a hate symbol by the Anti-Defamation League.

Why Trump?

We have already explored some of the personalities in the information war (Bannon and Jones), now it's time to explore the rise of Donald Trump, the 45th President of the United States.

Trump very artfully controls the narrative that surrounds him, unlike any of his recent predecessors. He created his own persona, not risking the public to form their own opinions. He tells everyone how smart he is, "super-smart", "a genius", and then backs that up by describing his successes. He cultivates a persona of being a very successful businessman, underscoring this point by highlighting his time spent at the Wharton School, at the University of Pennsylvania, considered to be one of the world's best business schools. He led his followers to believe that he was basically the messiah, the chosen one, sent to help save America. This notion was cultivated by Christian conservative groups and evangelists, who formed an unlikely yet significant part of Trump's base support.

Trump identified and focused on the deep-seated anger of the white working class in America, who felt they had been left behind in a world that was getting more politically correct and favouring the disenfranchised, immigrants (hence the "build the wall" slogan and effort) and minorities, at the cost of their own livelihoods. Trump targeted the high arousal emotions of anxiety and anger, highlighting how jobs had been moved abroad as part of a globalist agenda, and that immigrants were replacing them in the jobs that remained at home. The youth migration to cities furthered rural decay and rot. Trump was seen as someone who understood this, and someone who would take on the establishment in Washington, and redress the balance of these trends. And if nothing else, Trump was simply *not* the establishment, and the establishment was perceived as corrupt and self-interested by many of those who followed Trump. So if he couldn't address the issues, in the least he would wreak havoc to the current order; this was particularly favoured by 4channers.

Trump is not politically correct. He is brash and crude. He is a charismatic showman at his events, and oftentimes would make statements that would be seen as real gaffs in the eyes of veteran political commentators, but only seemed to enhance his connection to his followers. Trump never admits fault or failure, he never shows signs of weakness. This strongman appeal is also attractive to his audiences who are looking for a strong leader to recover their livelihoods.

Trump displays the dark triad of personalities: narcissism, psychopathy, and Machiavellianism. These same traits are evident in authoritarian leaders. There is no better example of Trump's narcissism than his release of his Trump NFTs in December 2022. Targeted to his followers, these are an opportunity for those followers to buy digital assets of Trump, which may then afford them an opportunity to meet Trump, personally. It is also clear that Trump loves his rallies, and loves the adoration he receives from his most ardent followers at these rallies.

In terms of psychopathy, Trump shows little care with the personal consequences of his actions. He surrounds himself with people who will agree with him, and does not worry about firing people who he feels are not 100% behind him. He has no problem demonizing his opposition, or even mocking the weak, as was the case when he mocked a disabled reporter, as well as the war hero, John McCain. In essence, Trump displays a clear lack of empathy which is characteristic of those who have psychopathic tendencies.

And Trump's Machiavellian attributes are apparent for all to see; Machiavellianism is the idea that the end justifies the means. In the case of Trump, this is best exemplified by his disregard for the truth, including repeating conspiratorial aspersions from InfoWars during his rallies, to animate his followers. Trump has no qualms in using fear as a motivator, in order to get his messages across.

Trump was very pragmatic about building his follower base, developing slogans that they could easily support, such as "Make America Great Again" (MAGA). Americans find it difficult to argue with that sentiment, especially those who feel left behind by the globalist agenda. His organization sells merchandize like his bright red "MAGA" hats, which became a form of identity for his followers, helping further brand his persona and forge the MAGA movement, and as some would describe, a cult.

The MAGA organization provides a sense of belonging to its members and became vital to Trump's base. As a Trump supporter, you could identify with other Trump fans, which then builds a sense of an "us" versus "them" in a battle of good versus evil, stoking fear amongst the MAGA movement and the need for its "strongman" leader. And Trump helped foment this in-group versus out-group metaphor with slogans like "crooked Hillary" and encouraged chants at his rallies such as "Lock her up". Us: we are the aggrieved. Them: the RINOS (Republicans In Name Only), the Never Trumpers, the Unselect committee embroiled in another "witch hunt"; these were all common Trump refrains. Despite the obviousness and somewhat hockiness of these tactics, they were effective. One can see how Trump (or his team) is a marketing genius.

Over time, Trump has failed to deliver on many of the promises that he made, including his assertion that COVID 19 would go away without a vaccine, that he would repeal and replace the Affordable Health Care Act, cut taxes, and so on. But it becomes harder for his followers to give up on him, as, psychologically, they've become part of his movement. The movement provided them a sense of identity, and threats to Trump and his movement are a threat to each of his followers. And as Trump's followers forgave him of one mis-step or missed promise, they slide down the slippery slope of continuing to forgive as they turn on those who attack Trump; the logic goes, because he has already established that they are evil, fake news, etc., his followers should continue to support him,

or else they seem like hypocrites. This is a form of habituation that the movement has created amongst its followers, continuing to support Trump even when he is unable to keep his promises. It's easier to stay cognitively consistent with the Trump movement than to admit they have been wrong all along.

After the mid-term elections of 2022, Trump announced his third run for the White House. Since that announcement, Trump's legal troubles appear to be coming to a boiling point. And the legal case that Dominion has brought against Fox News for their defamation case has revealed fractures between Trump and Fox News, while it's clear Fox News still panders to Trump's MAGA base.

