Fake News in an Era of Social Media: Tracking Viral Contagion

Book · February 2020

CITATIONS
11

READS
1,660

2 authors, including:

Fadi Safieddine
University of East London
58 PUBLICATIONS **279** CITATIONS

SEE PROFILE

Fake News in an Era of Social Media

Fake News in an Era of Social Media

Tracking Viral Contagion

Edited by Yasmin Ibrahim and Fadi Safieddine

ROWMAN & LITTLEFIELD
INTERNATIONAL
London · New York

Published by Rowman & Littlefield International Ltd
6 Tinworth Street, London, SE11 5AL, United Kingdom
www.rowmaninternational.com

Rowman & Littlefield International Ltd.is an affiliate of Rowman & Littlefield
4501 Forbes Boulevard, Suite 200, Lanham, Maryland 20706, USA
With additional offices in Boulder, New York, Toronto (Canada), and Plymouth (UK)
www.rowman.com

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: HB 978-1-7866-1421-6

Library of Congress Cataloging-in-Publication Data

978-1-78661-421-6 (hardback)
978-1-78661-422-3 (electronic)

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI/NISO Z39.48-1992.

Contents

Foreword

FAKE NEWS—ANXIETIES AND NEUROSIS IN THE ERA OF SOCIAL MEDIA

During nearly every age and in every society, there has always been an enterprise to counterfeit what society values the most. Even the most sacred of things, religious relics and faith, were not safe from replication and profiteering. These commodities were the victims of their own successes. Of course, information has always had a value, but in the information age, we find ourselves at the mercy of new forms that fabricate and harvest news that is distanced from the truth. Do we call it misinformation, disinformation, and mal-information (Hunt 2003), or plain bullshitting (Frankfurt 2009)? Today, fake news has assumed common parlance as an umbrella term for all of the above while propagating degrees of confusion as to what the term *fake news* actually means. In reality, fake news assumes a fluidity and slippage in the era of social media. Not so long ago, the news was about Internet criminals using spam emails and counterfeit websites to 'pharm' our personal details. In the 1990s and the turn of the Millenium, the talk was about copying and creating counterfeit credit cards and stealing identities through data. The 1980s saw an explosion in counterfeit brands, and for many years there has been counterfeit money. As a bizarre correlation in supporting our argument, in the late nineteenth century there was a great appetite for expensive raspberry jam, so much so that there was an industry around counterfeit raspberry jam using wooden pips to give the impression that it had high fruit content (Bender 1981)! Emulation and simulacra are intrinsic to human civilization and imagination.

One may argue that attempts to curb information, fake or otherwise, may be deemed as an attack on freedoms of speech and expression. With the

advent of the Internet and WorldWideWeb, we assumed that more information would lead to a more transparent and open society, which would help augment our democratic values. Moving from states and organizations to networked society, there was an underlying assumption that this information society would produce value and social relationships through the transaction and exchange of information as a valuable resource. In the era of information superhighways and transcending borders, inevitably information became another commodity that could be reprocessed by a whole new set of actors with varying agendas.

The intangibility of data and user-friendly technology that allows it to flow freely meant that information could be reconfigured and fabricated to target audiences on a global platform. What would ensue would be a slew of terminologies from counterfeit news, counterfeit facts, and counterfeit truth, dubbed *fake news*, *alternative facts*, and *posttruth*, marking a new era where boundaries between truth and falsehood would become tenuous and difficult to discern. While falsehood, misinformation, and counterfeit are not new in themselves, the intensity of their spread and the consequences of their making became something societies had to reckon with in 2016, and so hastily the world woke up to new realities and a new threat under the banner of 'fake news'. Abuse of trust and propagation of falsehood have become a threat to those freedoms represented in our democratic processes, the world order, and our health systems. Is technology to be blamed? How are we complicit in this? Can we solely blame foreign interference?

We've dealt with counterfeits in several ways in the past. One approach is to follow the money and disrupt the supply chain of offenders, thus making the counterfeiting less profitable. Some have used this analogy to suggest demonetization to cut the channels for funding for fake-news content generators. But not all fake-news generators are motivated by money, as there are those who have strong ideological and political motivations. We can also focus on the *quality* of the commodity and educating the masses. That would be desirable, yet not an instant panacea. We can see many news outlets attempting to distinguish themselves by presenting the quality and value of the information they provide to be unique and authentic, and in many cases linked to an online subscription fee. Counterfeiting can be mediated or even eliminated by learning how to identify them, creating tools to support us, and having deterrent policies in place. This may have worked in the past, but will these measures work for fake news?

In this book, we review the notion of fake news from multiple perspectives ranging from the technical and technological to the philosophical. We start to dismantle the chaos surrounding the term with our introductory chapter, 'History of Fake News'. The chapter presents the different types of news that have been labelled as fake news and how they differ from each other.

Importantly, the first chapter locates fake news through a platform economy of data and content transaction, as enabled through social media and its data capitalism. The definition of fake news is followed by a historical outline that demonstrates how fake news has had a vital trajectory and role in human history.The second chapter in this collection, 'Fake News, Media, and the Fourth Estate', examines how the emergence of fake news in the digital age immanently destabilizes the foundations of democracy and equally our conceptualizations of journalism governed through ethical codes of objectivity, veracity, and accountability forming the backbone of any functioning democratic polity. Fadi Safieddine in 'Political and Social Impact of Digital Fake News in an Era of Social Media' examines the effect of social media and fake news in the 2016 US election and Brexit, as well as the role of Russians, tribalists, and other interest groups. Gheorghita Ghinea, Suzana Cardoso, and Marian Caparelli argue in their chapter 'Case Study: Fake News and Threat to Brazilians' Health' that fake-news incidence in the Brazilian health field is high and has a huge impact on social groups' behaviour towards public health, highlighting the detrimental effect on its vulnerable diverse population. Seth Porter focuses on 'The Psychology of Fake News' by asserting that fake news is linked to the manipulation of audiences' cognitive processing where it affects memory recall, decision making, belief in misinformation, and collective memory and cultural norms. Pardis Pourghomi in 'Modelling the Propagation of Fake News in Social Media' argues that the ways in which individuals react to false news are comparable to epidemiology, or the science of spreading diseases. He demonstrates four different scenarios in which studies have modelled the spread of misinformation empirically, mathematically, or through lab simulations to showcase their results. Milan Djordjevic turns his attention to the 'Corporate Attempts to Combat Fake News' where he reports attempts to combat fake news by different social media and search engines, from Google's Fact-Checkers website links to Facebook *Report Fake News* and *War Room* strategies. Djordjevic asserts that while partial research has been done in the area of recognizing programs that can manipulate video and image footage, future studies should investigate people's long-term conviction in false news after preliminary exposure to it and how our manipulations affect our opinions. In 'Academic Research into Combating Fake News', Pardis Pourghomi looks at understanding the underlying incentives behind the creation and consumption of fake news, characteristics for its success, initiatives to educate audiences, filters for fake news, the use of algorithms, and the employment of automated and semiautomated validation tools. Pourghomi primarily focuses on academic research highlighting the differences of these approaches in comparison to that of the industry, illuminating the ways in which researchers attempt to apply a level of pragmatism to gain an insight into human processes and how these can be

paired with technological solutions to mediate behaviours around fake-news consumption. Milan Dordevic and Fadi Safieddine in their chapter 'Variable Identification and Approaches to Validate Fake News' examine four variables—human factors, interaction factors, platform factors, and content factors—involved in the spread of fake news to better understand ~~factors~~ that contribute to the success of fake news. Their starting point is that fake news as an issue remains open-ended as the solutions have not been developed, and several researchers have been unsuccessful in reaching a clear approach because academics and industry are failing to consider the larger environment in which news on social media thrives. Moufid El-Khoury updates on the legal perspectives and its limitations in the chapter 'Fake News: A Legal Perspective'. In deliberating the legal dimensions, he reiterates that global proliferation of fake news makes the legal response more challenging, especially when different legal systems do not share the same underlying policies in relation to speech regulation. In 'The Ethics of the Posttruth Society: Where Are We With This?' Yasmin Ibrahim provides a meta-perspective of the ethical and moral ramifications involved in living in what we term as the 'posttruth' era. She opines that the slide into disorientation and dissonance is even bigger as we seek to unentangle ourselves from these machinations while working out the ethical and moral ~~ramifications~~s of such a bind with the digital economy.

Fake news drags humanity and societies into many other crucial realms, from the protection of our freedoms of speech and expression, the moral role of media as a watchdog, of belief systems, and in sustaining human relationships and political structures through the construct of trust and its vital bind with truth as a virtue. Truth must remain a valuable construct, and that extends to the quality of information we receive and are afforded today and in years to come. Truth is the basis of our science, research, arts, medicine, social justice, and the perception we hold of ourselves and the universe. This collection opens up the struggles, strategies, and challenges ahead of us while recognizing that the debates on fake news will remain open-ended, unsettling, and malleable in the era of the digital.

Dr Fadi Safieddine, University of East London
Professor Yasmin Ibrahim, Queen Mary, University of London

REFERENCES

Bender, A. E. 1981. 'The Mythology of Food'. In *Good Health—Is There a Choice?* edited by P. H. Fentom, 27–35. London: Palgrave.

Frankfurt, H. G. 2009. *On Bullshit.* Princeton, NJ: Princeton University Press.

Hunt, D. P. 2003. 'The Concept of Knowledge and How to Measure It'. *Journal of Intellectual Capital* 4 (1): 100–13.

1

History of Fake News

Dr Fadi Safieddine

'Fake news' has become a buzzword in recent years, and as such has been defined, labelled, and classified in a variety of ways. This chapter introduces the term *fake news* starting with the general perceptions people have about the news that can be labelled as fake. Many of the examples presented here help demonstrate that the sophisticated techniques of faking news. The chapter presents the different types of news that have been labelled as fake news and how they differ from each other. Thus, establishing that beyond its nonfactual nature or lack of veracity, fake news is located through a platform economy of data and content transaction, as enabled through social media and its data capitalism. A review of the literature helps identify a variety of definitions of fake news, and is followed by an analysis which seeks to build a more encompassing definition, one that accounts for the key characteristics to distinguish fake news from other terms that have been conflated with fake news in recent times such as *post-truth*, *propaganda*, *alt-truth*, *satire*, *parody*, and more. This chapter identifies fake news as a phenomenon in which news or narratives have been manufactured to deliberately deceive a targeted audience, and designed to be transacted through a networked economy which disseminates through the trust and the attendant moral or psychological disposition of that network. Critically, this chapter presents tools to help readers better distinguish and correctly label fake news. Finally, a timeline of fake news is presented for readers to appreciate how fake news evolved through a myriad of forms: word of mouth, printed material, radio, TV and film, and digital media. The chapter aims to provide a snapshot of key and significant points of references where fake news has had an impact in the lead-up and well into the configuration of the social media era.

OVERVIEW

The term *fake news* can mean many things to different people. Recently, and following the unanticipated results of several elections and referendums, the topic of fake news started getting more attention. One can argue that fake news has become synonymous with election meddling (Berghel 2017), while others argue fake news has a long history that extends far back centuries but now assumes a different manifestation in the social media age. The reason fake news continues to attract added attention is the concern that it's being used and refined as a political weapon of choice (Berghel 2017). The concerns centre on the speed of dissemination and impact it can have. In fact, the speed of its dissemination is exponential compared to how factual news travels. In a study conducted by Vosoughi, Roy, and Aral (2018), they discovered that predominantly politically motivated fake news is reshared by more people and expanded significantly faster than factual news. The problem of combatting fake news is further complicated by the fact that by the time the rebuke to the fake news is generated, the damage is already done. Baum and Lazer (2017) reported in the *Los Angeles Times* that 'A long-standing body of research shows that people are more likely to attend to and later recall a sensational or negative headline, even if a fact checker flags it as a suspect.'

Of course, there is a psychological side to the success of fake news. If the information fits with people's prior beliefs, then it is going to be very hard to change it by simply correcting the misinformation (Rubin, Chen, and Conroy 2015). There is also a rewarding aspect of being part of the fake-news phenomena. Social networks provide incentives for people to share fake news by doing so 'an act rewarded by social media platforms by metrics such as attention, popularity and visibility' (Mejias 2017). For creators of fake news, there are added financial rewards from monetizing it, as it attracts a large number of visitors linked to higher advertising returns. Finally, the ideologists and agents involved in covert operations care little about the validity of the content and more on the strength of the intended outcomes. These ideologist operators would have their moral compass focused on the belief that *the ends justify the means* and less influenced by facts. And while much of the literature has focused on fake political news, it should be mentioned that fake news has covered topics such as vaccination, health, financial dealings, history, science, religion, local affairs, celebrity gossips, and more (Reilly 2017).

EXAMPLES OF FAKE NEWS

Before exploring the different definitions of fake news, this section presents a variety of cases and incidents that researchers and journalists have described as fake news. Reading through these allows one to discern how diverse these examples are and shows how challenging it is to find one clear definition for fake news.

Alternative-Truths

Alternative-truths are assertions that tend to be statements made with little or no evidence to back them up. For example: 'Ted Cruz was an anchor baby', 'Barack Obama wasn't born in the US,' 'The 2017 inaugural crowd was the largest in history', 'We need guns in schools to prevent grizzly attacks', 'There's no doubt that Saddam Hussein now has weapons of mass destruction', and many more statements that are best described as opinion-based. These make more sense as an ideological statement meant to appeal to a specific base. Attempts to reconcile these statements with reality would not likely bear fruit in changing ideologist views. 'For ideologues, the meanings of the words are their effects on the base' (Berghel 2017).

Hoax (Fake) Websites

Hoax websites are online sites that claim to be factual but are forgeries. Burkhardt (2017) gives several examples of fake websites. For example, DHMO is a parody website that claims to uncover hazardous material called DHMO. DHMO actually stands for H_2O—otherwise known as water—and is described on the website as a breakthrough chemical compound that has such amazing properties as to speed up corrosion and causes suffocation. Another example is a paper titled: 'Feline Reactions to Bearded Men'. It is a made-up scientific article that would appear at first instance as a genuine article about the reaction of cats to human beards. The paper has an abstract, finding, methodology, and event references. The whole article is fabricated and has citations to fictional characters such as Dr Seuss.

Social Media

Social media can be defined as using online social networks as a means of communications and the sharing of news. Responding to false reports of an

Israeli nuclear threat to Pakistan, Pakistan's defence minister Khawaja Muhammad Asif, tweeted on 23 December 2016 an ominous response to Israel stating, 'Israeli def[ence] min[ister] threatens nuclear retaliation presuming Pak[istan] role in Syria against Daesh[ISIS], Israel forgets Pakistan is a Nuclear state too' (Goldman 2016).

Sharing Photos

Where visual manipulation is added to a post, it is influential in giving credibility to a fake story. For example, a touching photo shared in the aftermath of the Nepal earthquake went viral on social media (see figure 1.1). The picture shows a boy and girl holding each other, with a moving caption: 'Two-year-old sister being protected by her four-year-old brother in Nepal' (Pham 2015). The image was not from the 2015 Nepal earthquake but from Vietnam, taken in 2007. It has been reused several times, including as a supposed depiction of Burmese orphans and victims of Syria's civil war.

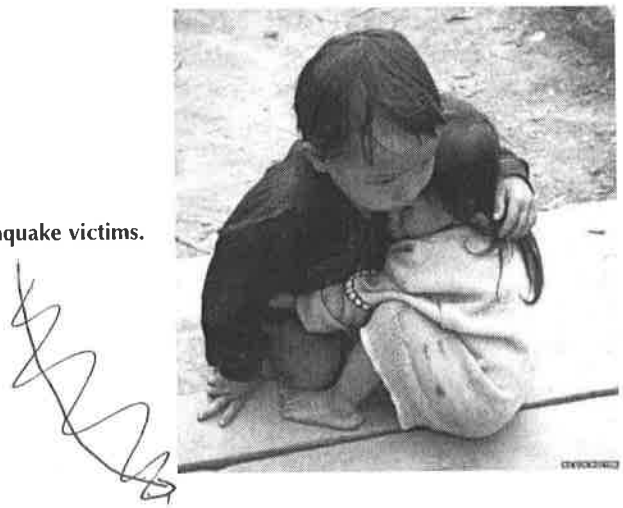

Figure 1.1. Nepal earthquake victims.

Targeted Fake News

The impact of fake news is amplified when combined with targeted user profiling, as explained by Christopher Wylie, the Cambridge Analytica's analyst turned whistle blower (Hern 2018). For example, in the 2016 US presidential elections, the topic of gay marriage had divided the opinion of the US electorates. Profiling Facebook users, Donald Trump's campaign sent targeted messages depending on people's views on topics of LGBTQ,

Figure 1.2. Two contradictory social media posts.

religion, immigration, and the economy that were tailored to each group. LGBTQ supporters and families received posts showing Trump carrying a Rainbow flag with the slogan 'LGBT for TRUMP' (Guilford 2016), whereas people with strong anti-gay-marriage views received messages that suggested Trump was committed to overturning the gay marriage decision.[1] Neither post turned out to be factual (Lacapria 2016). In reality, Trump remained evasive on this topic during the election.

Mass Media

There are instances where fake news goes viral and is shared by mass media. According to McNair (2017), the first-ever modern appearance of the term 'fake news' in the context of abusing social media news posts is in an article published by the *Washington Post* in 2014 profiling Paul Horner, a prolific hoaxer who infiltrated mainstream news media with fake stories. At the time, 'Horner's greatest coup remains a fake story he wrote during the government shutdown roughly a year ago. . . . a ridiculous piece titled *Obama uses own money to open Muslim museum amid the government shutdown*' (Dewey 2014). In that article, Horner claimed that Obama used his own money to keep the federally funded Muslim cultural museum open in Mississippi. Even though there is no museum of this kind in Mississippi, and the story is completely fabricated, it was reported by some mainstream news channels as factual.

Nonpolitical News

Fake news also affects social media posts about nutrition and health. Rowe and Alexander (2017) show how claims of miracle cures proliferate in nearly every medical subject, from mental health to cancer. People who create these posts are not necessarily the shadowy figures that some may assume. The fact-checking website Snopes looked at an example of fake-news posts claiming that sour honey can cure cancer and that pharmaceutical companies, cooperating with Hillary Clinton, know this and are hiding this fact from the public (Kasprak 2017). While this is not unusual to see in a fake news context, what is insightful is the response Snopes got from Agora Inc., the company promoting the post: 'We are extremely reticent to censure our analysts and writers. Instead, we encourage them to speak boldly. And let readers decide for themselves. Of course, we will be wrong often. And embarrassed occasionally' (Kasprak 2017).

News Agencies

There are instances where individuals label news agencies, as opposed to news reports, as fake. For example, Donald Trump had labelled news organizations rather than news reports as fake news when they reported news he disliked; for example, the fact that he did not win the popular votes in 2016, and the observation that there was a low turnout at his inauguration. Gelfert (2018) explains that there is a threat that influencers may use the term 'fake news' to crush dissent. Therefore, it is incorrect to use the term 'fake news' because doing so may lead to it being misused as a form of tactical slandering of opponents. Of course, these are just a sample of what academics and journalists have cited when discerning what is fake news. However, this sample gives readers an understanding of the concept of fake news.

TYPES OF FAKE NEWS

Another way to explore fake news is to look at the different types of fake news as reported in the academic literature. Our review identified seven types of fake news: satire news, parody news, propaganda news, manipulated visual content, fabricated content, alternative truth, and false connection.

Satire News

Satire news is factual news presented using humour, irony, exaggeration, ridicule, and mockery (Brodie 2018). Satire news programmes typically use

humour or exaggeration to present viewers with news updates. Examples of TV satire programmes are *Saturday Night Live* in the United States, *Les Guignols de l'info* in France, and *Have I Got News for You* in the United Kingdom. Humour could be considered a form of news broadcasting as it provides a critique of news, including political, economic, and social events. In principle, satire is based on a factual story. However, satire news should not be taken as factual or true. There are certain topics that are nearly impossible to write in a truly satirical way without someone mistaking it for a genuine article. One can argue that when the news is being presented in a satirical way, it cannot be considered fake news. On social media, a screenshot of a tweet created to appear as a genuine Donald J. Trump tweet poked fun by suggesting the president tweeted on the day the Dow Jones dropped 1,175 points: 'If the Dow Joans ever falls more than 1,000 "points" in a Single Day the sitting president should be "loaded" into a very big cannon and Shot into the sun at TREMENDOUS SPEED! No excuses!' (News Literacy Project 2018). The author of the tweet, British blogger Shaun Usher, did not think anyone would believe the tweet to be genuine, but the screenshot of Usher's tweet went viral. According to Brodie (2018), the challenges for satire stories happen when these stories are shared out of context or no longer link to the original satire context. These stories may contain some markers identifying them as satire, but these could easily be lost or missed by the less-informed readers. In the case of the tweet, the markers included misspelling for Dow Jones and link to the drop in the market. One may ask, is satire fake news? *The Onion* is a satirical digital newspaper that publishes articles that are a mix of reality and fiction, with a satirical tone set in a traditional news-organization format. The power of this approach is to connect the satirical story with current affairs. According to the original editor of *The Onion*, Scott Dikkers, satire is not fake news. In defending his newspaper, Dikkers said, 'It's satire. It's totally different from fake news, and it bothers me when those fake news organizations are basically out there printing likes or propaganda [but] label themselves satire' (Hutchison 2017)—insisting that the key difference is the intention. Dikkers maintains that *The Onion* does not intend to pass its news articles as factual, and when that has happened, it was accidental. Purcell (2017) agrees, insisting satire news adds an important perspective to the current news but is not intended to be deceitful. In fact, one of the most successful news articles in *The Onion* to be passed around as factual is the 1998 article titled *98 Homosexual-Recruitment Drive Nearing Goal*. The article was, in fact, a satire aimed at the Westboro Baptist Church anti-LGBT campaign. The problem is that the article was being circulated as proof of a homosexual agenda, which prompted *The Onion* editors at the time to adjust the link

of the article to direct the readers to the site's home page, so the story could be seen within the context of satire it was meant to be (Schillinger 1999).

Parody News

Parody news may also be referred to as 'hoax stories'. Parody news tends to be stories that are completely fabricated but intended as a joke, with clear disclaimers and clues. Brunvand (1963) defined parody as 'jokes [that] defraud the listener into believing that he is hearing a true narrative and then suddenly turn out to be all nonsense.' Parody, in effect, plays a practical joke with readers. There are trickster(s) and target(s). The trickster engages the target into playing without their knowledge or permission. Parody in media has been around and recognized for a very long time. However, as it has moved from print to digital, and the same issues concerning contextualization of satire and inability for readers to relate the given news to its sources, parody news can easily be turned into fake news. For example, there is the case of a 1995 digital story titled *Clinton Deploys Vowels to Bosnia* published by *The Onion* (*Onion* 1995), where someone copied the article and circulated via email. It spread fast and wide, and in that process lost its source with *The Onion*, until it was being read as factual without contextualization and attribution. There are certain topics that some have argued cannot be used in the form of parody or satire because people with the predisposition to believe the topic will likely to believe it as factual, a phenomenon called *Poe's Law* (Brodie 2018). For example, it is nearly impossible to write a parody article on Creationists in such a way that no one mistakes it for the genuine article. This phenomenon seems to be particularly common in areas of belief associated with politics. In 2012, Iran's Fars news agency reported with all seriousness *The Onion* report that suggested rural white Americans would prefer to elect Iranian President Mahmoud Ahmadinejad than Barack Obama as president of the United States (BBC News 2012).

Propaganda News

Propaganda news is defined as news that has some elements of fact, but which are deliberately false, twisted, or exaggerated to a point where it does not reflect reality (Reilly 2017). In other words, 'propaganda' refers to creating news stories with strong biases to an entity. This entity could be a political party, public figure, organization, government, or cause. Propagandists may fabricate facts, be selective with facts, and present opinions as facts (Edgar and Sedgwick 2002, 313–14). Propaganda intends to deceive and mislead the audience. Bakir and McStay (2018) named the propaganda news category a

Misleading Content type of fake news where the variability of dishonestly focuses on what is not presented or missing from the article. An example of misleading content is hypothetical reporting that names and shames politicians for not voting on a motion but fails to mention that a number of these politicians were ill or unable to attend on that day. In this context, Bakir and McStay talk about two types of propaganda, the first being *overt propaganda*, as a more open and transparent form of propaganda like that done historically by military and dictatorships with the intention of influencing public opinion. For example, Dmitry Tulchinskiy, bureau chief of Russian state news agency Rossiya Segodnya, argued that 'propaganda is the tendentious presentation of facts . . . It does not mean lying' (Troianovski 2014). The second type is *covert propaganda*, which tends to present itself as factual news, concealing its true intentions, importance, and source. For example, in the run-up to the invasion of Iraq, CNN and the *New York Times* are said to have unwittingly published false information pushed by the US military as part of a psychological operation (Love 2007). Thus, with covert propaganda, readers are not aware that the purpose of the news post is to influence them. These two types of propaganda can operate separately or together.

Manipulated Visual Content

This category of fake news relates to when an image or video is manipulated by either false content, contextualization, or editing to be presented as factual evidence to back a claim. There are several instances of this appearing on social media, and it is popular with those targeting users who have less time to read, as it allows for a faster time to react. Potentially, a large portion of fake news on social media could be attributed to manipulated visual content, such as the case of a fake screenshot of a Donald Trump tweet described earlier. On some social-media platforms, such as Instagram and Snapchat, the only effective way to spread news is to incorporate visual content. Several approaches have been used in creating manipulated visual content: a news post may be presented as a screenshot that appears to be coming from a genuine news article, a picture can be presented out of context, or an image can be altered. The term *Photoshopped*, a derivative from Adobe Photoshop software that enables image editing, is commonly used to describe images that have been edited. This category, however, extends the definition of manipulation of factual images to include videos created, edited, or manipulated to present a false narrative as a way to use visual deception to add credibility to a news story. This form has become common in social media, fuelled by easy-to-use image-manipulation tools that can produce changes ranging between simple to complex—for example, borrowing an image from a different article

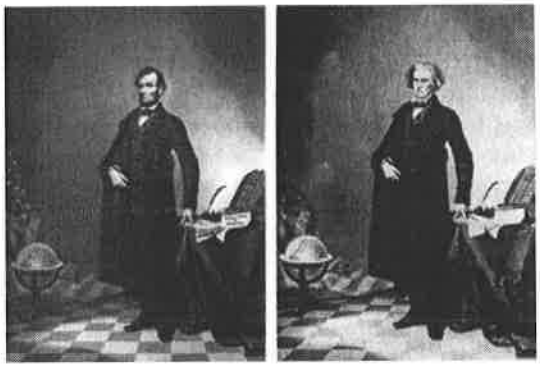

Figure 1.3. Fabricated image of Abraham Lincoln.

and pairing it with a fake story, adjusting changes to an image's colour and removing minor elements, to significant changes that involve removing or inserting humans or objects to an image. The following are early examples of these manipulations.

In 1865, following the assassination of Abraham Lincoln, journalists could not find a heroic-like photo of the sixteenth President of the United States. In figure 1.3, photographer Thomas Hicks used a combination of other photos and imposed Abraham's head to create a photo of the dead president.

In the first photo shown in figure 1.4, the man on the right is Nikolai Yezhov. Nikolai was once Joseph Stalin's righthand man and head of the Russian secret service. When he fell out of favour with the Soviet regime, images were edited to remove him and erase him from history. The photo of Joseph Stalin continued to be used as a propaganda tool.

Figure 1.4. Dotted image of Joseph Stalin and Nikolai Yezhov.

Fabricated Content

This is the case when the content is completely fabricated, with the story having no relation to reality in any way. The post should be viewed as a piece of fiction but is presented as factual news. Burkhardt (2017) called this category of fake news *Canards*, where the 'news' is intended as misinformation or misdirection and its creators have sinister motives. To distinguish it from parody news, fabricated-content news would have no indication whatsoever that it is fictional. Examples of completely fabricated content that have been discussed already: sour honey as as a cancer cure and Obama using his personal funds

to keep a Muslim museum open in Mississippi. Much has been said about the intentions that drive individuals and organizations to create, manage, and promote such content. This type of fake news will most definitely be central to the discussions throughout this book.

Alternative-Truth

The term is abbreviated as *alt-truth* but has a close association with terms such as *alternative-post*, *alternative facts*, *posttruth*, and *postfacts*, and at least in some circles, and before all these terms exploded online, Frankfurt (2009) called it 'bullshit' (BS). There are minor variations in how these terms differ. For example, alt-truths are described as statements that present explanations with little evidence to their validity. Alt-facts, on the other hand, are what is created when 'an ideologue offers an account of the facts that have no basis in reality; they are not false statements but rather alternative facts' (Berghel 2017, 41). Posttruth is a description of events that plays footloose with the truth. The commonality between these terms is that the description given to a scenario goes beyond the truth, rendering the truth irrelevant. Given that there is little variation between these terms, this chapter refers to them as *alt-truth*. The authors of alt-truths do not think what they are producing is any different in approach to the products of what they consider the deceptive traditional media. In fact, the inventors of alt-truth may see their attempts at twisted explanations as being no different to how films, movies, plays, and other forms of art get their audiences to willingly suspend disbelief sufficiently for a meme or post to be accepted as a fact. This form of fake news is probably the most challenging type. Alt-truth is a mix of truth and lies in a blend designed to provide *semantic elasticity* (Frankfurt 2009). An illustrative example of this is the case in the United States of the so-called Death Tax, which really is inheritance tax. By giving it a misleading name, politicians opposed to the tax used a mixture of facts and fabrications to suggest that it would be harmful to middle-class Americans. In reality, the tax does not apply to middle-income Americans and is meant to help them. In alt-truth, the facts are irrelevant; the pitch that is mixed with biased views is much more important. It is more based on emotions and less on fact and evidence. Examples of alt-truth include claims suggesting Donald Trump had one of the largest inaugurations in history, that £350 million would be given to the NHS every week upon the United Kingdom leaving the European Union, and that Obama's birthplace is still in dispute even when his birth certificate proves he was born in the United States. The effectiveness of this type of fake news is that the focus is on a panoramic view rather than the particular view. While most people would agree with US Senator Daniel Patrick Moynihan who

said, 'Everyone is entitled to his own opinion, but not his own facts', the reality is that alt-truth supporters take a different view; they believe everyone is entitled to their own facts as long as these facts serve their interests.

False Connection

False connection is described by Bakir and McStay (2018) as news or post titles that do not match their content. Traditionally, one would associate these with selected tabloid papers. With some tabloids, there is a tendency to exaggerate a claim. Sometimes tabloids have been accused of inventing titles that are from thin air, and more fitting with the fabricated-content category of fake news. Some false-connection articles have just enough truth to make them somewhat believable, thus falling into the category of alt-truth. However, the most common approach of false-connection fake news is to provide an eye-catching title that is an exaggeration of key facts. Sources can be easily found to back any given claim to add credibility to the story and the connection. In March 2016, the *Sun* newspaper in the United Kingdom ran a headline that read, 'Queen Backs BREXIT', a claim said to be linked to 'two unnamed sources'. Later that year, the Independent Press Standards Organisation (IPSO) deemed the title very misleading and forced the newspaper to publish its findings (The Sun 2016). Other forms of false-connection fake news include public relations posts where the title gives a false impression that it is a news article. The article is really an advertisement, or it's a press release intended to appear as independent reporting or news.

Familiarity with these seven fake-news types should help readers navigate the different definitions of fake news and where some researches' definition scopes vary based on their view of the topic.

DEFINITIONS OF FAKE NEWS

To better understand the term *fake news*, there needs to be an agreement as to what one means by the term *news* and the term *fake*. For example, the term *news* could be defined as recent accounts that are of significant interest (Kershner 2005) and that may have an impact on people (Richardson 2007), or even a dramatic account of something unusual or different (Jamieson and Campbell 1997). Traditionally, news was seen as a product of journalism, a field that is expected to deliver 'independent, reliable, accurate, and comprehensive information' (Kovach and Rosenstiel 2007, 11), and as such, journalists are expected to be impartial, honest, independent, free, and self-governed to provide this kind of service. With journalistic power comes

responsibility (Schudson 2003): 'That responsibility lies mostly on being subjective in judging which elements of the news are included or excluded. Beyond journalism, the news is affected by external forces such as political establishment, readers, and advertisers' (Shoemaker and Reese 2013). While journalists make the news, this does not mean they make it fake. *Fake* is associated with being forged, counterfeit, unauthentic, imitation, deceitful, dishonest, misleading, or fraudulent. In 2010, Abbasi et al. investigated fake websites and categorized them into two types: (1) *spoof* sites, which aim to imitate or copy existing websites but make their intention somewhat clear; and (2) *concocted* sites, which have far the more sinister aim of appearing as the legitimate entity for commercial or criminal gain. Another study looked at fake consumers' reviews websites that attempt to deceive users into believing that they are viewing customer-provided feedback when in reality all the reviews are faked (Zhang et al. 2016, 457). Thus, one could consider fake news in a wider context of existing whenever the intention is to mislead or misinform users (Wardle 2017). In the same way one sees variations in how researchers view fake news, the review of the literature shows variations in the definition of fake news. In total, three categories of fake-news definition have been identified.

Fake News as an Internet Phenomenon

The initial impression of fake news as an online problem may have been a reason why some defined fake news as purely an Internet issue. Lawyers Klein and Joshua (2017) define fake news as 'the online publication of intentionally or knowingly false statements of fact' (p. 6). However, their definition excludes satirical websites and posts, such as *The Onion*. There is some merit in considering fake news as purely an online problem. Online social media has provided the dry conditions for the fire of fake news to spread unheeded. The Internet reduced the entry cost and barriers for individuals and organizations to assume the role of journalists. It also allowed foreign covert operations to take advantage of data mining and data analytics to create targeted posting using troll armies. As shown later in this chapter in the timeline of fake news, there was no previous time in human history that these tools were at the disposal of fake-news generators. It is, however, short-sighted to confine fake news to being an online problem only. While many have associated fake news with social-media sharing of news, it is a mistake to suggest that this is exclusively the case. As Gelfert (2018) explains, a post that is fake news on social media is not suddenly not fake news if it is published in a newspaper—as such, defining fake news as exclusively online or social media is incorrect.

Fake News Definitions as a Benign Problem

Some researchers believe fake news is simply part of the Internet. Thus, fake news is defined as harmless, regardless of its intentions. Lazer et al. (2016) define fake news as fabricated information that mimics news media content in form but not in organizational process or intent. Reilly (2017) describes fake news as media hoaxes. The authors suggest fake news is the mischievous end of a broader project and probably aligned with what others describe as satire and parody, 'popped up as shorthand for journalism that should not be taken seriously because it was false, fabricated, or little better than fiction' (McNair 2017, 6). Secor and Walsh (2004) offer a succinct overview: 'Something is made public, people react, taking it seriously, then somehow the rug is pulled away, and people first suspect, then realize that they have been fooled. Sometimes a state of uncertainty prevails, and the event just fades from public consciousness; sometimes the hoaxer gets unwillingly unmasked much later; sometimes the hoaxer is exposed to public opprobrium; more often, the hoaxer claims credit to construct public notoriety for himself or herself' (p. 72).