Anonymity changes behaviour (just look at 4chan)

We all act differently depending on the setting and who we are with. We act in different ways in front of our parents, our work colleagues and our friends; we act differently online, where some will develop an entirely different persona; the disinhibition effect (John Suler) explains why we do this. A principal reason is the anonymous nature of the online environment, which eliminates the consequences of behaviour to the perpetrator, even if there are real world consequences for others. Thus if the rules of the offline world are requiring you to act in a way that you are not comfortable with, then you can circumvent those rules when you are online and anonymous; changing one's behaviour is particularly attractive to those who feel threatened by the changes wrought by the movement to a politically-correct world, the "woke" movement. It is also the case for anyone who abhors authority of any kind.

Being online lowers your inhibitions and can reveal a more authentic self, which can be good. This is discussed in "Everybody Lies" (Seth Stephens-Davidowitz). Being anonymous can also help people explore their identities, if those identities are not part of their mainstream culture. But it can also become a playground for those who want to act out their fantasies, which can manifest in a variety of negative behaviours. For those who exhibit the dark triad of personality traits that favour narcissism, psychopathy and Machiavellianism, harmful behaviours then include online bullying and a variety of troll behaviours.

We should also consider that there are 1.5 billion online gamers, nearly a fifth of the world's population, who are acting out their play in an online environment, allowing them to escape from the constraints of the real world in immersive games that generally focus on violence and skill. Cultural toxicity is an ongoing issue within gaming communities, and in this regard, toxic behaviour can be contagious and somewhat normalized. This helps explain why PewDiePie (real name, Felix Kjellberg), the first YouTube star to eclipse 100 million followers on his channel, has published content that included antisemitic, racist or misogynistic references in several videos between 2017-18. He would then apologize for each instance.

Anonymity also provides cover for bad actors to sow their seeds of destruction and confusion, as Russia has perfected over several decades of information warfare - is that blacktivist tweeting about social injustice real, or are they part of a Russian troll farm as part of a disinformation campaign or are they 4channer using a fake Twitter account, doing it for "the lulz"?

Troll behaviour

Troll behaviour is enabled by anonymity; because we are anonymous, we can say things to other people we wouldn't dare to say if our name was revealed.

The misogynistic movement, Gamergate, is an example of very toxic troll behaviour. Gamergate, which lasted from 2014-2015, was essentially a loose-knit online misogynistic movement campaigning against feminism, targeting women in the gaming industry, notably video game developer Zoe Quinn and feminist writer Anita Sarkeesian, who had been a victim of prior troll attacks. The controversy began when Quinn created a game which received positive reviews from the gaming media, but was panned by some of the gaming community. These gamers determined that the press coverage that the game received was biased - a blog post from a former boyfriend of Quinn accused her of sleeping with one of the reviewers.

This sparked a culture war, with gamers claiming to be fighting for gamer identity in the fight against the political correctness in the gaming industry and the ethical breaches in gaming journalism. Harassments and threats against Quinn and others included rape, death threats, and doxing - revealing personal information and nude photographs of the targets.

Fake accounts were set up on Twitter to share and spread the scandal. Bragging about these exploits occurred on 4chan, and subsequently 8chan, when the conversation was ultimately banned from 4chan. Other neo-Nazi groups joined the trolling to increase the mob size and further wreak havoc. The reality of this culture war was that there were likely only a small number of brutal trolls, but this small minority was able to simulate a loud voice. And they were able to silence voices that criticized the traditional gaming culture, who would fear for their own safety and privacy. Trolls would call out "social justice warriors", those who express socially progressive views.

This works the same with troll attacks from 4chan that may target other platforms like Tumblr, and other communities like the LBGTQ+ community. 4channers can decide upon a collective action and undertake it without sponsorship from outside interests or even any real organizational structure of their own; individuals then brag about their trolling exploits outside of 4chan when they return to the platform, thus inspiring others to do the same.

One example of this type of action occurred in 2016, when 4channers were able to turn Microsoft's AI chatbot, Tay, into a sexist racist. Basically, the bot learned its behaviour from the online environment in which it engaged, Twitter, Instagram, Facebook and Snapchat. 4channers were able to swamp these platforms and engage Tay with racist and sexist comments and content related to conspiracy theories, turning Tay into a racist, sexist conspiracy theorist. And they did all this for "the lulz". Microsoft Tay was effectively an earlier version of ChatGPT, launched in 2022 to significant response, and better guard rails, presumably.

Victim trolling is another example of trolling. Conspiracy theorist Alex Jones claimed the Sandy Hook shooting was fake, and the parents of the victims of the shooting were 'crisis actors'. This then inspired trolls to target those parents with vicious troll attacks. Because of the harm caused to the parents in this particular case, Jones finally admitted he was wrong, and now has to pay over $1 billion in damages because of the abhorrent behaviour of the trolls that he inspired.

In the information wars, troll behaviour is used to silence voices, and sway decision makers from making the political decisions they desire and feel is right. Trolling politicians, on Twitter, for example, with misogynistic and vial language, reduces debate and can lead to those on Twitter to leave the conversations. Meghan Markle is a high-profile victim of troll attacks. Markle is a strong black female voice - an obvious target for internet trolls.