Fake News as Completely Fabricated Content

Among the type of fake news discussed earlier, the majority of researchers rightly associated fake news with fabricated content. Thus, the following definitions of fake news are identified.

Fake news is:

> news articles that are intentionally and verifiably false, and could mislead readers. (Allcott and Gentzkow 2017, 213)

> a completely fabricated claim or story created with an intention to deceive, often for a secondary gain. (Fake News Challenge 2017)

> information that is designed to be confused with legitimate news and is intentionally false. (Oremus 2017)

> either wholly false or containing deliberately misleading elements incorporated within its content or context. (Bakir and McStay 2018, 1).

> the presentation of false claims that purport to be about the world in a format and with content that resembles the format and content of legitimate media organizations. (Levy 2017, 20)

> one that purports to describe events in the real world, typically by mimicking the conventions of traditional media reportage, yet is known by its creators to be significantly false, and is transmitted with the two goals of being widely re-transmitted and of deceiving at least some of its audience. (Rini 2017, E45).

The problem with many of these definitions is that they assume that fake news is completely fabricated and has no element of truth. This may not always be the case. The other concern with some of these definitions is the scope being limited to political news. Again, this is not always the case. One element that seems to be common is the intention behind the fake news. As such, the most concise definition that encompasses the key elements of fake news has come from Gelfert (2018), whereby fake news 'should be reserved for cases of deliberate presentation of false or misleading claims as news, where these are misleading *by design*' (p. 1), with an emphasis on the term *by design*. The argument is that the term *fake news* should be applied to channels intent on the manipulation of their audience's cognitive processes. Thus, we define fake news as a phenomenon in which news or narratives have been manufactured to deliberately deceive a targeted audience and are designed to be transacted through a networked economy that disseminates through the trust and ethical or psychological disposition of the network.

HOW TO IDENTIFY FAKE NEWS?

Given all the definitions listed above, there is rightly fear that the term *fake news* is turning into 'a catch-all term with multiple definitions' (Lilleker 2017, 1). Of course, one of the key differences between online news generators and established news agencies is that the latter have procedures in place to review and validate the information they receive. 'Newspapers employ fact-checkers to vet a reporter's article before it is published. They often require more than a single source before publishing an article and limit reporters' reliance on anonymous sources. These practices seem likely to raise the veritistic quality of the reports newspapers to publish and hence the veritistic quality of their readers' resultant beliefs' (Goldman 2008, 117). While newsagents are not 100 per cent failproof, by way of these processes, the majority of newsagents remain more reliable than online blogs and social-media content-generator groups. Tandoc, Lim, and Ling (2018) suggest that while fake new can have the same look and feel of factual news presented in the form of a website or post, the intention is a critical component in this definition. With fake news, the intention goes beyond just amusement or harmless fun to become weaponized when combined with news bots, newsgroups, and targeted posting, building a network of pages, groups, and supporters. The risk with fake news is that it undermines journalism's legitimacy, and the source of news is no longer given enough weight. To better understand fake news, Tandoc, Lim, and Ling (2018) reviewed academic publications about how academics view fake news. By doing so, they identified two dimensions that characterize the definition of fake news: level of facticity and level of deception. 'Facticity'

refers to how factual is a news article or a post. This allows for variations as sometimes the key elements of the news are correct, but one minor misquote would not necessarily render it as fake news. 'Level of deception' refers to the intention behind the post, where an error of judgement or humour would not render a piece of work as fake news. Both dimensions can be ranked from high to low. The facticity scale measures the degree in which fake news relates to facts, with the low end of the scale being completely invented to the high end being factual, while intention measures the content creator's immediate intention in misleading the audience—for example, satire and parody may score differently in their facticity, with satire being based on facts but parody being completely fabricated. However, satire and parody both score low on their intention to deceive their audience, as they both aim to entertain or amuse. Whereas the cases of fabrication, false connection, alternative truth, propaganda, and manipulated visual content may be considered high on the intention to mislead the readers scale yet differ from one case to another based on their facticity. In these latter cases, the intention is most definitely planned to be highly deceitful. As such, we can discount satire and parody as fake news (see table 1.1).

This way of assessing fake news allows a better way to judge incidents where genuine mistakes happen in reporting by assessing if the motivation to deceive is sufficient enough to label it as fake news, or to rather consider it in the realm of professional mistakes. Also, the clarity we get from this definition allows us to discount attempts to discredit news that some may simply dislike for political or personal reasons. Another consideration is how to classify articles presented as opinion pieces. Can an opinion be judged using the same scale as factual or fake news? Tandoc, Lim, and Ling (2018) believe opinions cannot be assessed using the same scale. An opinion piece can be high and low on either dimension, and it asserts itself by declaring directly or indirectly its intention and could, to varying degrees, be factual but inherently biased. As such, opinion pieces cannot be considered to be fake news—they are just an opinion. In the same context, one can argue that a work of fiction

Table 1.1. A Typology of Fake-News Definitions

	Author's Immediate Intention to Deceive	
Level of Facticity	*High*	*Low*
High	False Connection Propaganda Alternative Facts Manipulation	Satire
Low	Fabrication	Parody

is not fake news until it starts to be shared as factual. Because intentions here may well be innocent at first, such as mistaking a piece of parody as news, these should not be classified as fake news. However, the intentional removal of the source, removing the contextual links to parody, and intentional sharing of the article could render the final product as fake news.

Can the same criteria apply to the enhancement of images for news purposes? In the Iraq War of 2003, the *Los Angeles Times* used a large photograph showing an intense scene in which British soldiers motions to Iraqi civilians to stay put while a father carries his child across the road (Carlson 2009). Several newspapers around the world shared the image before it was discovered to be a photoshopped amalgamation of two slightly different photos. The photographer, Brian Walski, blended two pictures on his laptop. In reality, the resulting image did not change the context of the situation. One may consider that the image may be fake, but is the photo necessarily fake news? Could it fall into the category of visual repair of an image? After all, the incident did happen, but unfortunately, the man carrying the child had his head turned away in the first photo, making the image less photogenic. Could one argue that editing images is permissible for photogenic purposes? That said, where do we draw the line? When the photo was exposed as a forgery, the *Los Angeles Times* fired Walski immediately. In the news, nothing but factual reporting is acceptable, and while one can sympathize with Walski, preserving the integrity of news is far more important. And while mainstream media has been caught many times with sloppy journalism and photographers faking their images, social-media promotion of citizen journalism means the checks and consequences of faking an image are no longer the deterrents that they once were. Given the criteria above, one cannot accuse the newspaper of publishing fake news, as clearly this was not intentional, and the intention was never to deceive. Going back to the scale model presented in table 1.1, one can now be better at assessing some difficult cases of classifying fake news. Researches have disagreed on how to categorize gossip, rumours, and April's Fool hoaxes as either fake news or not:

- Gossip is usually characterized as possessing 'relevance only for a specific group', thus 'is disseminated in a highly selective manner within a fixed social network' (Bergmann and Bednarz 1993, 70). As such, if presented as factual news, then it is fake news. Where it is not presented as factual, gossip becomes rumour.
- Rumours may sometimes become very different from their specific origins, which tend to be factual or partly factual, but evolve to be reported as news or urban legends (Gelfert 2018). Where they are treated as just

rumours or urban legends, then they are not news. As such, rumours are not to be considered fake news.

- April's Fool hoaxes are a deliberately fabricated falsehood presented as factual and could easily fit into many definitions of fake news. However, unlike fake news, these hoaxes are typically designed to be found out eventually. As such, April's Fool hoaxes fall mostly into the category of parody or satire.

THE TIMELINE OF FAKE NEWS

This section introduces a brief history of fake news along the timeline that allows readers to see how fake news evolved. Importantly, one can start to realize why fake news has become a significant challenge, more than any time before. According to Burkhardt (2017), fake news has existed in the form of rumour and false stories ever since humans lived in groups and power mattered. Burkhardt sets the timeline of fake news into the following categories: pre-printing-press era, post-printing-press era, mass media era, and Internet era. The fake-news timeline is summarized in figure 1.5.

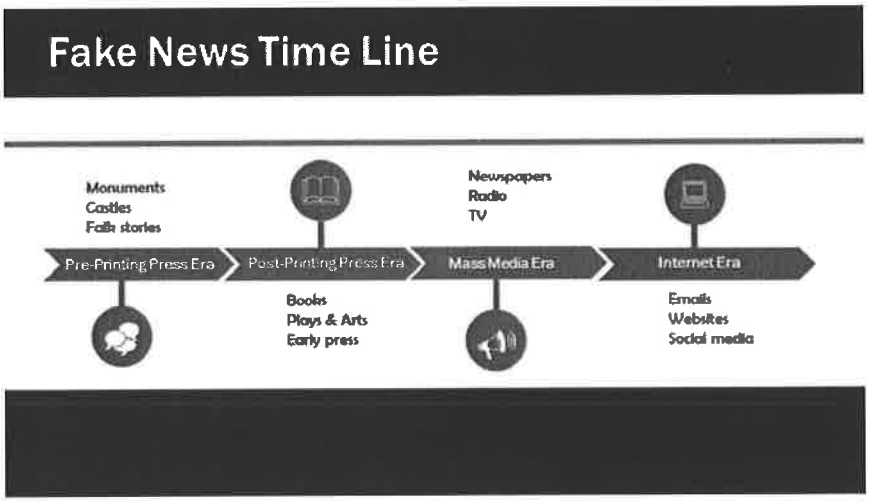

Figure 1.5. Fake news time line.

Pre-Printing-Press Era

During the pre-printing era, much information was carved and written in a variety of small-scale records, but some were intentionally displayed in large monuments, such as the Pyramids in Egypt and castles across the world.

Whatever records were received were contaminated with fake news that pictured the leaders and emperors of their times as heroes in battle. For, after all, history is written by the winners. One example from sixth-century Byzantium historian Procopius of Caesarea (500–554 AD) provides insights to some these practices. Procopius initially supported Emperor Justinian. However, after the death of the emperor, Procopius attacked him, releasing a treatise called *Secret History* that refuted the many achievements accredited to the emperor and his wife. Procopius was pandering to the new emperor who disliked Justinian, and therefore *Secret History* is an example of the type of fake news that occurred in that time. Challenges in spreading fake news before the printing press probably involved interception of the news. The news relied on the powerful; there was very little that could be done to challenge the validity of the news, although this did not stop rumours and satire from flourishing.

Post-Printing-Press Era

As the saying goes, 'knowledge is power' and true to that, power remained in the control of those who could read and write. But as literacy increased, the spread of books and newspapers helped fuel the need for information. Burkhardt (2017) presents two interesting stories in relation to this. The first involves the famous satirist Pietro Aretino, who in the 1500s in Italy, used to blackmail the Italian nobility with slandering sonnets, pamphlets, and plays in exchange for payments. The power Aretino held shows the influence information and, in some cases, misinformation could have for people who controlled information. The second story takes place two centuries later, with the suggestion that the French Revolution could have been fuelled in part by fake news when a picture of Marie Antoinette was superimposed onto a photo of a dragon. At the time, the fake-news story of a dragon-like creature captured in Chile that was being shipped to France further increased the unpopularity of the queen and may have contributed to the harsh treatment she received during the revolution. Perhaps what is most striking is how similar the problems of the eighteenth-century post-printing-press era describe our current challenges with fake news. Jonathon Swift wrote an essay about fake news back in 1710 titled *The Art of Political Lying*. In the article, he describes the real challenge as 'Falsehood flies, and the truth comes limping after it so that when men come to be undeceived, it is too late; the jest is over, and the tale hath had its effect' (Swift 1710). In 1898, the writer J. B. Montgomery-McGovern wrote a scornful essay titled *An Important Phase of Gutter Journalism: Faking*, in which he criticized what he called fake journalism 'resorted by news bureaus to supply their readers with daily news and over sensationalised stories' (Montgomery-McGovern 1898, 240). The

advent of photography did not slow down but rather fuelled some aspects of fake news. Photography manipulation became an art of illusion, as presented in figures 1.3 and 1.4. And more so, manipulated photography played an important role in promoting fake news (Fineman 2012). In fact, Lazer et al. (2018) suggest that fake news became so endemic during World War I that there was a widespread backlash against the propaganda that coincided with the rise of corporate public relations. Shortly after, key regulations passed to ensure the freedom of the press. In addition, pressure built from the public that confirmation and authentication of news stories are needed. It is in the backdrop of that period of serious and rather dry news that *The Onion* found a market and established itself as a light-hearted entertainment newspaper that hovered between satire and parody news. It is clear from this early review that fake news is not recent and has been the subject of concern far before mass media and the advent of the Internet.

Mass Media Era

With mass media came the challenges of radio and TV when used in everything from political propaganda, misinformation, and displacement. Two famous examples of innocent radio broadcastings turning into fake news are *Broadcasting the Barricades* in the United Kingdom and *War of the Worlds* in the United States. In January 1926, BBC radio broadcasted *Broadcasting the Barricades*, a show that implied that London was being attacked by Communists, Parliament was under siege, and Big Ben was on fire. Despite the clear disclaimer at the start of the show, listeners who tuned in having missed the start genuinely believed the show was broadcasting actual events. In 1938, the Columbia Broadcasting System radio network in the United States broadcasted a radio adaptation of H. G. Well's drama *War of the Worlds*. The drama recording presented the story of a Martian invasion that appeared to be live reporting of actual events (Tandoc, Lim, and Ling 2018). While the intention was to entertain the radio listeners, the adaptation assumed the form of live news reports in a time when radio was the main source of information. It is estimated that the broadcasting frightened one million US residents (Cantril 2005). All to the good that they were not intending to spread fake news, but the impact was otherwise. World Wars I and II and the Cold War of the late twentieth century saw propaganda employed in mass media by all sides.

Internet Era

Just because something is printed in a newspaper does not mean it is necessarily true, and the same sentiment should be applied to the Internet. The fact is that the Internet lowered the cost of entry to new competitors, and this

includes those who reject abiding by the traditional standards of the media (Lazer et al. 2018). As such, the technology which came with the Internet era has changed the playing field. With Web 1.0, online users saw incidents of emails being used to spread fake news, and at the time, many categorized this type of fake news as spam. Users also witnessed on several occasions fake websites facilitating the spread of fake news (Cunha et al. 2018). However, the sources were still identifiable and controllable in the form of specialist blogs and portals of fake information. Web 2.0 and the arrival of social media gave rise to the term *citizen journalism*. In the past, citizen journalism was restricted to blogs, but social media provided a platform to present this form of journalism as newsworthy. Social media changed how we view the news. 'Now, a tweet, which at most is 140 characters long, is considered a piece of news, particularly if it comes from a person of authority' (Tandoc, Lim, and Ling 2018, 139). Marchi (2012) found through a study of sixty-one racially diverse high schools in the United States that teens are relying on social-media news, and that they lean towards opinionated news rather than objective news. More and more people get their news from social media; in fact, a survey from 2016 found 44 per cent of Americans get their news from Facebook (Gottfried and Shearer 2016). With Web 2.0, the difficulty in combating fake news became exponentially more challenging as the sharing expanded beyond one's email list to strangers. In addition, sharing moved away from having to include a sources link that would keep the article in context. Many posts repositioned the news article as their own posts (Brodie 2018). The use of social bots (automated accounts impersonating humans) that can *like*, *share*, and *comment* significantly amplified the spread of fake news (Lokot and Diakopoulos 2016; Safieddine, Dordevic, and Pourghomi 2017). One of the less-talked-about elements of fake news is a byproduct of social-media obsession with visual content: the term *memes*. Memes represent an image, video, short text, or a combination of these, typically funny or critical. However, characteristically memes combine a message with a picture that conveys a powerful message, making it easy to copy and share online. Memes are something of a landscape bullet in how a message gets picked up and reposted. Interestingly, McNair (2017) has suggested that satirical memes that poke fun at fake news are the most effective countermeasure against hard-headed supporters who are unlikely to be swayed by facts. The art historian Mark Jones suggested, 'Each society, each generation, fakes that thing it covets most' (Jones 1990, 13). And since we are living in the *information age* where information has premium value, we have reached a point where faking information earns a big prize. There are significant economic incentives for creating fake news that in turn increases traffic and helps generate income from advertisers. People are attracted to gossip, rumour, scandal, innuendo, and the hard-to-believe. The more outrageous yet believable a post is, the more likely it would go

viral, providing the content producers financial revenue from advertisement. 'A man running a string of fake news sites from the Los Angeles suburbs told NPR he made between $10,000 and $30,000 a month. A computer science student in the former Soviet republic of Georgia told the *New York Times* that creating a new website and filling it with both real stories and fake news that flattered Trump was a "gold mine"' (Holan 2016). The incentive to make money seems to override the incentive to do good. That is why social media has been accused of benefiting financially from the monetizing of fake news pages and newsgroups. As a countermeasure, some of social-media giants have moved towards demonetizing some of these outlets, as opposed to closing them and thereby being accused of censorship.

CONCLUSION

This chapter has described fake news and how it is applied in newspapers, news posts, sinister websites, and most recently, social-media newsgroups. It was, however, in the run-up to the 2016 US election that the term *fake news* started getting international attention (Cunha et al. 2018). During the year that followed, newspapers and media outlets reshaped the term *fake news* 'to represent misinformation and manipulation', and 'fake news' became Collin's dictionary most used word of the year for 2017 (Cunha et al. 2018). With the financial, ideological, and political stakes set so high, many concerned stakeholders have upgraded fake news from being a benign risk to being a critical risk. The minds that have worked tirelessly to weaponize fake news and enjoyed the taste of victory are unlikely to walk away from the prize. One should expect that fake-news techniques will evolve to exploit every possible loophole and bypass countermeasures taken to stop them. Social platforms, news agencies, politicians, governments, and researchers will need to learn and adapt faster than any time before. In fact, to collectively succeed, they need to work together in order to be one step ahead.