As the online environment begins to normalize trolling and other anti-social behaviour as a result of the anonymous nature of the online environment, this normalized behaviour spills over into the offline world; there are several examples of this, including Pizzagate, The 2021 Capitol riots of

January 6, the attack on Paul Pelosi leading up to the 2022 midterm elections in the United States, the mass shooting in Australia in December 2022 that left 6 people dead, and the suicide of Molly Russell. This is not just an American problem, but it's a problem that is exacerbated in America because of its more liberal interpretation of Freedom of Speech, and the easy access to guns to act out crimes.

We are lazy, too

As humans, we are **cognitively lazy;** we have to be, given the amount of information we see every day; we suffer from decision fatigue. As such, we tend to rely on heuristics, cognitive biases, and shortcuts to help us make decisions and process information; ultimately, this makes us more susceptible to disinformation, fake news, conspiracy theories and marketing messages.

Known cognitive biases include *confirmation bias*. Because we are in online communities of similarly-minded people, we see more content that supports our existing beliefs, and less content that contradicts those beliefs; these communities essentially become echo-chambers, where similar beliefs echo around the community. We are also more inclined to consume and share content that already supports the beliefs that we have, because this requires less cognitive load, less effort and aligns with our existing beliefs, which is appealing. It is also hard to change beliefs; once a conspiracy theory has been exposed as false, that information doesn't generally gain traction in related communities, the communities are already vested in their beliefs, so will tend to ignore new information that is contradictory.

The *anchoring bias* means that we tend to be overly influenced by the first source of information that we hear or read about an event. Thus, if the first report that we see of an event is from our favourite online source, we will put more weight on that version of the story, and it biases how we interpret following news stories on the same event. This has wide implications as we spend more time on social media and gain more exposure to news from sources other than traditional media. It also means that media that spends more time gathering all the facts for a story will be at a disadvantage as more swift media, which just wants to use the story to support an on-going narrative, selectively targets a few facts from an event and gets the story out sooner. Trump uses this bias to great effect in order to get out in front of a story which may be damaging to his campaign.

The *mere exposure effect* (availability heuristic) means that if something is repeated often enough, we start to believe it; Joseph Goebbels, Nazi minister of propaganda under Adolf Hitler, understood this. Propaganda works this way under authoritarian regimes. In Russia, if the only news source you have is from the regime itself and you are not exposed to a counter-narrative, you will start to believe the prevailing narrative, or at least start questioning your own beliefs. This is known also known as gaslighting.

The mere exposure effect also normalizes extreme behaviour, whether it's a teenager who commits suicide after being exposed to suicide-related content on Instagram and Pinterest, or a virulent troll from 4chan who is stalking a target with vile content.

The *bandwagon effect*, another cognitive bias, manifests itself online as social proof. If something looks popular, then it must be right. Thus, content that has received many likes gains social proof and is more credible as a result. Others who see that content assume that it is a popular point of view, and will believe it, simply because of its popularity. This has significant implications for content consumption and sharing, especially when popularity can be simulated by troll behaviour.

The *affect heuristic* is where we rely on our emotions, rather than rational thinking, when making decisions. There are things you've said or done in the heat of the moment and later regretted. That's triggered by the affect heuristic. As community members are exposed to content that supports their point of view (confirmation bias) that makes them angrier, then they are more likely to take emotionally-driven actions, which may help explain some trolling behaviour.

The *framing effect* is a marketing tactic that looks at how we frame an idea in order to make that idea more appealing to the target audience. This effect is the reason why the same news event can be covered by two different news outlets, offering contrasting versions of the same story, led by deliberately different headlines highlighting different aspects of the story. It is also a tactic used by the "Make America Great Again" (MAGA) movement, using slogans like "drain the swamp" and "crooked Hillary". Boris Johnson used it to great effect during the Brexit campaign, asserting that £350m a week could be diverted back to the U.K.'s National Health Service (NHS). This argument may have had a sliver of truth (the £350m was money going to the European Union (EU)), but only works because the vast majority of the population do not question it.

The *nostalgia bias* explains our sense that everything was better at some point in the past. While many advertisers use this tactic, one current example is the slogan of the MAGA movement, Make America *Great Again*. The "great again" appeals to our sense of nostalgia. And it also appeals to our sense of loss, and *loss aversion* is another important cognitive bias. The idea is that we fear loss more than we seek gain. "They're going to take our guns" is a common refrain from the gun lobby in the United States. Here, the National Rifle Association (NRA) is focusing on the loss of guns to agitate its membership, rather than the benefits of having more guns.

There are literally hundreds of cognitive biases that can short circuit our critical thinking when we are exposed to new content and ideas.

Belief in **conspiracy theories** is enabled by a number of cognitive biases, such as confirmation bias; we give more weight to evidence that supports our existing beliefs, than weight to evidence that contradicts those beliefs. *Projection* is another bias that supports conspiracy theories; those who believe in conspiracy theories tend to be those who believe in rumours and are suspicious of others. *Jumping-to-Conclusions bias*, where we tend to make up our mind about a particular situation without examining all the facts is another bias that helps explain why some people support conspiracy theories. The *illusory pattern perception bias* is where we look for patterns in seemingly unrelated events, to develop meaning that's not there, turning unrelated events into coincidences, and coincidences into correlations; this is essentially how the Pizzagate conspiracy theory was developed, tying together information from Clinton's leaked emails with prior conspiracies and Jeffrey Epstein, making connections between paedophilia and pizza - **c**heese **p**izza (mentioned in the email exchanges) is code for **c**hild **p**ornography, for example.

While belief in a conspiracy theory may be difficult to understand, the believers fundamentally believe that they are fighting against efforts of deliberate government control, and a conspiracy theory like QAnon supports that idea, which forms some of the basic tenets of the conspiracy. There will be elements of truth within the conspiracy that will be enough for the theory to take root. Repetition of the theory is then exacerbated in the echo-chambers of the communities that are discussing the theory. Thus the conspiracy theory seeps into the recipients' conscience. Another example of this happening is with the anti-vaxx communities, where there were some elements of truth within the broader conspiracies, and these were used to manipulate a narrative of increasing government control - the core idea propagated is that vaccines are another means of more authoritarian control.

Part 4: Social Media

Social Media is a big problem

More than half the world is on social media, and we are spending an increasing amount of time on social media.

Social media has done wonderful things in terms of connecting us, and reconnecting us to lost friendships. But there's been a growing dark-side, driven by the business models of the social media platforms.

The popular social media platforms like Facebook, Instagram, Twitter and TikTok rely heavily on advertising revenue for their business model. To increase advertising revenues they need to increase users on their platforms and *engagement* among their active users. The more engagement, the more time on site, and therefore the greater the advertising revenue for the platforms. And they are not just competing with each other for our time, but they are competing with Netflix, Spotify, Fortnite, Call of Duty, and other forms of entertainment. Our attention is a finite resource, and these platforms are geared to try to increase their exposure to our attention. It's a zero sum game, the more attention we give to one platform, the less we give to others.

One way to increase user engagement is the use **algorithms** to improve the personalization of the newsfeed content that each user sees. The algorithms prioritize content that they believe will be more engaging and elicit greater reactions and activity among users. They do this using artificial intelligence/machine learning, learning what content and content types create greater engagement for each user. Algorithms also consider the engagement that a piece of content has already received, and expose content that has a high level of engagement to a wider audience, to increase its engagement further; more engagement yields more exposure, which can lead to more engagement, an ever evolving feedback loop.

Because fake or distorted news and vile content that is designed to appeal directly to specific audiences is generally more engaging, leading to more likes, loves, comments and anger responses, that content gets prioritized by the algorithms. This explains how misogynistic content on TikTok from Andrew Tate went viral and started appearing in teens' newsfeeds.

Content can also be "hijacked" if the comments section of the content is swamped by trolls - comments attacking the original poster or comments that contradict the original post. This type of engagement will trigger the algorithms, highlighting that the original creator essentially lacks control over how their content is spread on social media.

Spreading and sharing disinformation earns you more engagement and gets you more attention online; this is part of our social reward. And if you are gaining more attention through sharing false news, you may do it more often, whether you are doing it deliberately or the information aligns with your beliefs and you genuinely do believe it. In addition, the disinformation might be so outrageous, it gets spread for the irony of it.

Of course, Fox News understood the power of emotionally-fuelled storytelling as far back as the 1990s. A critical difference here, though, is the connectivity that the internet and social media adds to the mix, enabling the spread of disinformation.

Social media platforms make more money through more activity on their sites. They also learn more, and that learning translates into better personalization, enabling the platforms to serve more

accurate content and advertisements to stimulate more engagement. Altogether, this increases the influence of social media as a source for news; because we are spending more time on their platforms, we are prone to see our first source of a news event on these platforms, triggering our anchoring bias.

A second method of increasing engagement by the social media platforms is the use of **gamification** tactics to keep us on their platforms and engaged; how far away are you from your phone, and the next ping, does it trigger a visceral reaction? How many likes did that latest photo you posted gain for you? The *degree of variability in the reward* for your engagement on social media is a deliberate tactical copy from the gambling industry. Social media platforms are designed to become **addictive**, dripping out dopamine hits as you see the reactions of others in your newsfeed. It is noteworthy that the only two industries which call their customers "users", are technology platforms and the drug trade.

Combine the addictive nature of social media on its users, with cognitive biases like confirmation bias and the mere exposure effect, and we are prone to lose agency; we lose control of our ability to make rational decisions. Unfortunately, this is exacerbated in teenagers, who have not reached the levels of maturity necessary to make good decisions. Mental health is coming under near-constant attack for those who are more prone to inhabiting the online world.

A third crucial flaw for social media platforms is their struggle for **appropriate censorship**. A little context is important here.

Social media platforms are not considered the publishers of the content that they host. In the United States, they are protected by Section 230 of the Communication Decency Act (1996). This is unlike traditional media, which is liable for the content that they publish. The rationale for this law is twofold: one - it was to help a nascent industry to grow, and two - the content on the platforms is created by its users. This law is being challenged in the U.S. Supreme Court with a case being brought against YouTube, for recommending violent content to an Islamic State gunman who participated in the Paris attacks in 2015. Some of the justifications for keeping Section 230 in place in this case revolve around the economic impact of removing the protection, rather than the true merits of the case. This case is scheduled to be decided in June, 2023.

Social media platforms are private companies, and are free to develop their own rules, and can therefore censor individuals as they see fit; furthermore, the first amendment, in the United States, protects individuals from government censorship, not censorship from private enterprises like social media sites.