NOTE

1. Source: https://twitter.com/mikederoode/status/772468230451494912.

REFERENCES

Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., and Nunamaker Jr, J. F. 2010. 'Detecting Fake Websites: The Contribution of Statistical Learning Theory. *Mis Quarterly* 34 (3): 435–61.

Allcott, H., and Gentzkow, M. 2017. 'Social Media and Fake News in the 2016 Election'. *Journal of Economic Perspectives* 31 (2): 211–36.

Bakir, V., and McStay, A. 2018. 'Fake News and the Economy of Emotions: Problems, Causes, Solutions'. *Digital Journalism* 6 (2): 154–75.

Baum, M. A., and Lazer, D. 2017. 'Google and Facebook Aren't Fighting Fake News with the Right Weapons'. *Los Angeles Times*, 8 May. www.latimes.com/opinion/op-ed/la-oe-baum-lazer-how-to-fight-fake-news-20170508-story.html.

BBC News. 2012. 'The Onion Spoofs Iran News Agency on Obama-Ahmadinejad Story'. http://www.bbc.com/news/world-middle-east-19764166.

Berghel, H. 2017. 'Alt-News and Post-Truths in the "Fake News" Era'. *Computer* 50 (4): 110–14.

Bergmann, J. R., and Bednarz, J. 1993. *Discreet Indiscretions: The Social Organization of Gossip.* New York: Aldine de Gruyter.

Brodie, I. 2018. 'Pretend News, False News, Fake News: *The Onion* as Put-on, Prank, and Legend'. *Journal of American Folklore* 131 (522): 451–59.

Brunvand, J. H. 1963. 'A Classification for Shaggy Dog Stories'. *The Journal of American Folklore* 76 (299): 4–68.

Burkhardt, J. M. 2017. 'History of Fake News'. *Library Technology Reports* 53 (8): 5–9.

Cantril, H. 2005. *The Invasion from Mars.* Princeton, NJ: Princeton University Press.

Carlson, M. 2009. 'The Reality of a Fake Image: News Norms, Photojournalistic Craft, and Brian Walski's Fabricated Photograph'. *Journalism Practice* 3 (2): 125–39.

Cunha, E., Magno, G., Caetano, J., Teixeira, D., and Almeida, V. 2018. 'Fake News as We Feel It: Perception and Conceptualization of the Term "Fake News" in the Media'. In *International Conference on Social Informatics* (September): 151–66.

Dewey, C. 2014. 'This Is Not an Interview with Banksy'. *Washington Post*, 22 October. https://www.washingtonpost.com/news/the-intersect/wp/2014/10/21/this-is-not-an-interview-with-banksy/?utm_term=.693cfa1d1d8a.

Edgar, A., Sedgwick, P., and Sedgwick, P. R., eds. 2002. *Cultural Theory: The Key Concepts.* Abingdon: Psychology Press.

Fake News Challenge. 2017. 'Exploring How Artificial Intelligence Technologies Could Be Leveraged to Combat Fake News'. http://www.fakenewschallenge.org/.

Fineman, M. 2012. *Faking It: Manipulated Photography Before Photoshop.* New York: Metropolitan Museum of Art.

Frankfurt, H. G. 2009. *On Bullshit.* Princeton, NJ: Princeton University Press.

Frye, N. 1944. 'The Nature of Satire'. *University of Toronto Quarterly* 14 (1): 75–89.

Gelfert, A. 2018. 'Fake News: A Definition'. *Informal Logic* 38 (1): 84–117.

Goldman, A. 2008. 'The Social Epistemology of Blogging'. In *Information Technology and Moral Philosophy*, edited by M. J. van den Joven and J. Weckert, 111–22. Cambridge: Cambridge University Press.

Goldman, R. 2016. 'Reading Fake News, Pakistani Minister Directs Nuclear Threat at Israel'. *New York Times*, 2 March.

Gottfried, J., and Shearer, E. 2016. *News Use Across Social Media Platforms 2016.* Pew Research Center. http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/.

Guilford, G. 2016. 'Donald Trump's "Support" of LGBT Communities in One Image'. Quartz, 31 October. https://qz.com/823649/donald-trump-unfurled-a-rainbow-flag-with-lgbt-written-on-it-at-a-rally-in-greeley-colorado-to-express-his-so-called-support/.

Hern, A. 2018. 'Cambridge Analytica: How Did It Turn Clicks into Votes?' *Guardian*, 6 May. https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie.

Holan A. D. 2016. 'Lie of the Year: Fake News'. *PolitiFact*. www.politifact.com/truth-o-meter/article/2016/dec/13/2016-lie-year-fake-news/.

Hutchison, B. 2017. 'The Onion Founder: We Do Satire, Not Fake News'. *WISN 12* (Milwaukee), 15 February. http://www.wisn.com/article/the-onion-founder-we-do-satire-not-fake-news/8940879.

Jamieson, K. H., and Campbell, K. K. 1997. *The Interplay of Influence: News, Advertising, Politics, and the Internet (with InfoTrac)*. Belmont, CA: Wadsworth Publishing Company.

Jones, M., Craddock, P. T., and Barker, N., eds. 1990. *Fake? The Art of Deception*. Berkeley: University of California Press.

Kasprak, A. 2017. 'Sour Honey Is a Cure for Cancer?' Snopes.com. https://www.snopes.com/fact-check/sour-honey-cure-cancer/.

Kershner, J. W. 2005. *The Elements of News Writing*. Boston: Pearson Allyn and Bacon.

Klein, D., and Wueller, J. 2017. 'Fake News: A Legal Perspective'. *Journal of Internet Law* 20 (10): 5–13.

Kovach, B., and Rosenstiel, T. 2007. *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect*. New York: Three Rivers Press.

Lacapria, K. 2016. 'Trump Plans to Reverse Marriage Equality as President? Snopes Fact Check'. Snopes.com, 1 February. https://www.snopes.com/fact-check/trump-plans-reverse-marriage-equality-elected-president/.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Nyhan, B., Pennycook, G., Rothschild, D. and Schudson, M. 2018. 'The Science of Fake News'. *Science* 359 (6380): 1094–96.

Levy, N. 2017. 'The Bad News about Fake News'. *Social Epistemology Review and Reply Collective* 6 (8): 20–36.

Lilleker, D. 2017. *Evidence to the Culture, Media and Sport Committee 'Fake News' Inquiry*. Presented by the Faculty for Media and Communication, Bournemouth University.

Lokot, T., and Diakopoulos, N. 2016. 'News Bots: Automating News and Information Dissemination on Twitter'. *Digital Journalism* 4 (6): 682–99.

Love, R. 2007. 'Before Jon Stewart: The Truth About Fake News'. *Columbia Journalism Review* 45 (6): 33–37.

Marchi, R. 2012. 'With Facebook, Blogs, and Fake News, Teens Reject Journalistic "Objectivity".' *Journal of Communication Inquiry* 36 (3): 246–62.

McNair, B. 2017. *Fake News: Falsehood, Fabrication and Fantasy in Journalism*. New York: Routledge.

Mejias, U. A. 2017. 'Fake Becomes Legit: Disinformation, Social Media and Democracy'. *The Disorder of Things*. http://www.democraticaudit.com/2017/02/20/fake-becomes-legit-social-media-and-the-rise-of-disinformation-in-democracies/.

Montgomery-McGovern, J. B. 1898. 'An Important Phase of Gutter Journalism: Faking'. *The Arena 19* (99): 240.

News Literacy Project. 2018. 'A Fake Trump Tweet Goes Out of Control'. https://newslit.org/get-smart/rumor-review-a-fake-trump-tweet-goes-out-of-control/.

Onion, The. 1995. 'Clinton Deploys Vowels to Bosnia'. https://www.ling.upenn.edu/~beatrice/humor/clinton-deploys-vowels.html.

Oremus, W. 2017. 'Facebook Has Stopped Saying "Fake News".' Slate.com. http://www.slate.com/blogs/future_tense/2017/08/08/facebook_has_stopped_saying_fake_news_is_false_news_any_better.html.

Pham, N. 2015. 'Haunting "Nepal Quake Victims" Photo from Vietnam'. BBC News, 4 May. https://www.bbc.co.uk/news/world-asia-32579598.

Polage, D. C. 2012. 'Making Up History: False Memories of Fake News Stories'. *Europe's Journal of Psychology* 8 (2): 245–50.

Purcell, C. 2017. 'The Onion: Inside the Biggest "Fake News" Outlet in the World'. *Morning Herald Sydney [Australia]*, 26 August. http://www.smh.com.au/entertainment/the-onion-inside-the-biggest-fake-news-outlet-in-the-world-20170822-gy1chg.html.

Reilly, I. 2018. 'F for Fake: Propaganda! Hoaxing! Hacking! Partisanship! and Activism! in the Fake News Ecology'. *The Journal of American Culture* 41 (2): 139–52.

Richardson, B. 2007. *The Process of Writing News: From Information to Story*. Boston: Pearson.

Rini, R. 2017. 'Fake News and Partisan Epistemology'. *Kennedy Institute of Ethics Journal* 27 (2): E-43.

Rowe, S. B., and Alexander, N. 2017. 'On Post-Truth, Fake News, and Trust'. *Nutrition Today 52* (4): 179–82.

Rubin, V. L., Chen, Y., and Conroy, N. J. 2015. 'Deception Detection for News: Three Types of Fakes'. In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*. American Society for Information Science.

Safieddine, F., Dordevic, M., and Pourghomi, P. 2017. 'Spread of Misinformation Online: Simulation Impact of Social Media Newsgroups'. In *2017 IEEE Computing Conference*, London.

Schillinger, L. 1999. 'Award-Winning Local Journalists Reflect Own Self-Hatred Back on Nightmarish World: The Onion's Secret Formula: Tragedy Equals Comedy Equals Success'. *Wired—San Francisco* 7:106–10.

Schudson, M. 2003. 'The Sociology of News'. In *Contemporary Societies*, edited by Jeffrey Alexander. New York: Norton.

Secor, M., and Walsh, L. 2004. 'A Rhetorical Perspective on the Sokal Hoax: Genre, Style, and Context'. *Written Communication* 21 (1): 69–91.

Shoemaker, P. J., and Reese, S. D. 2013. *Mediating the Message in the 21st Century: A Media Sociology Perspective*. New York: Routledge.

Swift, J. 1710. 'The Art of Political Lying'. *The Examiner*, no. 14 (9 November), para. 9. Reprinted in Richard Nordquist, 'The Art of Political Lying, by Jonathan Swift'. *ThoughtCo.* https://www.thoughtco.com/art-of-political-lying-by-swift-1690138.

Tandoc Jr, E. C., Lim, Z. W., and Ling, R. 2018. 'Defining "Fake News": A Typology of Scholarly Definitions'. *Digital Journalism* 6 (2): 137–53.

Troianovski, A. 2014. 'Russia Ramps Up Information War in Europe'. *Wall Street Journal*, 22 August. https://www.wsj.com/articles/russia-ramps-up-information-war-in-europe-1408675046.

Vosoughi, S., Roy, D., and Aral, S. 2018. 'The Spread of True and False News Online'. *Science* 359 (6380): 1146–51.

Wardle, C. 2017. 'Fake News. It's Complicated'. *FirstDraft*. https://medium.com/1st-draft/fake-newsits-complicated-d0f773766c79.

Zampoglou, M. 2017. 'Detecting Malicious Tampering in Digital Images'. *Information Technology Institution*. http://pacific.jour.auth.gr/wp-content/uploads/2017/11/Zampoglou.pdf.

Zhang, D., Zhou, L., Kehoe, J. L., and Kilic, I. Y. 2016. What Online Reviewer Behaviors Really Matter? Effects of Verbal and Nonverbal Behaviors on Detection of Fake Online Reviews'. *Journal of Management Information Systems* 33 (2): 456–81.

2

Fake News, Media, and the Fourth Estate

Dr Yasmin Ibrahim

The media, particularly print media and journalism, through time has constantly battled new forms of crisis. Whether it is the increasing plurality of media formats or the vertical integration of media organizations through concentrated ownership, the media is faced with a number of ongoing crises which threaten its watchdog functions and the public's trust in it. These have induced charges of tabloidization and 'dumbing down' of debates in the public sphere and in the shrinking of the 'political' in the age of satellite media and intense print capitalism. In our contemporary digital landscape, the media as we know it through its multimedia formats is in renewed crisis in the guise of fake news. The salience of fake news as a form of viral dissemination through digital platforms targeting audience through data profiles threatens the very foundations of media as the 'fourth estate' and a moral guardian of the democratic polity. This chapter examines how the emergence of fake news in the digital age immanently destabilizes the foundations of democracy and equally our conceptualizations of journalism governed through ethical codes of objectivity, veracity, and accountability forming the backbone of any functioning democratic polity.

We are often presented as living in a postdigital world. The chronotope of the 'post' within a digital architecture implicates seismic changes and lives subsumed into an architecture situated with and without human cognition, and as such premising intense transformation of societies and human identities. Within this configuration is this notion that information, content, and postings generated from human action and thoughts enmeshed with machines in digital platforms will overwhelm us. Drowning us through its sheer velocity, rates of exchange, and recalibration by algorithms, reordering these as data while profiling us through our consumption and production habits and

preferences. The postdigital world swarmed with the leaking of content from a utopian perspective should ideally empower us, enlightening us relentlessly through instant modes of offering news. Yet today the very notion of being saturated with news and posts points to a beleaguered human form, painting an image of vulnerability and presenting an overarching crisis in assembling and consuming truth as the foundation of our political and civic lives.

The notions of truth, authenticity, veracity, and verifiability of news have become a contested ambit framed against the contentious issue of fake news. Appearing through networked digital platforms and online sources, media as a commodity of consumption is under intense moral scrutiny, dragging it into a realm of renewed crisis. Its position as a guardian in safeguarding the foundational ideals in a democratic society as the Fourth Estate, holding powers of the executive, legislative, and judicial arms of governments accountable, is seemingly in disarray against the big threats posed by fake news.

Fake news is related to a wider digital ecology and architecture today and can be defined as content which is deliberately 'misleading, containing significant omissions, falsehood or is designed to deceive its intended audiences' (Dentith 2016, 66). As such, as Matthew Dentith points out, there is an expectation that those who investigate and report the news are bound to an ethical and moral code to accurately report their findings. While the notion of fake news may be perceived as intrinsic to the digital archaeology of news production, it is countenanced against our utopian visions of democracy where news media perform a watchdog role of accountability, accruing a large measure of trust with the public through this paradigm. While the media and press in particular have had to adapt to a digital economy, the various challenges from information overload to the demise of trust in the media to the diminished role of gatekeeping online, these narrate the Fourth Estate as being in a diminished and depraved state while working to a new publishing models. The complex phenomenon of fake news has further added to the notion of the media as being in a moment of sustained and insurmountable crisis in the digital economy where news is handled through a range of actors and platforms without the moral or civic responsibility of safeguarding truth or veracity.

NEWS MEDIA, DEMOCRACY, AND THE FOURTH ESTATE

According to Thomas Carlyle, Edmund Burke identified the press as a Fourth Estate in the eighteenth century with reference to the three realms in France: clergy, aristocracy, and commoners. In a British context, the three estates in Parliament refer to Kings, Lords, and Commons (see Dutton 2009; Hampton

2009). Since then, radio, television, and other mass media have been constituted into the Fourth Estate even though it mainly emphasized the normative role of the press, in particular in the functioning of a democratic state. With the demise of feudal society, as Dutton (2009) points out, the estates have come to be most often linked to the separation of powers into the legislative, executive, and judicial branches of government, with the press defined as the Fourth Estate through its independence and watchdog qualities of holding governments and different parts of the estate accountable. In contrast to the formal institutions of power, the power of the press constituted genuine informal power to represent the popular will of its readers, forging an indirect relationship between the powers of the state and the governed, legitimizing its role and prowess within the democratic polity. The role of journalism constituting the Fourth Estate is deemed so important that a number of states, including the United States, offer some privilege and protection to those working in the media (Felle 2016, 2). The First Amendment in the United States is designed to safeguard the press from government interference, with key Supreme Court cases setting the precedent in blocking the government from interfering with the press in allowing it to exercise editorial discretion and in publishing government secrets. In the US constitution, the First Amendment proclaims that 'Congress shall make no law . . . abridging freedom of the press' (Federal Government of the United States of America 1787). Similarly, the implicit right to be informed and the right to free expression are underpinned by the First Amendment in the US Constitution. These protections nevertheless may not be sufficient in protecting the watchdog role of journalism today with the transformations in the economic models of publishing on digital platforms. While journalism remains legally shielded from government interference, it is nevertheless not exempt from the encroachment of the private sphere with the growing prowess of technology platform companies and their ability to disseminate and control content in the era of fake news (Carroll 2018, 15).

In delineating the compounded threat of fake news unleashed through the coalescing of human and nonhuman processes online, there is then a need to review the utopian modes in which we have envisioned mass media within any healthy functioning democratic polity. In essence, the notion of the Fourth Estate is morally and ethically bound and packaged through media and its output; truth is seen as having a value proposition, as binding the other estates and holding these accountable to us. As such, truth needs to be defended and fake news combated as an entity which can manipulate our democratic processes such as elections or referendums designed to collect the popular will of the populace.

A vital characteristic of the Fourth Estate is its independence from the government and other forms of power or oppression. Its independence would

then predicate its ability to hold powers accountable or to represent public opinion by providing unbiased information or news. With the movement of news production and consumption online and the ability for ordinary citizens to post content online, the Internet as a public platform for political discussion or a form of virtual public sphere marked the early celebratory discourses of the digital realm and its possibilities for public deliberation.

There was rife speculation that the Habarmasean public sphere could be reconfigured through a vision of universal access to this medium. In terms of the initial imagination of the Internet, it propounded new ways to enact political engagement where communication can be instantaneous and reciprocal rather than the one-way communication of mass media such as print and broadcasting. Jürgen Habermas (1991) in normative terms conceived media as providing a sphere for discussion and deliberation. This space of rife political possibilities and the harnessing of the political for the augmentation of democratic politics invoked the idea of the Internet as the 'Fifth Estate' (Dutton 2009), premising the idea of a newly empowered citizen, enriched with the agency to be networked, to source the relevant information through search and social media. This birthing of the enlightened human form or the normative virtual sphere though seductive in the initial theoretical ruminations slid into new levels vulnerabilities and moral depravities which would confound society, where acritical life forms and machines will find new modes to enmesh with news production and its representation. The emergence of content sliding into a category that is neither factual yet designed though omissions and misinformation would start to further strip the patina of trust hitherto invested in the Fourth Estate in the data economy of the Internet in the coming years.

THE FOURTH ESTATE AND ITS NORMATIVE IDEALS

The notion of the Fourth Estate encapsulated the relationship between the press and readers imposing an 'educational ideal' of serving as an agency for public discussion without foreclosing the possibility of competing ideas and interests to circulate while to some measure prevailing through its own ideological orientations (Hampton 2009, 4). It signified a type of polis where citizens may not see each other face to face but conceive a wider arena of participation and public deliberations as envisaged in the Habermasean (1991) public sphere. Beyond the educational model, the press could also be conceived through a representative model signifying and capturing people's engagement with politics. Within these two paradigms, what is illustrative about the function of the press is its bridge between representative government and its populace. The press then garners its gravity from its ability to present seri-

ous issues and to remain independent to be able to function without bias and with objectivity. Though the reality of such a proposition has been questioned over time, particularly with the alignment of the press with political parties in the United Kingdom, the threat to press independence is seen as affecting the healthy functioning of a representative democratic society. Beyond our contemporary preoccupations with the challenges posed by fake news, the role and power of the media has come into scrutiny over years in terms of its political economy and equally in terms of its agenda, power, and influence.

The growing concentration of media ownership in the early twentieth century as well as the barriers to entry, the dilution of news content in favour of lifestyle coverage or what has been termed 'tabloidization', or the close alignments of the press with the agenda of the state have challenged the concept of independence as necessary in normative theorizations of the Fourth Estate. The Fourth Estate as such may be deemed as mythic (Boyce 1978) for not integrating the realities of financial survival of media firms which are inordinately expensive to run and as such may converge with political elites and the state agenda. The dumbing down of the political agenda of the press for commercial viability, to attract advertising revenue, and to increase circulation reaffirm press organizations as dependent on the market. As such, the normative ideals of the Fourth Estate often overlooked the commercialization, ownership, and the role of the state (Barnett and Gaber 2001), curbing its moral role as the watchdog.

In delineating the threats to the Fourth Estate through the encroachment of market economics or the shaping of its agenda with political power, the power of media underpinned through its moral role has been increasingly questioned since the twentieth century. In tandem, in the digital age, the notion of news as a commodity of exchange is further proscribed through different processes not previously accounted for, in the digital economy, including fake news.

NEWS WITHIN THE ECOLOGY OF 'PROSUMERISM'

Today, in terms of the Western world, we perceive the Internet and social media to be embedded in our everyday lives, merging our notion of our online and offline worlds. The Internet and social media are seen as transcending the age of gatekeeping which was a key characteristic of mass media where there were barriers to entry due to prohibitive investment costs. The Internet is seen as empowering for its ability to manufacture audiences who can not only rate and distribute content but also produce and upload through their individual orientations. The audience as producing and consuming seamlessly has resulted in a hybrid term and phenomenon known as 'Prosumerism'.

On the one end of the spectrum, this is seen as endlessly empowering and this new consumer with the possibility of creating content was reconfigured into news production as the newly reconstituted 'citizen journalists' (Allan and Thorsen 2009), particularly in crisis situations where they can contribute to news and event creation when media were not able to access these (see Ibrahim 2007). During elections, opinions from vloggers and bloggers start to create new modes of engaging with election campaigns and political figures. At points, these bloggers could be seen as influencers with their own followers willing and able to retweet and repost content, amplifying these political opinions through a strong social-media following. Thus, the Fifth Estate is also a potentially potent political force, but without the centralized institutional foundations of the Fourth Estate (Whitten-Woodring 2012).

Ulrich Beck et al. (1992) and Anthony Giddens (2010) premise the rise of modernity with the increasing role of expertise as the foundation of useful knowledge but in late modernity there is growing contestation over knowledge formation and the role of experts. In tandem with this, Andrew Keen's (2007) notion of the 'cult of the amateurs' signifies the emergence of laypeople as experts, offering opinions on technical matters or issues of importance such as medical opinion on the Internet, representing both the opportunities for publishing offered by the Internet but also the lack of gatekeeping where the layperson may not be curtailed from offering opinion on any issue. This hybrid entity of consumers as producers or 'prosumers' became a double-edged sword where audiences in the face of reduced gatekeeping online had to become more discerning about how they consumed content.

With social media and its architecture of postings conjoined to data mining and data profiling, the digital economy after the Dot-com bubble burst in 2000 witnessed an intensity to monetize the Internet. The virulent and ferocious conversion of humans, their networks, and their postings as data and profiles regenerated the digital economies in renewed modes of acute monetization. In the process, the political economy of mass media was already going through immense changes in terms of its operations with shrinking advertising revenues and dwindling subscriptions. The recovery after the bubble burst brought forth the centralization of data extraction and its assemblage as the corner stone of digital capitalism in the forthcoming years.

The implications of the intense proliferation of user-generated content (UGC), the embeddedness of social networking sites, and the ways in which news organizations had to reconfigure the modes in which they produced news meant that news as a commodity online was being increasingly enmeshed into a data and dissemination economy premising social networks and social ties as conveyors of news content. These would then bring further conundrums for the so called 'Fifth Estate' and humans fatigued through

information overload and news as offerings which came through social networks and their newsfeeds with the attendant editorial oversight or gatekeeping. The industrial collection of data and profiles and algorithms to reorder content through its rate of consumption and virality would again throw a dark shadow on the press in the digital age with fact and fabrication swimming in an avalanche of the trivial and the political without an attendant means to automatically root out the false from the authentic online.

NEWS ORGANIZATIONS AND THEIR SURVIVABILITY IN THE DIGITAL AGE

The conceptualization of the Internet as the Fifth Estate for its capacity to be a repository for news and the creation of public spheres also represented a vast degree of decentralization of news or news as a commodity. If the mass media was narrated as some form of a monolithic entity as the Fourth Estate, the pressures for news media to operate with and through the relevance of the digital economy imposed new pressures on the news media industry where news consumption was being reshaped and sustained, and subscription to news needed rethinking and restrategizing. News organizations had to reshape and resize their modes of operation in view of this with the proliferation of content online and the ways these may be shared on social networking sites. Additionally, UGC from bloggers and comments from readers were also adding to the ways in which news was being produced. As such, the veracity of news sources, the speed in which news travelled and leaped platforms, or the means through which news became viral meant that news production and consumption had morphed into a new beast where readers' input and their ability to share immaterial content through social-networking sites and social media meant they played a large part in amplifying news content globally beyond the platforms of news production or the oversight of editorial judgement.

The earlier model of centralization of the mass media transposed onto the digital age alluded to a more fractured landscape of news production. This journey into a vast landscape of content as news being offered through social-networking sites also posited the risk of audiences being more selective to news, particularly where these are unmediated by editors and professional journalists and as such create virtual 'echo chambers' where one seeks news that reaffirms one's points of view or belief system (Sunstein 2009). The narrowing of perspectives to mirror one's own views and the lack of diversity in news offerings also shone a light on the dwindling normative concepts and hitherto sacred concepts such as objectivity, balance, and nonbias valued as part of news

media ethics and their code of practice. If the demise of 'gatekeeping' was seen as empowering in the digital economy, the lack of editorial judgement and ethical standards pointed to news as entering an era of intense commodification and exchange as part of the algorithmic assemblage in the digital economy.

The repositioning of news as part of a commodity of exchange within social networking sites and mechanisms to connect people and refresh ties on an everyday basis meant that news as a form of information to educate citizens and update them on issues in society veered into another paradigm. News as a commodity was being inserted into the sociality of exchange which social-networking sites enabled. Here new online norms of transaction were sitting cheek by jowl with news in its utopian conceptualization of the 'Fourth or Fifth Estate'. New social norms and pressures were driving news consumption and production. Newsfeeds were enabling social connections and the rates of exchange and consumption were also facilitating a virality with the news. Users' observations and comments were remaking an original news story through these interactive modes and in the process, news was being recombined with algorithms designed to monetize on the rate of exchange and traffic. This remaking of news as a commodity on social-networking sites inevitably points to the vulnerability of news organizations and the pressures to retain journalistic standards in a changing political economy of news production. As such, it renewed questions of whether the rise of the Internet and related information and communication technologies, such as social media, will undermine the Fourth Estate, as an important institution in liberal democratic societies (Whitten-Woodring and James 2012).

NEWS IN THE DIGITAL AGE

The Internet is seen as fostering the emergence of nontraditional journalism and in so doing enabling a new form of 'networked Fourth Estate'. This premises a news environment where various actors and processes can shape news content. Yochai Benkler, through the example of WikiLeaks, argues that this form of networked Fourth Estate is a 'set of practices, organizing models, technologies that together come to fill the role that in the 20th Century we associated with the free press' (Freedom of the Press Foundation 2013, 29). Driven by a hybrid environment which interweaves traditional media with the digital platforms and technologies, it changes the nature of delivery and dissemination and equally widening the pool of participants and actors contributing to news-making (Dutton and Dubois, 2015). It merges and hybridizes media forms, practices, and genres, igniting struggles in the ethical practice of journalism as institutions and actors seek to maintain and defend particular internal practices (Russell 2017).

The transformations in the media industry, including the press, in the United States since the mid-1980s include increasing competition, lower spend on newsrooms, and fragmentation of the market, leading to strategies of differentiation (Benkler 2013). This, along with the Internet, led to a decentralization of production. According to Benkler (2013, 12), in this period the Internet rapidly shifted from being primarily a research and education platform to being a core element of our communications and information environment. These changes happened while news production was also gearing up to receive user-generated content from conflict zones such as the Iranian reform movement of 2009, where amateur photography and uploads onto social-networking sites meant news organizations also had to respond to the instant modes of relaying content from different parts of the world. In contextualizing WikiLeaks within the economy of the networked Fourth Estate, Benkler (2013, 16) contends that the leaks were about bringing to light direct, documentary evidence about government behaviour so that many others, professional and otherwise, can analyse the evidence and search for instances that justify public criticism. Acting through political conviction, leveraged on a combination of volunteerism, global presence, and decentralized action, Wikileaks represents an integral part of the networked Fourth Estate.

With WikiLeaks, and amidst controversies about experienced journalists versus amateurs and the overall prognosis for sound journalism in America, Benkler (2013, 19) argued that the networked Fourth Estate exemplified through WikiLeaks demonstrated the emergence of a complex relationship between a host of actors (i.e., nonprofit academic and professional groups, and small commercial professional publishers), including the lay public, presenting a networked alternative to the more traditional models of media checks and balances and new modes to enact the watchdog function. Though this was seen as a new model of collaboration and cooperation, despite the criticism of it weakening media ethics or not completely dismantling the power of traditional media in garnering the public's attention against the decentralization of the Web, the leaks could be read as an empowering moment for the networked Fourth Estate.

This networked press is an interconnected web of actors with differing agendas and values that can include both journalism and like-minded individuals who seek to check on the government, but also other actors such as technology platforms, software engineers, algorithms, and users who may not share the same values and commitments. The encroachment of platform capitalism into this mix of actors warrants renewed scrutiny, bearing in mind Google and Facebook now take most of the advertising revenue that fuelled the press in the twentieth century (Carroll 2018, 2). These seismic changes have to be reconciled with the changing landscape of journalism and the press in terms of fragility and survival. From 2001 to 2016, more than half of the news industry

jobs in the United States disappeared, with the term 'news deserts' coined to describe many communities without local journalism, representing a rapid withering of the Fourth Estate over a short time (Carroll 2018, 2).

News is then made and distributed through this enmeshing of the human and nonhuman actors with the seamless incorporation of the latter leveraging closely and intimately on personal and big data. A handful of giants or platform corporations such as Facebook, Google, Apple, and Twitter controlling industrial complexes of data can perform as information gatekeepers, as opposed to the Fourth Estate controlling the information to the public. As such, news curated and ordered through users and algorithm without a transparent account of how information reaches users emphasizes the new modality of the press working to and within the digital economy (Carroll 2018, 3). With news values determined by users and algorithms and not necessarily by news editors, news production and dissemination perform to a different imperative. The demise of gatekeeping online is supplanted through the intense activity of users or audiences or 'the people formerly known as the audience' (Rosen 2011) who are actively engaged in sharing, curating, and creating news and posts without being delimited to only consuming it.

From the preoccupation of WikiLeaks to the empowering moment of holding powers accountable through a networked economy, news online would again shift into new forms of vulnerabilities with the intense manipulation of data profiles, data breaches, and the dissemination of fake news with varying agendas and actors. In contrast to the leaks making power elites vulnerable, fake news as a process which unleashes immeasurable harm and uncertainty in terms of the magnitude of its consequences would hit at the heart of democracy, thwarting the notion of a rational public informed through a press through editorial oversight. News would now veer into an economy which targets people through their data profiles and affective states while weaponizing false information for hidden and visible agendas. News, propaganda, and falsehoods will manipulate and experiment with people's behaviour and responses to news as a form of commodity exchanged through the communication architecture of the everyday.

FAKE NEWS AS A THREAT TO DEMOCRACY

The notion of misinformation hits at the heart of a functioning democracy, as the citizen exercising her rights and participating in a polity is envisioned through her access to information that is accurate and in some manner or form verifiable. The notion of something being nontruth also affects the normative dimensions of the media, as a Fourth Estate being a crucial part of a democratic society where its circulation of information is seen as vital in informing

and educating citizens about elections and encouraging their participation in a democratic society and its processes. The prognosis of a networked Fourth Estate and its multitude of opportunities and dilemmas to a world manipulated through fake news and where our discernment between truth and the authentic becomes problematic in a content-saturated digital world plunges us into a very murky period of confusion and lack of clarity as to where the world may be heading to.

Jonathan Albright (2017, 87) points out that 'social interaction is at the heart of fake news' where news is part of a platform capitalism which enables interactivity between audiences, coupled with the disinhibitory effects of an online environment fostering echo chambers and tribalism. These modes of news consumption can be emotional as well as lending to a sharing economy that offers news to its networks as a mode of sustaining relationships. Fake news thus is dependent on the emotional and affective response to news and, as such, fake news works within the ecology of the digital realm, drawing on data mining, profiling, affective interactions, and networked communication.

Fake news in itself as a form of untruth brings its own existential challenges to both the notion of democracy and journalistic ethics. But it is not the truth position per se which is unsettling. Fake news is propelled into digital architecture where very many things can happen in tandem such that it can be difficult to extrapolate and isolate those harms individually. It is bound with the technological design of filtering, ranking, and algorithms remediating access to information while leveraging on user interactions to embed fake news such that its travel through platform capitalism is not linear but invokes different touch points which affect users both as individuals and as part of collective consuming communities. The coding of desires, preferences, and relationships between people into algorithms engineer and manipulate connections (Van Dijck 2013, 12). The rapid amplification of emotionally charged messages in social-media platforms and the use of sentiment-based sharing tools can manipulate emotions such that verification of content may be a secondary issue rather than responding to it or passing it on to others. The advancements in data mining and the pervasive gathering of information means that platform capitalism develops profiling systems on its consumers such that news becomes a commodity which is then enmeshed with these tools and technologies.

THE NETWORKED FOURTH ESTATE AND CAMBRIDGE ANALYTICA

The Cambridge Analytica scandal, as a moment of collaboration between the media and various volunteers in holding powers accountable, pointed a lens into extreme data breaches, the targeting of audiences through their profiles, and grooming through targeted and specially designed content. In 2013, the

University of Cambridge psychology professor Dr Aleksandr Kogan created an application called 'thisisyourdigitallife', which provided a personality quiz to Facebook users. When a Facebook user downloaded the app, it would start collecting that person's personal information such as profile information and Facebook activity (e.g., what content was 'liked'). Additionally, it also collected information about the user's network. With three hundred thousand people downloading the app, it managed to collect data from about eighty-seven million people (see Kozlowska 2018). Cambridge Analytica used this data to discern voter personality traits and behaviour to help conservative campaigns in the 2016 elections, including Donald Trump's campaign. This constituted a serious data breach, violating Facebook's terms of service which prohibit the transfer or sale of data 'to any ad network, data broker or other advertising or monetization-related service' (cf. Kozlowska 2018). Cambridge Analytica reportedly generated personality profiles of millions of individual voters, which meant that narrowly targeted political advertisements could be sent (Gonzalez 2017). Some described Cambridge Analytica's tools as 'mind-reading software', a 'weaponized AI (artificial intelligence) propaganda machine' that 'turned the world upside down' by saturating voters with carefully crafted messages (Gonzalez 2017).

The revelation of the Cambridge Analytica scandal put the notion of the networked Fourth Estate again under intense scrutiny. The scandal highlighted the power of platform capitalism where algorithms, intense data mining, brokerage of data, profiling, and targeting was beginning to experiment with people's news-consumption behaviour to further influence voting patterns or national referendums. With a handful of social-media firms owning vast quantities of data, the scandal revealed how machines and algorithms were co-opted in the process. As such, their ability to control content and target information mean they control not just industrial amounts of data but also content with their vast reach and wealth in terms of advertising revenue.