Censorship efforts of social media platforms comprise a combination of artificial intelligence and human intervention, but this combination is ill-equipped to appropriately censor content at a massive scale when bad actors are using the platforms to manipulate behaviour; for example, when vaccine emojis are replaced by carrot emojis in vaccine discussions, an artificial intelligence system needs to learn of the switch before potentially over-correcting harmless or unrelated food discussions. A whole new category of algospeak has evolved with the specific purpose of evading censorship or the algorithm downgrading of target keywords. Examples of algospeak include unalive for suicide, dead or kill, cornucopia for homophobia and pron for porn.

There are essentially two broad areas of censorship that platforms undertake: removing fake accounts and moderating harmful fake news content.

Fake accounts, which on Twitter, for example, can be created by Twitter bots, are often used to spread fake news, create fake news, and generally troll. They will also form botnets, which are networks that share and like each other's content, making the content appear to be more engaging than it actually is. This additional level of activity creates social proof for the content in a manner that misleads real accounts to believe the content is genuine and interesting, due to its high level of engagement.

The elusive number of fake accounts on Twitter was allegedly a reason put forth by Elon Musk for his hesitancy to buy Twitter.

That Twitter deal ultimately went through, leaving Musk in charge; Musk is a self-proclaimed free-speech absolutist, stoking fears that Twitter may relax its censorship activities, allowing more hate speech and disinformation. Musk has yet to establish a pathway forward on censorship, but he has determined that they will set up a content moderation council, which will rule on issues related to accounts that are banned regarding censorship issues. Facebook had previously established a similar council. That said, Musk has already announced an amnesty for all banned accounts that did not break U. S. law. This decision was based on a Twitter poll initiated by Musk.

Trump is one of the users banned from Twitter (and other platforms) after the Capitol attack, and it is widely understood that this is a decision that Musk had not supported. Trump has now been invited back to Twitter, Facebook and YouTube, but he has thus far only posted one on Facebook. He has established his own social media platform, ironically named, "Truth Social", and Trump is financially obligated to favour Truth Social for his social media posts, in the short term. But with his pending bid for the White House, he may well wish to return to networks with much larger audiences.

In an effort to better monetize Twitter, Musk has decided to charge a fee for verified accounts. This decision has been widely criticized as it presents opportunities for fake accounts to gain verification, which will only create additional confusion and the erosion of trust in the platform. Facebook and Instagram have now followed this initiative, although they require a document to support the authenticity of the account to be verified. This move is considered an initiative to create additional revenue streams beyond paid advertising.

In general, when platforms do take appropriate censorship actions, they are demonized by users and accused of overly censoring speech, in an attempt to erode the trust of the medium by those who feel they are being censored.

Platforms like Twitter and Facebook have increased their resources in terms of adding fact-checking facilities. They can then append posts that they determine do not represent the truth with a notification stating so. But the reality of this process is that the content has already been published, is out there, and has been shared. Therefore, chasing it with labels that question its truth is almost akin to shutting the stable door after the horse has already bolted. Political speech is also protected from fact checking, so content from politicians, including Trump, are not fact-checked.

There is also a disincentive for platforms to remove content, which, in turn, reduces engagement, the fuel that feeds the platforms' business models; if they ban one high profile account, that account's followers become enraged, and some leave with the banned account to another platform with more permissive free speech rules.

The counter argument to this, however, is that advertisers and community members may start leaving the platform as a revolt against perceived increases in hate speech. The increase in hate

speech may also trigger the interests of regulators. This presents a dilemma; moderators must achieve a delicate balance to get the moderation right and balance this with the platform's business interests.

The game "whack-a-mole" is a good analogy for the current state of play with regard to censorship on these platforms; they shut one fake account down, and another fake account starts posting almost immediately.

Developing viral hits

Content marketers try to create content that consumers want to share. Their hope is to produce the type of content that's spread far and wide, as Barilla has done with its Roger Federer YouTube clips, as Dove has done with its Real Beauty campaigns and as Ariel has done with its #Sharetheload campaign. These types of content include certain characteristics that increase the likelihood that they will be consumed, shared, and go **viral**.

Jonah Berger (The Wharton School, University of Pennsylvania) has studied this phenomenon and identified six pre-conditions that increase the likelihood of virality: Social currency, Triggers, Emotion, Public, Practical value, and Storytelling (STEPPS).

Emotion, the E in the Berger acronym, is a key pre-condition. Emotive content that is "high-arousal" content, is more engaging, and thus more likely to go viral. High-arousal emotions include anger, anxiety, surprise, humour, shock, love and fear.

Fake news is more likely to be consumed and shared because it is more likely to stimulate these high arousal emotions than traditional news, which, quite frankly, can be rather dull. Absurdist humour and the use of memes, which enables easy consumptions in an age where our attention span is limited, are deliberate attempts to increase emotional virality.

Social currency represents the first S in the Berger acronym and plays another significant role in the sharing of content. The idea is that we create and share content that makes us look good to our network. It adds to our credibility and online persona. If we see a meme that we think is witty, aligns with our point of view, and is novel to us (we haven't seen it before), we may be tempted to share it. We will then look to see how many likes and reactions we get, our social reward. It can elevate our status among our network if we appear to be someone who provides entertainment and new insight. Similarly, high profile people sustain their profile by sharing content they think resonates with their audience. Donald Trump Jr. describes himself as a "meme war general" on his Twitter profile; he is a prolific tweeter.