The Cambridge Analytica scandal threw out controversies on many fronts. Facebook members had to contend with the fact that their profiles were not secure and were used in experiments on user behaviour in either Donald Trump's US presidential campaign or the United Kingdom's Brexit referendum. At a wider societal level, new threats abound with a sense that our democratic institutions may be manipulated by targeted content, profiling, and data breaches.

This enormous power of a few platform giants in producing a whole new ecology in which news production is subsumed into then seamlessly integrated into social-media platforms, search engines, and social networks. This economy is not only about news production but works to a wider information dissemination, production, and exchange, making news swim

amidst other forms of activity, particularly political campaigning and the grooming of audiences for various agendas. Platform capital seeks to harness the connection between content and traffic. According to a report in *Wired*, each week about six hundred million people saw a news story on Facebook (Greenberg 2016) with about 50 per cent of referral traffic from publishers' sites coming from Google and 25 per cent from Facebook between February 2018 and February 2019 (cf. Carroll 2018, 15). Other sites where news is exchanged and reposted include Twitter, Reddit, Instagram, and Snapchat. Together they demonstrate the erosion and adaptation strategies of the press to survive where their advertising revenues are under siege by platform capital and they have to make new partnerships to share content with platforms to survive in the marketplace.

With the press in an economically vulnerable position, the phenomenon of fake news and its wider threats to democratic ideals and values casts a longer and much more insidious shadow on the Fourth Estate. With lower budgets for staff and cuts in the size of editorial rooms, and with platform capital appropriating a bigger role in censorship and gatekeeping, fake news repremises a long trope of vulnerability, equally in discerning what is news, propaganda, or misinformation online. Cyberspace as vulnerable to hacking and leaking also threatens our notions of state security and sovereignty. Reports of the Russian government interfering in the 2016 US presidential election by, among other things, hacking into the e-mail system of the Democratic National Committee (DNC) and releasing its e-mails constitutes foreign transgression into a national sphere and as such can be deemed a cyber-attack (Ohlin 2016, 1579). As such, spying, espionage, and hacking can then threaten the very foundations of state security. Within this premise of cyberspace as being prone to violations and transgressions, fake news enters as one of the various entities which will destabilize and thwart our political and civic institutions.

CONCLUSION

Journalism encapsulated through the normative idea of the Fourth Estate through time has faced different challenges from the industrial revolution to the digital age. Its rise, dominance, and cynicism in modernity followed through in postmodernity with new challenges for economic survival and increasing loss of trust in the press and authorities. In the digital age, new pressures abound for the press where it has had to adapt agile strategies to remain in a competitive digital marketplace. As a networked Fourth Estate in the digital economy, it has become conjoined to other actors and processes which again shed light on its diminishing ability to be a gatekeeper and retain

editorial oversight in a rapidly changing marketplace where audiences as consumers and producers were becoming crucial to its modes of dissemination.

The dominance of a handful of platform corporations handling industrial amounts of data and their prowess in the digital economy and the ways in which news is implicated through these platforms again infuses new vulnerabilities for journalism. With the conjoining of complex processes and the capitalist agenda, the appearance of fake news as an intrinsic commodity designed to manipulate human behaviour and actions unleashes another conundrum for both the press and humanity. The 'fake' parading as news, inadvertently targets the values and ideals invested in the Fourth Estate while drawing on it to reach targeted audiences and beyond. The damage it can wreak to societies and polities is consequential, but in the era of information overload and platform capitalism it has woven itself into the very fibre of everyday communication and exchange. Expelling and clearing it out of the system will be not only down to individuals or communities or even the press (as we used to know it) but to hidden algorithms and the hand of invisible capital and equally new laws and legislation coupled with public education and the will to repudiate fake news as a social ill.

REFERENCES

Albright, J. 2017. 'Welcome to the Era of Fake News'. *Media and Communication* 5 (2): 87–89. In *Citizen Journalism: Global Perspectives*, vol. 1, edited by S. Allan and E. Thorsen. New York: Peter Lang.

Allan, S., and Thorsen, E. 2017. *Citizen Journalism: Global Perspectives.* New York: Peter Lang.

Barnett, S., and Gaber, I. 2001. *Westminster Tales: The Twenty-First-Century Crisis in Political Journalism*. London: Bloomsbury.

Beck, U., Lash, S., and Wynne, B. 1992. *Risk Society: Towards a New Modernity*, vol. 17. Thousand Oaks, CA: Sage.

Benkler, Y. 2011. 'A Free Irresponsible Press: Wikileaks and the Battle over the Soul of the Networked Fourth Estate'. *Harvard Civil Rights-Civil Liberties Law Review* 46:311. https://dash.harvard.edu/bitstream/handle/1/10900863/Benkler.pdf?sequence=1.

———. 2013. 'WikiLeaks and the Networked Fourth Estate'. In *Beyond WikiLeaks*, 11–34. London: Palgrave Macmillan.

Boyce, G. 1978. 'The Fourth Estate: The Reappraisal of a Concept'. In *Newspaper History from the Seventeenth Century to the Present Day*, edited by G. Boyce, J. Curran, and P. Wingate. London: Constable.

Carroll, E. 2018. 'Platforms and the Fall of the Fourth Estate: Looking Beyond the First Amendment to Protect Watchdog Journalism'. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3300966.

Dentith, M. R. X. 2016. 'The Problem of Fake News'. *Public Reason* 8 (1–2): 65–79.
Dutton, W. H. 2009. 'The Fifth Estate Emerging Through the Network of Networks'. *Prometheus* 27 (1): 1–15.
Dutton, W. H., and Dubois, E. 2015. 'The Fifth Estate: A Rising Force of Pluralistic Accountability'. *Handbook of Digital Politics*, 51–66.
Felle, T. 2016. Digital Watchdogs? Data Reporting and the News Media's Traditional 'Fourth Estate' Function'. *Journalism* 17 (1): 85–96.
Freedom of the Press Foundation. 2013. *Transcripts from Chelsea Manning's Trial: In the Matter of: United States vs. PFC Bradley E. Manning (Vol. 17, afternoon session)*. San Francisco, CA: Freedom of the Press Foundation. https://freedom.press/chelsea-manningtranscripts.
Giddens, A. 2010. *Modernity and Self-Identity: Self and Society in the Late Modern Age*. Second edition. Cambridge: Polity Press.
Gonzalez, R. J. 2017. 'Hacking the Citizenry?: Personality Profiling, "Big Data" and the Election of Donald Trump'. *Anthropology Today* 33 (3): 9–12.
Greenberg, J. 2016. 'Facebook Has Seized the Media, and That's Bad News for Everyone but Facebook'. *Wired*. https://www.wired.com/2016/04/facebook-seized-media-thats-bad-news-everyone-facebook/.
Habermas, J. 1991. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Boston: MIT Press.
Hampton, M. 2009. 'The Fourth Estate Ideal in Journalism History'. In *The Routledge Companion to News and Journalism*, 47–56. New York: Routledge.
Ibrahim, Y. 2007. The Technological Gaze: Event Construction and the Mobile Body. *M/C Journal* 10 (1). http://journal.mediaculture.org.au/0703/03-ibrahim.php.
Keen, A. 2007. *The Cult of the Amateur: How Today's Internet Is Killing Our Culture*. New York: Crown Business.
Kozlowska, I. 2018. 'Facebook and Data Privacy in the Age of Cambridge Analytica'. The Henry M. Jackson School of International Studies, 30 April. https://jsis.washington.edu/news/facebook-data-privacy-age-cambridge-analytica/.
Ohlin, J. D. 2016. 'Did Russian Cyber Interference in the 2016 Election Violate International Law?' *Texas Law Review* 95:1579.
Rosen, J. 2011. 'The People Formerly Known as the Audience'. In *The Social Media Reader*, 13–16. New York: NYU Press. https://www.huffingtonpost.com/jay-rosen/the-people-formerly-known_1_b_24113.html.
Russell, A. 2017. *Journalism as Activism: Recoding Media Power*. New York: Wiley.
Sustein, C. R. 2009. *Republic.com 2.0*. Princeton, NJ: Princeton University Press.
United Nations. 1966. *International Covenant on Civil and Political Rights*. Geneva: United Nations Publications.
US Government. 1787. *Constitution of the United States of America*. Washington, DC: US Government Printing Office.
Van Dijck, J. 2013. *The Culture of Connectivity: A Critical History of Social Media*. Oxford: Oxford University Press.
Whitten-Woodring, J., and James, P. 2012. 'Fourth Estate or Mouthpiece? A Formal Model of Media, Protest, and Government Repression'. *Political Communication* 29 (2): 113–36.

3

Political and Social Impact of Digital Fake News in an Era of Social Media

Dr Fadi Safieddine

In 2014, the World Economic Forum linked fake news to the top-ten threats to society (World Economic Forum 2014). Despite the long history behind fake news, several reasons are currently helping the growth of its importance. First, the methods by which one can enter the media industry has become easier than any time before, where individuals with no computing or journalistic skills are able to create a professional-looking news outlet online. For example, many web-hosting platforms provide ready and customizable media templates to mimic well-established news outlets that cover health, politics, celebrity, and local news. Thereafter and supported by mass tribalists and/or cyber-bots, they attract traffic that in turn allows the monetize publishing of content by using advertising platforms. Due to reputational concerns and in compliance with ethical journalistic practices, mass media outlets guard against reporting fake news. Secondly, social-media platforms are considered as suitable for disseminating fake news, and their use has risen sharply. Fake news is not restricted to political content; some of the reported concerns around media news is that it can be harmful and deadly. Fake news is linked to an increase in hate crime, youth crimes, the lynching of innocent people due to mistaken identity, election fraud, and more. This chapter focuses on the social and political impact of fake news. Enlisting conspiracy theories regarding fake medical news in Poland, the chapter traces how misinformation affected public belief in the dissemination of false news about cancer, vaccinations, and common diseases. The chapter examines the effect of social media and fake news in the 2016 US election, Brexit, and last but not least, the role of Russians. Tribalists and other interest groups and fake-news spread are also discussed based on

research conducted at the University of Oxford. The main emphasis of this chapter is to demonstrate cause and effect through case examples.

MEDICAL FAKE NEWS: FROM CANCER CURES TO ANTIVACCINATION

According to Fu et al. (2016) and Fahy et al. (2014), health information that can be found online is often unreliable. As early as 2013, 60 per cent of adults in the United States looked for health information online (Fox and Duggan 2013), and this number is bound to have increased since. There have been studies investigating user reactions to health information online, and these show that most users do not assess the reliability of health information they see online (Allam, Schulz, and Nakamoto 2014; Knapp et al. 2011; Kutner et al. 2006). Specifically, several researchers claim that individuals do not consider the credibility of online content regarding vaccination (Nan and Madden 2012; Betsch et al. 2010, 2013; Allam et al. 2014). Thus, health fake news is said to be a rising potential threat to public health. However, the magnitude of the threat that these phenomena pose remains hard to quantify. A pilot study conducted by Waszak, Kasprzycka-Waszak, and Kubanek (2018) examined the top health-related fake-news stories published in the Polish language on social media. Researchers assessed fake news through shared health web links in the Polish language published between 2012 and 2017 by using an application called BuzzSumo. In their research, they reviewed eighty of the most frequently shared pages, and the objective was to find the presence of fake news on each page. In general, 98 to 100 per cent of the shares and engagements were found on Facebook. Waszak et al. (2018) analysed the topics of public interests with an unequal distribution (total shares, means, in thousands): cancer (34), neoplasm (18), vaccinations (15), heart attack (7), AIDS/HIV (7), hypertension (5), stroke (5), and diabetes (2). The study found that vaccination was the most contaminated topic in terms of false reporting (90 per cent), followed by heart attack and HIV/AIDS at 70 per cent. Altogether, links containing fake news were shared 451,272 times between 2012 and 2017 and accounted for 40 per cent of the studied material. The study by Waszak et al. (2018) also found that 40 per cent of the most frequently shared links contained text they classified as fake news. These were shared more than 450,000 times. The sources of these fake-news links provided interesting clues; in one case, 20 per cent of these fake-news links came from one source. For example, the most frequently shared link in the studied population was a story about a miracle cancer treatment, which claimed to cure any cancer in just a few days. This completely fallacious article was shared nearly 65,000

Figure 3.1. Social media fake news copied and translated into multiple languages and formats.

times. In fact, our research shows that this particular claim has been translated and disseminated in multiple languages, as demonstrated in figure 3.1.

Public health is confronted with a risk of patients' exposure to fallacious and misleading information. It is believed that such a phenomenon is affecting health literacy and is spreading medical conspiracy theories. Indeed, one study confirmed that health fake news is resulting in behavioural changes and is now emerging as a serious threat to public health (Schwitzer 2017). For example, media, health organizations, and medical reports have linked the increase in measles infection rates and death to fake news shared on social media. The United States experienced two major outbreaks of measles in Minnesota (CDC 2017) and New York (CDC 2019a). However, the bigger picture paints a much more worrying trend. By 8 August 2019, the data from the US Centers for Disease Control and Prevention shows the outbreak of measles in thirty states and reported cases set to double previous ten-year records (see figure 3.2).

Many parents regard vaccination to be one of the most common preventive health measures, but others do not perceive it to be safe and necessary. In their

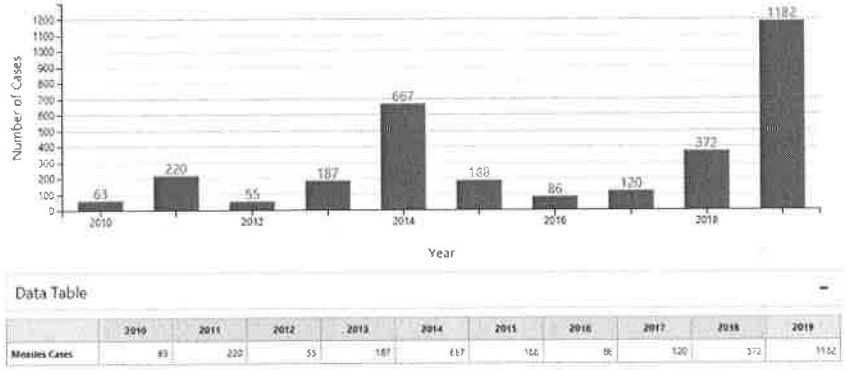

Figure 3.2. Measles outbreak cases from 2010 to August 2019, USA (CDC, 2019b).

research, Dube et al. (2015) **[AQ]**looked at the factors that influence parents' decisions about vaccination, studying the impact of fake-news dissemination about vaccination on social media. In high-income countries where immunization programs are well established, the successful deployment of vaccines has caused a decline in vaccine-preventable diseases (VPD), which has led to a lack of direct experience with these types of illnesses by parents (Larson et al. 2011). VDP, however, pose a more imminent threat to health in low- and middle-income countries (Taylor 2015). Of course, the controversy around the safety of vaccination has been around long before the arrival of the Internet. Back in 1802, a British satirist, James Gillray, mimicked a scene at the Smallpox and Inoculation Hospital at St. Pancras, as shown in figure 3.3. The cartoon shows the cowpox vaccine being given to terrified young women, and cows emerging from different parts of peoples' bodies.

AQ: Please add an entry to the References for Dube et al. 2015.

Done

The 1970s saw a pertussis (whooping cough) vaccine controversy that is believed to have started the modern conspiracy theories about immunization (Baker 2008). This was further exacerbated when Great Ormond Street Hospital for Sick Children in the United Kingdom published a report that suggested thirty-six children suffered serious neurological problems following DTP (diphtheria, tetanus, and pertussis combination vaccines) (Kulenkampff, Schwartzman, and Wilson 1974). There was an outpouring of concern, and it was around that time that the Association of Parents of Vaccine Damaged Children was founded in the United Kingdom, drawing further attention to what it saw as safety problems with the pertussis vaccine. In the following years, the United Kingdom experienced a major decline in vaccination rates,

Figure 3.3. Anti-vaccine cartoon, 1802. James Gillray, The Punch.

from 77 per cent to 33 per cent, and three major epidemics resulting in over one hundred thousand cases and at least thirty-six deaths (Baker 2008). The United States experienced similar antivaccine problems. In 1982, the Emmy-winning documentary *DTP: Vaccination Roulette* alleged that the pertussis component in the DTP vaccine was causing severe cerebral damage to children. Similar controversies have surfaced in Asia, Africa, and the Middle East (Feldman-Savelsberg, Ndonko, and Schmidt-Ehry (2000). Some of the arguments that were used in the 1800s are still being used now online and on social media to discourage parents from vaccinating their children. Some of these arguments include that the vaccines are either ineffective or could cause harm; vaccines are made by large pharmaceutical companies for profit only; harm caused by vaccines is being hidden from the public; forced vaccination is a violation of human and civil rights; alternative therapy works better (Wolfe and Sharp 2002). There is very little doubt that the Internet and social-media platforms have amplified the power of the antivaccination movement, otherwise known as 'Anti-Vax' (Kata 2012). That said, these groups have become well established in claiming that they are not antivaccination per se, but rather pro 'safe' vaccine and pro 'informed-decision' about vaccination (Offit 2015). Two tactics used in antivaccination marketing are ads and critical posts. Figure 3.4 illustrates a screenshot of an ad on Facebook

Figure 3.4. Screenshot of advertising of fake news on Facebook (Chiou and Tucker, 2018, p. 6).

for a fake-news story about vaccines. The word 'sponsored' that appears on the top the screenshot indicates that post is an advertisement that in this case is linked to vaccinesrevealed.com, which claims that vaccines are 'neither safe nor effective'.

An example of antivaccination social-media posts is shown in figure 3.5, where a post appears in an antivaccine group on Facebook, claiming that there is a 'significant link' between children who are vaccinated and mental disorders. Each user who sees the post may react to it by liking it, writing a comment on it, or maybe sharing it. The post suggests a study has found a link, but a closer look would show that the title of the post is an exaggeration. In fact, the post links to a study that shows a very small correlation between the MMR vaccine and the subsequent diagnosis of brain-related disorders. Nevertheless, the article concludes that children should not be vaccinated because of high health risk that vaccines supposedly present. The study received heavy criticism for the very small sample of only twelve, lack of a solid methodology, and the fact that ten of the twelve original authors wrote a retraction (Rao and Andrade 2011). Despite the overwhelming body of research that followed, and assurances from multiple health agencies and medical bodies that showed MMR vaccine is safe, the content of that article continued to spread, mixing with conspiracy theories and displaying typical characteristics we see today of fake news.

Furthermore, Chiou and Tucker's (2018) study shows that advertising has an impact on the popularity of fake news on social media. To combat the

Figure 3.5. Screenshot of post on anti-vaccine Facebook groups (Chiou and Tucker, 2018, p. 10).

popularity of health fake news on social media, the authors recommend the development of a policy to ban advertising health news on social media.

SOCIAL MEDIA AND FAKE NEWS IN THE 2016 US PRESIDENTIAL ELECTION

The impact of fake news has led to an increased concern as to its impact on democratic systems. In the aftermath of several fake-news stories going viral, researchers speculated how these stories might have had an impact on voters in the 2016 US election (Allcott and Gentzkow 2017). There are tangible reasons for this concern. First, recent surveys show that social media is the source of 62 per cent of US adults for getting the news (Gottfried and Shearer 2016; Eady et al. 2019; Middaugh 2019). Secondly, Facebook has been the primary platform for those sharing fake news. And thirdly, a large proportion of people who see fake news claim that they believe it (Allcott and Gentzkow 2017). Allcott and Gentzkow (2017) presented a list of twelve conspiracy theory statements with political implications that have been dispersed during the

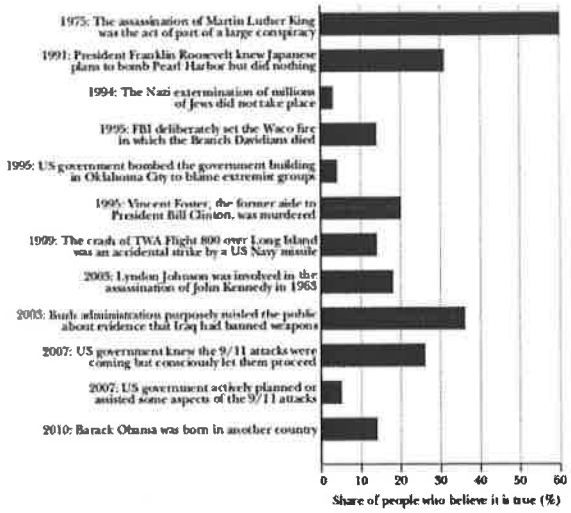

Figure 3.6. Polling data on the percentage of people who believe statements that are not true (Allcott and Gentzkow, 2017, p. 215).

past half-century, while Bowman and Rugg (2013) used polling data to show what proportion of people believe each statement is true. A summary of their findings is presented in Allcott and Gentzkow (2017, 215) (see figure 3.6).

Some websites intentionally publish fabricated and misleading international news articles, and their names often originate from the name of legitimate news organizations; for example, 'denverguardian.com'. Following the 2016 election, anecdotal reports were established to provide the picture behind these sites. BuzzFeed and the *Guardian*, in another investigation, found that most of these articles originated from Veles, Macedonia. This small town was linked to more than one hundred websites publishing fake-news articles. These websites coordinate in sharing and propagating each other's fake-news articles, allowing them to trend higher and go viral (Subramanian 2017). Townsend (2016) reports that a twenty-four-year-old Romanian man was in charge of running Endingthefed.com, which was responsible for four out of ten popular fake-news stories on Facebook. In another study, Sydell (2016) reports that many fake-news sites such as WashingtonPost.com.co, NationalReport.net, and USAToday.com.co are owned by Disinfomedia—a US-based company—which claims to employ between twenty and twenty-five writers. There are two primary motivations for those writing and publishing fake-news stories. The first motivation is ideological. The fake-news writers provide false news to favour their preferences. For example, in the case of the Romanian man who managed endingthefed.com, he claims that he started to

the website to support Donald Trump's campaign (Townsend 2016). The second motivation is financial, where significant advertising revenue is gained when fake-news stories go viral on social media and users click on the story, which takes them to the original site. Teenagers in Veles, for example, created stories favouring both Trump and Clinton that earned them tens of thousands of dollars (Subramanian 2017). Researchers continue to study the impact of fake news on the US presidential elections. There is a new indication as to the influence of total exposure to fake news. During the US presidential elections of 2016, the average US adult read and remembered on the order of one or several fake-news stories with a stronger influence on pro-Trump stories than pro-Clinton stories (Allcott and Gentzkow 2017). The question then: *To what degree did the election results depend on the effectiveness of fake-news exposure in changing the way people vote*? For instance, researchers demonstrate that for each TV campaign ad, voters shares would have been changed by approximately 0.02 percentage points (Spenkuch and Toniatti 2016). Spenkuch and Toniatti (2016) assessed this fact against their database of fake news. They determined that it would have changed vote share by an amount on the order of hundredths of a per centage point. Thus, if one fake-news article was about as persuasive as one TV campaign, the impact on US voters would have been measurable. Therefore, one could conclude that this change could have affected the US presidential election results.

UNITED KINGDOM LEAVING THE EU REFERENDUM OF 2016

During the 2016 UK referendum, otherwise known as the Brexit vote, it was found that political bots played a small but strategic role in shaping Twitter conversations (Hern 2017). Hern (2017) found that a family of hashtags associated with the argument for leaving the European Union dominated, and that almost a third of all the such messages were generated from less than 1 per cent of sampled accounts. In recent years, reports claiming varying levels of Russian activity in the Brexit referendum have appeared in the media. The main research into Russian interference during the Brexit referendum was carried out at the British universities Swansea University, City University, University of Edinburgh, and the University of Oxford.

The research group based at Swansea collected Brexit data on Twitter between May and August 2016, and showed that public opinion on Twitter predicted the outcome of the referendum (Hern 2017). To investigate the spread of public opinion before, during, and after the referendum, researchers used the Twitter streaming API with *#Brexit* to collect a total of twenty-eight million tweets between 24 May and 17 August 2017. The data sample contained

details such as usernames, date of account creation, date of tweeting, number of retweets, and number of followers. Then they examined both proleave tweets and proremain tweets by extracting the relevant hashtags. The results helped to classify 20 per cent of the accounts as cyber-bots, using three criteria that looked at an abnormal time of tweeting (00:00 hours–06:00 hours BST), an abnormal number of tweets per day, and tweets from platforms. Their research indicates that bots on Twitter can influence public opinion, and that the outcomes of major events that involve public participation can be predicted by studying bot interactions on Twitter. Using the list of cyber-bot accounts that Twitter has made public and linking these to the data sample collected during Brexit, media reports claim that they have discovered 150,000 bot accounts linked to Russia that were active during Brexit (Payne 2017). In this particular case, researchers concluded that cyber-bots did influence the outcome of the Brexit referendum. However, the researchers have yet to produce a formal publication where these claims linking the accounts to the Russian-based Internet Research Agency have been verified.

The researchers at City University also used the Twitter-streaming API to search for Brexit-related tweets between April 2016 and August 2016 (Bastos and Mercea 2019). Bastos and Mercea investigated cyber-bot activity to establish the following: (1) dissemination of hyperpartisan and polarizing content; (2) to characterize the lifecycle of a cyber-bot—with a period of high posting activity followed by a sharp drop in activity levels; (3) to determine the influence cyber-bots had on Brexit discussions; (4) to determine if cyber-bots caused faster cascades than human users, looking at the impact through spread, reach, and intensity; and (5) to distinguish cyber-bots from human users and other bots in the botnet, by measuring if the botnet was located in a network of human users or if was restricted to only to clusters of botnets. The study analysed ten million tweets that referenced the referendum using relevant hashtags, focusing on a period between 10 June 2016 and 10 July 2016. The researchers used several metrics like the presence or absence of geographical metadata, account-creation date, followers-to-following ratio, and activity levels to distinguish human users from cyber-bots. According to these researchers, some positive predictors of cyber-bot activity include tweets to user, uncommon words in username, use of web interface to post content, retweet reciprocity, and account-creation date. Out of a total of 794,949 users, only 30,122, (37 per cent) were located in the United Kingdom, and 40,031 users have either been deleted, changed to private, blocked, deactivated, or changed their usernames since the referendum. Furthermore, 17 per cent of tweets had the keyword 'remain' while 31 per cent contained the word 'leave'. Bots were eight times more likely to tweet leave slogans than other Twitter users, and 63 per cent of URLs in cyber-bot

tweets do not exist any longer. The researchers have identified 13,493 cyber-bot accounts with Russian links.

In a study conducted by researchers at the University of Oxford, Narayanan et al. (2017) sought to identify highly automated accounts that were either linked to Russia or tweeting pro-Russian content. In addition, the study looked at classifying the type of content that was shared by these accounts and analysed YouTube videos shared by a sample of Twitter users who were active during the referendum. Finally, the study looked at classifying the type of political news stories shared into high-quality, professional news, extremist, sensationalist, conspiratorial content, masked commentary, fake, and other forms of junk news and news stories linked to Russian sources. Their data set contained over five million Tweets collected between the 6–12 June and again 17–23 June 2016, using a combination of proleave, proremain, and neutral hashtags. This sampling strategy yielded tweets from 1,112,403 distinct Twitter user accounts. First, the researchers examined the Brexit-related Twitter traffic from a list of automated accounts linked to Russia. They found that altogether only 105 of these accounts produced any tweets or retweets about Brexit, and they generated only 15,964 tweets in that sample, which represents less than 0.3% of the total traffic. The researchers also found that overall, only 3 per cent of the Russian accounts initially 'outed' by Twitter generated any traffic during the Brexit debate. Second, looking at the news content shared during these Brexit conversations on Twitter, it appeared that the proportion of content from known Russian sources like Russia Today or Sputnik got little traction over Twitter. Only 511 of the 87,169 links shared were to these sources, representing under 0.6 per cent of the data. The largest proportion of content shared by Twitter users interested in UK politics came from professional news organizations, which accounts for 63.9 per cent of the URLs shared on Twitter in their sample. Within that, the *Telegraph* was most popular, with 8.1 per cent of professional news coming from this source. The BBC came next with 6.2 per cent of links directing to its website. Third, the researchers took a dedicated look at the YouTube videos shared by Twitter users during the referendum. Using their standard dictionary, they labelled over 11.0 per cent of the content like fake news and only 8.0 per cent of the content as sourced from professional news channels. Over 22.7 per cent of the content was created by citizens or civil society groups, and nearly 5.3 per cent of videos were classified as political humour or satire, and finally, only 1.4 per cent of the videos were linked to Russian news sources.

All four studies, done independently of each other, concluded that fake news, partly propped by cyber-bots, played a role in the 2016 United Kingdom leaving the European Union referendum. The actual impact is still not well established. One could link these four studies with the work of Spenkuch and

Toniatti (2016) on the impact of political advertising on public opinion to draw some anecdotal inference that would suggest that, in all likelihood, fake news did impact the results of that referendum; although further research is needed.

PROPAGATION OF LOCAL RUMOURS AND FAKE NEWS

In the last decade, communal violence linked to fake news on social media has been reported across the globe. Narrain (2017) studied several cases of public violence in India, in particular, those in which social-media apps played a major role in the spread of a rumour. For example, the Indian government banned text-message services in 'Northeastern exodus' where rumours were circulated on the upcoming violence against people of northeast origin. In another example, in 2014 Mohsin Shaikh was lynched by a mob, and the news spread all over social media (Narrain 2017), and subsequently in 2018 WhatsApp was said to be responsible for several lynching that happened in India. Researcher Chinmayi Arun (2019) argues that there are two types of problems associated with a rumour propagated on WhatsApp. First is the disinformation, and second is an incitement to violence. Chinmayi further describes how fake news has led to several incidents of lynching in India linked directly to content shared on WhatsApp. The government accused WhatsApp as being the main cause of the problem. The ministry of electronics and information technology began giving notices to WhatsApp to disable some of its features in India (Arun 2019). This resulted in the Indian government putting pressure on WhatsApp services in the country, and WhatsApp rectified some of its practices in response to these pressures. In all fairness, WhatsApp cannot be the only means of the spread of rumour and violent incidents. There must also be other factors that have an essential role in this regard; yet the question is to what extent has WhatsApp contributed to this concern of the Indian government. Simon et al. (2016) explain that people nowadays rely more on instant-messaging applications such as WhatsApp compared to other social-media apps because users feel these apps are safer and are a private means of communication. Another feature that encourages the propagation of news on WhatsApp is the fact that the identity of the original sender of a forwarded message is not known to anyone, and therefore the source from which a message has been created and forward to several users is not visible (Simon et al. 2016). While the Indian government succeeded in making WhatsApp indicate a forwarded message as 'flagged', nevertheless identification of the original source of a message remains a problem. WhatsApp has also limited quick forwards to five chats at once. WhatsApp has been forced to trace and block users who share harmful information and provide a method for the Indian government to track those users (Arun 2019).

However, there is no guarantee that the proposed solutions and requested modification in WhatsApp services would have a significant impact on the decrease in the propagation of violence via social-media instant-messaging services. On the other side of the argument, Narrain (2017) suggests that fake news is not the main cause of lynching at all, rather the culprits are those who incite violence and the amplification of harmful speech. Narrain suggests that despite social media providing a platform that can be potentially used for the circulation of rumours, the issue of critical concern is the potential collaboration of WhatsApp with the Indian government under the banner of security and the implications of this on privacy and civil liberties. We sympathize with both views, but we see parallels in trying to pin the blame on the tool or the user and the US debate on gun control. Do you blame the gun or the shooter? One should not deny that the tool has a major influence in amplifying the impact and destruction, be that lynching or killing, without discounting how it can mediate user behaviour or its potential for communal violence.

CONCLUDING REMARKS

The suggestion that fake news is just part of the Internet and part of the process democracies have to learn to adapt to and develop strategies of containment for seems far too sanguine considering the pervasive and extensive risk and harms it could pose to society and humanity through realms such as health, elections, referendums, and communal violence. In this chapter, we presented a sample of what will likely be a repeated experience in many countries where social media facilitates groups with tribal interests to get together, in some cases funded and supported by external powers. The reaction from India may be seen as harsh and underlines a threat to individual freedoms. On a similar scale, one cannot completely blame or acquit social-media platforms, but the potential solution may lie in tweaking technology such that we can either slow or stop the propagation of fake news while ensuring that freedom of expression remains protected. This is a tenuous call, particularly when today, private social-media giants have the power to take down or censor posts signalling power and influence beyond national and transnational boundaries.

REFERENCES

Allam, A., Schulz, P. J., and Nakamoto, K. 2014. 'The Impact of Search Engine Selection and Sorting Criteria on Vaccination Beliefs and Attitudes: Two Experiments Manipulating Google Output'. *Journal of Medical Internet Research* 16 (4): e100.

Allcott, H., and Gentzkow, M. 2017. 'Social Media and Fake News in the 2016 Election'. *Journal of Economic Perspectives* 31 (2): 211–36.

Arun, C. 2019. 'On WhatsApp, Rumours, and Lynchings'. *Economic and Political Weekly* 54 (6), 30–35.

Baker, J. P. 2008. 'Mercury, Vaccines, and Autism: One Controversy, Three Histories'. *American Journal of Public Health* 98 (2): 244–53.

Bastos, M. T., and Mercea, D. 2019. 'The Brexit Botnet and User-Generated Hyperpartisan News'. *Social Science Computer Review* 37 (1): 38–54.

Betsch, C., Renkewitz, F., and Haase, N. 2013. 'Effect of Narrative Reports about Vaccine Adverse Events and Bias-Awareness Disclaimers on Vaccine Decisions: A Simulation of an Online Patient Social Network'. *Medical Decision Making* 33 (1): 14–25.

Betsch, C., Renkewitz, F., Betsch, T., and Ulshöfer, C. 2010. 'The Influence of Vaccine-Critical Websites on Perceiving Vaccination Risks'. *Journal of Health Psychology* 15 (3): 446–55.

Bowman, K., and Rugg, A. 2013. 'Public Opinion on Conspiracy Theories'. *AEI Paper and Studies*.

CDC. 2017. *Measles Outbreak—Minnesota April–May 2017*. Washington, DC: Center for Disease Control and Prevention. https://www.cdc.gov/mmwr/volumes/66/wr/mm6627a1.htm.

———. 2019a. *Measles Outbreaks from Imported Cases in Orthodox Jewish Communities—New York and New Jersey, 2018–2019*. Washington, DC: Center for Disease Control and Prevention. https://www.cdc.gov/mmwr/volumes/68/wr/mm6819a4.htm.

CDC. 2019b. *Measles Cases and Outbreaks*. Washington, DC: Center for Disease Control and Prevention. https://www.cdc.gov/measles/cases-outbreaks.html.

Chiou, L., and Tucker, C. 2018. *Fake News and Advertising on Social Media: A Study of the Anti-Vaccination Movement* (No. w25223). Cambridge, MA: National Bureau of Economic Research.

Eady, G., Nagler, J., Guess, A., Zilinsky, J., and Tucker, J. A. 2019. 'How Many People Live in Political Bubbles on Social Media? Evidence from Linked Survey and Twitter Data'. *SAGE Open* 9 (1): 2158244019832705.

Fahy, E., Hardikar, R., Fox, A., and Mackay, S. 2014. 'Quality of Patient Health Information on the Internet: Reviewing a Complex and Evolving Landscape'. *The Australasian Medical Journal* 7 (1): 24.