Virality is also about good *storytelling*, the final S in the Berger acronym. Take, for example, the QAnon conspiracy theory. Started on 4chan in 2017 by Q (an anonymous supposedly high-level government official), the story is that a cabal of satanic, cannibalistic sexual abusers of children operate a global child sex trafficking ring and conspired against U.S. President Donald Trump. The Trump administration secretly fought the cabal of paedophiles and would conduct mass arrests of thousands of members on a day known as "the Storm".

4chan was already receptive to elements of this conspiracy theory; the Pizzagate conspiracy theory, for example, pre-dated QAnon by a year. Pizzagate falsely claimed that several high ranking democratic officials and restaurants were operating a human trafficking and child sex ring. This conspiracy theory, which was incubated on 4chan, 8chan, Reddit and Twitter, led to a shootout in Washington D.C.

Confirmation bias was also at play with QAnon, making this conspiracy theory more accessible to this community. But the story itself is extraordinary and infused with high-arousal emotion.

The QAnon conspiracy is equal parts bizarre, extraordinary, shocking, and absurd. It contains all the great ingredients for a story to go viral. Q's seeding strategy, with anonymous Q drops building the narrative, led to groups of users conspiring to interpret the meanings of the messages to further build the story. While the QAnon story is ridiculous in itself, it got wide coverage with its linkages to President Trump. It emerged from 4chan (via 8chan) through YouTube and Reddit, and on to mainstream media, to a Trump rally in Florida in July 2018 – normiefication at play. The QAnon Shaman, a QAnon conspiracy theorist from Georgia, who has a long history with the conspiracy theory, was identified as one of the leaders involved in the Capitol riots, on January 6, 2021.

Do people really believe the conspiracy theory? Or do they just believe enough of it for it to matter to that community of alt-right followers? The underlying basis for the theory is essentially that all those in authority are bad people and they need to be eliminated, or at least removed. That narrative aligns with the followers of QAnon and the alt-right. Within that community, confirmation bias and the echo-chamber effect help reinforce the narrative, as Q dropped plenty of fuel to further the evolution of the story.

When "The Storm" didn't materialized, this did not deter the followers of the QAnon conspiracy, as they were habituated to continue to make adjustments of their assumptions to support the wider narrative. QAnon remains a viable conspiracy movement years after its creation.

Adherents to the conspiracy believe their culture is under attack, their way of life is under attack, so the narrative of the conspiracy binds them to a movement of resistance - and it's an us (believers) versus them (non-believers) fight.

There is no doubt that Some QAnon conspirators are true believers of the conspiracy, some believe some aspects of the theory, and some are simply trolling, either for the lulz, or because the theory does speak to their desire to undermine authority. The feeling of belonging to something, being part of something bigger than oneself is also a critical factor in its success. While QAnon started as a conspiracy theory, it has become a political movement, and is basically a cult.

And anyone can simulate popularity

Troll farms like the Internet Research Agency can be deployed to either amplify conversations or silence opposing views, creating the effect of a larger enraged audience, social proof, or lack of opposing positions. Troll farms also create more fake news.

There doesn't have to be a connection between the troll farm, or any troll activity, with the publishers of the content. For example, Russia might have chosen to influence the 2016 Trump election and Brexit, with or without explicit ties with the respective political campaigners. The dense Mueller report neither exonerated or identified specific ties between the two parties. Regardless, Russia can simply choose a side, amplify news that is already being disseminated, and create their own fake news, in order to amplify conversations that favour the outcomes they desire - outcomes that diminish the cohesion of the west. After all, they are already experts in distorting and amplifying news to their own citizens through their control of their domestic narrative.

This same logic applies to conspiracy theories like QAnon. If China or Russia wanted to help spread the theory, knowing its destructive power, they would do so without any direct connection to those who established the conspiracy - maybe themselves - nor do they have to believe the theory. It's easy to set up a bot network to fan the flames of a conspiracy.

The cocktail of different actors, Russian or Chinese interests, nationalists, white supremacists, racists, and religious fundamentalists, sometimes with goals aligned, often with different goals, are able to exploit the mechanisms of digital media to create chaos. We are living in dangerous times, and this is clearly exacerbated by the exploitation of digital media to distort the truth, and create false narratives. We need to take a check in terms of where we are at with the evolution of media, and determine roads ahead that might help return the digital medium to the opportunities it was designed to present from its inception – connecting humanity for the good.

Part 5: Four potential solutions

So What Now?

In the days of fewer media outlets, often with rules to be fair or to not cover salacious stories and obvious lies, we had a shared culture of ideas. We are now in a situation where we don't view the same media, don't share the same stories, or get similar exposure to a common core of thoughts and ideas.

The shared culture thought freedom of speech was so important because, if all ideas could be debated in a common marketplace of ideas, truth would, more than likely, prevail. When conversations are only held in isolated echo-chambers, with no opposing narrative to debate the assertions, they are not challenged, and are able to take root, false or not.

At the conclusion of the Fairness Doctrine in the United States, in 1987, where freedom of speech advocates claimed that the rule forced speech, we saw the rise of conservative talk radio, led by Rush Limbaugh. This was an early warning of the fracturing of the common marketplace of ideas. But conservative talk radio and the introduction of Fox News in 1996 were only the precursors to what is happening now in the online world. Online conversations develop on permissive platforms like Reddit and 4chan. They are stoked by right-wing sites like Breitbart News and InfoWars, and state interests like Russia and China. They are then accelerated by the hungry algorithms of the social media platforms, and further amplified by the conservative news media.

Interest in mainstream media from social media platforms is also waning, with fewer references to it and greater attention focused on alternative influencer networks and newer media sites. Facebook's 'fake news' problem has added to concerns and has led to declining amounts of mainstream media references on the platform. We are getting our first news source in alternate ways which impacts how we understand news events because of anchoring bias.

If we are to assume there will always be bad actors in this world, seeking to manipulate our behaviour for personal gain, then there are few things we can truly control in the above scenario.

There is little will to change our freedom of speech rules in the western world, despite the significant shift in the landscape of conversations. And we know that while these rules, as they stand, do cause harm, we are fortunate not to be under an authoritarian regime where freedom of speech is essentially for the few and not the many. Ironically, authoritarian regimes in these countries are able to control speech at home, while manipulating the narrative abroad because our free speech rules allow them to do so.

We should also note the technology advancements in our online worlds that we inhabit; there are three broad issues here, Deepfakes, generative AI and the Metaverse.

Deepfakes are essentially the AI manipulation of video content, to create the impression that someone has said or done something that they haven't. This is just one common and worrying form of Deepfakes, otherwise known as synthetic media. Examples of this type of manipulation include an experiment from the Massachusetts Institute of Technology (MIT), showing a speech that was prepared for Richard Nixon, but thankfully never happened, entitled "Moon Disaster".

While memes have been the preferred medium to spread disinformation, and a very effective one at that, video content is generally considered to be more trustworthy in the minds of consumers. So, if video content can be effectively manipulated to spread disinformation, this is a potentially very worrying proposition. One recent example of this was in March 2022, when the President of Ukraine appeared to be instructing his military to surrender. This was a deepfake, and never happened. But the consequences of such a disinformation campaign could have been disastrous.

There are two potential outcomes to more common usage of deepfakes. The use of deepfake video becomes an effective weapon for disinformation campaigns because we tend to trust the video medium, or wider use of deepfakes further erodes our trust in media - basically a liar's dividend. Either consequence, or both, don't bode well. Deepfakes are rather novel, at this stage, and hard to create. The direction of travel of the advancement of technology only means that this type of content will become easier to create, especially with the advancements of generative AI.

With the release of ChatGPT in 2022, we are only now starting to see the power of generative AI, and its possible inclusion in the development of fake news in the information wars. Tools like this, deployed by troll farms like IRA, can only help troll farms in their production of fake news. We should also recognize how these "large language model" tools work - they learn from the content on the internet (among other sources). So if the content they are learning from is biased (or fake news), they will produce biased content and more fake news.

The Metaverse has also received a lot of hype over the last year, stimulated by Facebook's name change to Meta and its significant investments in this new medium. And while the Metaverse remains at a speculative stage of its evolution, many think that in the long run, the Metaverse will effectively become a more immersive internet experience for all. This will potentially increase our time in the online environment as we further develop our online identities and personas.

Even at this experimental stage of the Metaverse, we are seeing extensive trolling behaviour in metaverses like Horizon Worlds, Facebook's metaverse platform. The challenge to reducing trolling behaviour is going to be greater in the Metaverse because the content is ephemeral; the trolling activity is only visible as it is happening, unlike the text/image-based internet, where the trolling activity archives as evidence to that activity. So not only will a mature Metaverse likely mean we will spend more of our time online, but censorship operations will have to adapt to a more real-time focus.

There are four avenues to pursue

Humans will always take shortcuts in information processing and sharing, and these shortcuts make us prone to fake news consumption. In this regard, Finland is a country that has made inroads in modern media literacy, teaching children how to process the news more effectively. So bringing greater awareness to the importance of media literacy, and making it a key part of our school curriculum is fundamental. Finland's citizens also tend to trust their institutions more so than we do

in other parts of the world. Sadly, we may be too late to explore this pathway. Finland, as a society, is also more homogenous than the U.S. and the U.K, for example; there are less competing identities and cultures to harmonize.

A second avenue to explore is the need for government oversight regarding the algorithms of social media platforms. While these algorithms remain a black box super-charged for user engagement, we will continue to see our democracies trashed, and extreme behaviour flourish. So, like any competitive industry where harm can be done to the consumer, some form of regulation and scrutiny must be established. The social media platforms are, after all, enabling the pollution of our minds.

The United States Congress introduced safety standards for new cars in 1966. The United Kingdom introduced the Food Safety Act in 1990. It is time that big tech, with its huge influence on the world today, came under the same level of accountability - to reduce harm to the consumer.

A third area to explore is changing the laws regarding whether social media platforms are the publishers of the content that they host. In the U.S., Section 230 of the Communications Decency Act currently protects these platforms, and places the burden on the poster of the content. If this situation was reversed, then the social media platforms would need to rigorously review their censorship processes. In conjunction with this legal adjustment, governments around the world would then shape their own policies towards these platforms so those governments can take a more rigorous stands; The Online Safety Bill in the U.K. is a possible example. Of course, the social media lobbies are powerful and very resistant to change.

It is interesting to note here that Alex Jones, who has developed several conspiracy theories from his InfoWars platform, has been fined in excess of $1 billion for the harm he caused to the parents of the victims of the Sandy Hook school shooting. Jones had previously insisted that the school shooting was a fabrication and the parents were simply crisis actors; the fabrication, Jones alleged, was designed to try to increase the need for gun control measures. While Jones was free to promote this conspiracy theory because of freedom of speech, the harm that this speech caused its victims, who were trolled by Jones's legion of supporters, created the grounds for several lawsuits. Because the courts have now established a significant level of punitive damage, this may deter conspiracy theorists of the future. The Dominion case (regarding the integrity of voting machines in the U.S.) against Fox News, will be another case to follow, closely.

Finally, we can also consider how Google ranks its search results, where a key factor in its algorithm focuses on the authority of the content publisher, rather than only the engagement-level of the content. That authority score is harder to manipulate by new media that is motivated to spread disinformation. According to Moz.com, a site that provides authority scores for web publishers, The New York Times has an authority score of 95, while the Washington Post's is 94. In contrast, Breitbart.com's authority score is 90, and InfoWars's is 83. Google also looks at Expertise, Authority and Trust (E.A.T.) as a further means to establish the relevance of content in a search result. And Google search results, for the most part, still favour a broader spectrum of sources, or at least, don't overly favour the ideology of the alt-right. Of course, even Google's algorithms are not perfect, as the case of the alleged indoctrination of Dylann Roof proves.

In summary, we should consider a combination of approaches to combat the increasing threats of the internet on democracies, worldwide. The four recommendations above may not be an exhaustive list. And we should also consider that nefarious actors, who have enjoyed the freedoms

the internet has presented them to operate in such disastrous ways, will aggressively seek out alternatives, as we roll out any potential solutions.

Post script

During the drafting of this essay, Paul Pelosi, husband of Nancy Pelosi, the leader in the U.S. House before the 2022 midterm elections, was violently attacked in his home in California. The perpetrator was targeting Nancy Pelosi, who was in Washington D.C. at the time. This is a clear example of how online indoctrination spills over into the offline world. The attacker had two blogs and a Facebook page that showed he had been radicalised by alt-right ideology.

Soon after the attack, conspiracy theories spread across the internet alleging that the attacker and victim knew each other and were in a same-sex relationship. There was no evidence to support this, and it contradicted the known facts of the case, but it helped anchor and frame the narrative.

These conspiracy theories were retweeted by Elon Musk (although soon retracted), congresswoman Marjorie Taylor Greene, Donald Trump, Jr., and discussed by President Donald Trump. These individuals not only helped fan the flames and spread the disinformation, but they also served as tastemakers; they are leaders among their ideology and their actions either explicitly or implicitly endorsed the disinformation. All this happened one week before the midterm elections in the United States.

This essay draws from the work of others who I have admired and researched during my 30 years of studying the internet and teaching marketing. Recent works, in particular, have shaped this piece, including Feeding the Demons (Alex Alvarova), Mindf*ck (Chris Whylie), Social Dilemma (Netflix) and The Great Hack (Netflix). I recommend a review of all four of these works.

Many people have helped with the drafting of this essay, whether to help shape ideas or to improve the writing. I will be forever grateful for their contribution. They include Mukul, Leslie, William, Matthieu, Saumya, Daniel, Anders, Richard, Graham, and Sean.

And also lots of content and ideas came from the following resources:

https://journals.sagepub.com/doi/full/10.1177/1461444820948803

https://www.uab.edu/news/research/item/8840-study-says-fringe-communities-on-reddit-4chan-influence-flow-of-alternative-news-to-twitter

https://www.researchgate.net/publication/8451443_The_Online_Disinhibition_Effect

https://firstmonday.org/ojs/index.php/fm/article/view/10643/9998

https://en.wikipedia.org/wiki/QAnon

https://fivethirtyeight.com/features/why-people-fall-for-conspiracy-theories/

https://time.com/6076590/qanon-conspiracy-why-people-believe/

https://www.nytimes.com/2020/10/13/magazine/free-speech.html (subscription required)

https://www.nytimes.com/2021/05/07/world/asia/misinformation-disinformation-fake-news.html (subscription required)

https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo

https://edition.cnn.com/interactive/2019/05/europe/finland-fake-news-intl/

https://www.bbc.co.uk/news/uk-63330885?app=news.uk.story.63330885.page

https://www.washingtonpost.com/news/the-intersect/wp/2014/10/14/the-only-guide-to-gamergate-you-will-ever-need-to-read/

https://www.politico.com/magazine/story/2017/03/memes-4chan-trump-supporters-trolls-internet-214856/

https://journal.media-culture.org.au/index.php/mcjournal/article/view/1422

https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php

https://www.buzzfeednews.com/article/janelytvynenko/infowars-is-running-rt-content

https://www.rollingstone.com/culture/culture-news/meet-alex-jones-175845/

https://www.politico.com/magazine/story/2017/03/memes-4chan-trump-supporters-trolls-internet-214856/

https://www.rollingstone.com/feature/anatomy-of-a-fake-news-scandal-125877/

https://eu.usatoday.com/story/tech/talkingtech/2017/07/18/steve-bannon-learned-harness-troll-army-world-warcraft/489713001/

https://www.wired.com/2017/02/veles-macedonia-fake-news/

https://www.rferl.org/a/macedonia-fake-news-sites-us-election-conservatives/30906884.html

https://www.context.news/big-tech/opinion/social-media-companies-are-to-blame-for-andrew-tate

https://asharangappa.substack.com/p/inside-the-fox-news-sausage-factory