Feldman-Savelsberg, P., Ndonko, F. T., and Schmidt-Ehry, B. 2000. 'Sterilizing Vaccines or the Politics of the Womb: Retrospective Study of a Rumor in Cameroon'. *Medical Anthropology Quarterly* 14 (2): 159–79.

Fox, S., and M. Duggan. 2013. *Pew Internet and American Life Project*. Pew Research. https://www.pewinternet.org/2013/01/15/health-online-2013/.

Fu, L. Y., Zook, K., Spoehr-Labutta, Z., Hu, P., and Joseph, J. G. 2016. 'Search Engine Ranking, Quality, and Content of Web Pages That Are Critical Versus Noncritical of Human Papillomavirus Vaccine'. *Journal of Adolescent Health* 58 (1): 33–39.

* Dude, E., Vivion, M. and MacDonald, N.E., 2015. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. Expert review of vaccines, 14(1), pp. 99-117.

italic

Gottfried, J., and Shearer, E. 2016. *News Use Across Social Medial Platforms 2016*. Pew Research Center. http://www. journalism.org/2016/05/26/news-use-across-social-media-platforms-2016.

Hern, A. 2017. 'Russian Troll Factories: Researchers Damn Twitter's Refusal to Share Data'. *Guardian*, 15 November. https://www.theguardian.com/world/2017/nov/15/russian-troll-factories-researchers-damn-twitters-refusal-to-share-data.

Independent, The. 2017. 'Brexit: Russian Twitter Accounts Tweeted 3,468 Times About Brexit'. *Independent*, 15 November. http://www.independent.co.uk/news/uk/politics/brexit-latest-russian-twitter-accounts-eu-independence-referendum-tweets-influence-result-a8055746.html.

Kata, A. 2012. 'Anti-Vaccine Activists, Web 2.0, and the Postmodern Paradigm—An Overview of Tactics and Tropes Used Online by the Anti-Vaccination Movement'. *Vaccine* 30 (25): 3778–89.

Knapp, C., Madden, V., Wang, H., Sloyer, P., and Shenkman, E. 2011. 'Internet Use and eHealth Literacy of Low-Income Parents Whose Children Have Special Health Care Needs'. *Journal of Medical Internet Research* 13 (3): e75.

Kulenkampff, M., Schwartzman, J. S., and Wilson, J. 1974. 'Neurological Complications of Pertussis Inoculation'. *Archives of Disease in Childhood* 49 (1): 46–49.

Kutner, M., Greenburg, E., Jin, Y., and Paulsen, C. 2006. *The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy*. NCES 2006-483. Washington, DC: National Center for Education Statistics.

Larson, H. J., Cooper, L. Z., Eskola, J., Katz, S. L., and Ratzan, S. 2011. 'Addressing the Vaccine Confidence Gap'. *The Lancet* 378 (9790): 526–35.

Middaugh, E. 2019. 'Teens, Social Media, and Fake News'. In *Unpacking Fake News: An Educator's Guide to Navigating the Media with Students*, edited by W. Journell and J. Stoddard, 42. New York: Teacher's College Press.

Nan, X., and Madden, K. 2012. 'HPV Vaccine Information in the Blogosphere: How Positive and Negative Blogs Influence Vaccine-Related Risk Perceptions, Attitudes, and Behavioral Intentions'. *Health Communication* 27 (8): 829–36.

Narayanan, V., Howard, P. N., Kollanyi, B., and Elswah, M. 2017. 'Russian Involvement and Junk News During Brexit'. *The Computational Propaganda Project: Algorithms, Automation and Digital Politics*. University of Oxford. https://comprop.oii.ox.ac.uk/research/working-papers/russia-and-brexit.

Narrain, S. 2017: 'Dangerous Speech in Real Time: Social Media, Policing, and Communal Violence'. *Economic and Political Weekly* 52 (34) 26 August.

Offit, P. A. 2015. *Deadly Choices: How The Anti-Vaccine Movement Threatens Us All*. New York: Basic Books.

Ołpiński, M. 2012. Anti-Vaccination Movement and Parental Refusals of Immunization of Children in the USA'. *Pediatria Polska* 87 (4): 381–85.

Payne, A. 2017. 'Russia Used a Network of 150,000 Twitter Accounts to Meddle in Brexit'. *Business Insider*. https://www.businessinsider.com/russia-used-twitter-accounts-to-meddle-in-brexit-investigation-shows-2017-11?r=US&IR=T.

Rao, T. S., and Andrade, C. 2011. 'The MMR Vaccine and Autism: Sensation, Refutation, Retraction, and Fraud'. *Indian Journal of Psychiatry* 53 (2): 95.

Schwitzer, G. 2017. 'Pollution of Health News'. *British Medical Journal*: j1262. doi:10.1136/bmj.j1262.

Simon, T., Goldberg, A., Leykin, D., and Adini, B. 2016. 'Kidnapping WhatsApp—Rumors During the Search and Rescue Operation of Three Kidnapped Youth'. *Computers in Human Behavior* 64:183–190.

Spenkuch, J. L., and Toniatti, D. 2016. 'Political Advertising and Election Outcomes'. *Kilts Center for Marketing at Chicago Booth–Nielsen Dataset Paper Series*, 1–46.

Subramanian, S. 2017. 'Inside the Macedonian Fake-News Complex'. *Wired*, 15 February.

Sydell, L. 2016. 'We Tracked Down a Fake-News Creator in the Suburbs. Here's What We Learned'. *National Public Radio*, 23 November. http://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs.

Taylor, S. A. 2015. 'Culture and Behaviour in Mass Health Interventions: Lessons from the Global Polio Eradication Initiative'. *Critical Public Health* 25 (2): 192–204.

Townsend, T. 2016. 'Meet the Romanian Trump Fan behind a Major Fake News Site'. Inc.com. http://www.inc.com/tess-townsend/ending-fedtrump-facebook.html.

Waszak, P. M., Kasprzycka-Waszak, W., and Kubanek, A. 2018. 'The Spread of Medical Fake News in Social Media—The Pilot Quantitative Study'. *Health Policy and Technology* 7 (2): 115–18.

Wolfe, R. M., and Sharp, L. K. 2002. 'Anti-Vaccinationists Past and Present'. *Britsh Medical Journal* 325 (7361): 430–32.

World Economic Forum. 2014. 'The Rapid Spread of Misinformation Online'. *Outlook on the Global Agenda 2014.* http://reports.weforum.org/outlook-14/top-ten-trends-category-page/10-the-rapid-spread-of-misinformation-online/.

4

Case Study: Fake News and Threat to Brazilians' Health

Gheorghita Ghinea[1]
Suzana G. Cardoso[2]
Marian C. M. Caparelli[3]

In the contemporaneity of communicational phenomena, fake news emerges as a problem that has been causing a change in behaviour and the decision-making power of users navigating the various social media. The volume of real-time information circulating in the digital environment makes verifyifying the veracity of information a difficult task. In this study, a survey of the typology of fake news, taking as reference the information disorder conceptual framework proposed by Wardle and Derakshan (2017, 20), examines how mis-, dis-, and mal-information intersect around the concepts of falseness and harm. This chapter explores a fake-news category widely present in recent years in Brazil. It is a type frequently broadcasted through Twitter, YouTube and WhatsApp groups: health fake news. In a preliminary fake-news survey on the Internet, the Health Ministry website—the central organ responsible for federal policies relating to Brazilian public health—created, in August 2008, the campaign *Health Free of Fake News*. It aims to check messages sent by users through the institution's WhatsApp channel. The misleading data sent to be checked by the Health Ministry is related to various diseases, like measles, yellow fever, HPV, diabetes, and cancer, as well as supposed treatments like miraculous teas and nutritional restrictions. Fake-news incidence in the Brazilian health field is high and has a huge impact on social groups' behaviour towards public health, including a decrease in the number of citizens immunized from disease through vaccination. Health Ministry data show that vaccination rates in both the adult and child immunization calendars are below the coverage goal (Revista 2018). ~~The Brazilian federal government official report has also been added to the decision of the clipping of the empirical phenomenon of fake news and public health (Brasil 2018).~~ **[AQ]**Facing this reality, the research focus of this study

2018

AQ: The meaning of the preceding sentence is not clear.

is fake news about vaccination, investigating the harm caused to society by misleading messages shared on social media. This research aims to investigate how fake news can affect public health subjects in Brazil, which has a rate of 147,385,581 million (90.98 per cent) literate and 14,604,155 million (9.02 per cent) illiterate people. Also relevant to this analysis is that fact that, according to the Instituto Brasileiro de Geografia e Estatística (IBGE) Demographic Census of 2010, updated on 1 January 2019, 30,588,598 million (18.88 per cent) Brazilians have a monthly income of one to two minimum wages, R$510.00 to R$1,020,00, equivalent to US$137.25 to US$274.50 per month (IBGE 2010). Methods of data collecting include an interview with the Health Ministry Communication's social-media coordinator, and a general clinics' area analysis done by the coauthor Marina Caparelli. In a diverse country, with its national territory's continental proportion, [AQ] high illiteracy and low education rate, many citizens' reliance on a public-health system that has been suffering from purposeful scrapping, and many citizens with low income or even no income at all, its population has become vulnerable to the impact of health fake news.

AQ: The meaning of the phrase "its national territory's continental proportion" is not clear.

MIS-DIS-MAL-INFORMATION

Accompanying the social, economic, and political changes of contemporary society, the diversity among the actors assumes different formats, flows, and purposes. Preceding the era of technological and mobile devices, complex communication systems, and the high speed of information flow—made possible by digital network technologies—the traditional ways of transmitting information were oral speech, interpretation of print media, and by the sonority of the radio and television, both analogue. The ease and speed of information transmission have altered the profile of actors who in the past were only consumers of information but nowadays can reinterpret their role by producing and sharing online content in real-time, functions once limited to the great telecommunication conglomerates. Notwithstanding, despite all the technological advent, the current rise of problems calls into question the technical veracity of the news: disinformation. In the academic literature, there are variations in the nomenclature used for deceiving information or fake news. Researchers distinguish the terms *disinformation* and *misinformation* to designate and differentiate misleading information, where some researchers may refer to misinformation as any type of deceptive information but disinformation as only cases where the product is intended to deceive (Karlova and Fisher 2013; Keshavarz 2014). Farkas and Schou (2018, 299) also reflect that many authors may use misinformation and misleading

content unintentionally but to a limited extent, as opposed to intentional disinformation (Fallis 2015; Kumar, West, and Leskovec 2016). Claire Wardle (2017) acknowledges that

> the term 'fake news' is increasingly recognised as problematic, but to find alternatives . . . it is necessary to understand the current information ecosystem and take into consideration three factors: the different types of content; the motivation of the creators; the ways of dissemination.

Starting from the proposal by Wardle and Derakhshan (2018), which considers that the three categories of spectrum information disorder are misinformation, disinformation, and mal-information, in this study it is not possible to find out, through the website of the Ministry of Health, which users shared fake news in different social media platforms. For the authors, mal-information, for instance, consists of 'true like information that violates a person's privacy without public interest justification; goes against the standards and ethics of journalism' (Wardle and Derakhshan 2018, 46). During a search for fake news on the Ministry of Health website, a total of ninety-three fake news items were identified between 8 August 2018 and 10 July 2019. To ascertain how many users circulated the fake news assuming it to be real and how many circulated it knowing that it was fake news, we chose to analyse the phenomenon of the presence of the third element of information disorder spectrum: mal-information.

SOCIAL MEDIA AND VACCINATION-RATE DROP

It is noticeable that fake news is not only focused on the field of politics, as verified in the literature review. Another important phenomenon that has caused serious problems in society is the disinformation, misinformation, and mal-information within the public health area. Spreading deceiving content is one of the factors that has contributed to the reduction in child immunization, increasing the risk of serious diseases reemerging. Another serious problem is the performance of antivaccine groups (Ferrari and Figueredo 2018; Rocha 2019). Although the Brazilian government has expended massive effort through prevention campaigns, research has found that there has been a significant drop in the vaccination rate, which worries public health agents and institutions (Rocha 2019).

In an article published on 21 May 2017, the newspaper *O Estado de S. Paulo* identified five Brazilian antivaccine groups on Facebook with about 13,200 members. Here is a snippet of the article's findings:

> These movements are being singled out as one of the main agents responsible for a recent outbreak of measles in Europe, where more than 7,000 people have already been contaminated. In Brazil, the groups are driven by thematic pages on Facebook disclosing, without a scientific basis, the supposed side effects of the vaccines. (Cambricoli and Palhares 2017)

As seen above, these groups are not only a local happening, as they take place in different parts of the globe.

A FAPESP[4] research study conveyed thorough a journalistic report published on its website in August 2018 also points to the problem. It reports that experts do not dismiss the possibility that the causes for the drop in vaccination rates are due to fake news circulating on social media. The article highlights the incipient action of antiimmunization groups, also known as antivaccination movements. According to the article, in Europe and the United States, the more common reasons for those who do not adhere to vaccination include religious and philosophical matters; but for a segment of this group, the decision not to vaccinate comes from relying on untruthful information.

In an interview with Pesquisa FAPESP on 29 October 2018, Carla Domingues, general coordinator of the National Immunization Program of the Ministry of Health, reported that, in addition to measles, another current concern is the risk of poliomyelitis reemerging. 'The notification of a possible case of paralysis caused by the poliovirus in Venezuela in April was frightening', she declared. Until recently, public health authorities believed they had virtually eliminated measles from Europe. In an article published by BBC Brazil, it was noted that cases of the life-threatening disease are reappearing in Europe as a result of the sharp drop in vaccination rates. According to the article, the number of new cases has tripled in the past year, reaching 82,596. Most of them are concentrated in Ukraine, where 53,218 people contracted the disease. Brazil had zero cases in 2015, 2016, and 2017—and in 2016 even won a certificate from the World Health Organization (WHO) for having eliminated measles (Whewell, Shypko, and Kuryshko 2019). But according to the news report 'The Medical Student Who Died of Measles', 10,262 cases of measles were recorded in Brazil in 2018 (Whewell, Shypko, and Kuryshko 2019). The Ministry of Health, as reported in June 2019, confirmed that there were 123 cases in the country this year—showing a downward trend from the previous year.

THE STUDY'S APPROACH

With an emphasis on the official note of the federal government of Brazil and journalistic articles cited in this chapter about the drop in vaccination rates

die, says Rocha. To restrain fake news' impact on people with cancer, Brazil's Instituto Nacional do Câncer (INCA 2019) produced a booklet on the right diet to be consumed by cancer patients. When asked what measures or initiatives the Health Ministry is taking, Rocha stated that the institution has created the following actions: The Ministry's social media were created exclusively to fight fake news. The entity created a government profile on social media because of the H1N1 flu outbreak. Since 2011, it has been working on social media to transmit news to its followers from reliable sources. According to Rocha, during the study's period, the Ministry had received 10,034 analysis requests, replied to 9,102 of them, and found over ninety instances of fake-news content. To allow the reader an overview of the work done, in eleven months, the channel Health Without Fake News received around forty messages a day. Also, from experience, Rocha could conclude that the main media platforms in which such deceivable information circulates are Twitter and YouTube, especially because of the massive actions done by antivaccine groups. It is interesting to point out that as WhatsApp is an encrypted social media platform, Rocha declares it is easier to monitor Twitter and YouTube news' circulation. Regarding the areas most affected by the phenomenon, Rocha said that the Ministry of Health had given priority to vaccination campaigns fighting fake news, due to the significant drop in vaccination rates. Two programs were created to address this problem: Brazil Vaccine Movement and Health Against Fake News. The Brazil Vaccine Movement has created a yellow seal with a blue heart to encourage vaccination campaigns every year. The campaign is composed of several advertising pieces: printed posters, banners, vehicle stickers, T-shirts, and other items that are used by the population to propagate ideas favourable to vaccination. Since August 2018, the Ministry of Health has made available—in the Health Without Fake News channel[5]—a WhatsApp number to receive messages from the population to report fight fake news about health, adds Rocha. Any citizen can forward messages to this channel, with images or texts that have been received via social media, to verify the information before sharing it.

SELECTION OF FAKE NEWS FOR ANALYSIS ON THE BRAZILIAN MINISTRY OF HEALTH WEBSITE

In the period from 24 August 2018 to 10 July 2019, ninety-three instances of fake news were identified and verified by the Ministry of Health in the campaign Health Without Fake News. Given the technical notice from the Ministry of Health, which highlights the impact of fake news in the area of public health in Brazil, it was decided to focus on all the misleading messages about vaccination during the analysed period. This resulted in focusing on ten fake-news stories as described in table 4.1.

due to fake news, we decided to take ten articles belonging to the category of 'vaccine' for verification. To build the survey sampling, an initial inquiry was conducted using the fake-news section search tool of the campaign Health Without Fake News website, inside the Brazilian Ministry of Health page. A temporal filter was used to gather all fake news identified from 24 August 2018 to 10 July 2019, constituting approximately eleven months. The selection resulted in ten misleading articles about vaccines. The survey resulted in ninety-three articles, of which seventy-six were fake news and seventeen were reliable information. From the list of fake news of this period, we started with the classification of the subjects listed from the following categories: date of publication on the website, fake news verified, Brazil's Ministry of Health replica, dis-mis-mal-information category analysis (Wardle and Derakshan 2017, 20) and scrutiny from the General Clinic's point of view in relation to the potential impact of these materials. We also interviewed the social-media coordinator from the Communication Department Advisory of the Ministry of Health of Brazil. The aim was to understand what measures have been taken by the Brazilian Ministry of Health in combating fake news, as well as dis-mis-mal-information. Also, the ten selected fake-news articles were analysed to verify, from general practice, the potential consequences of such fake news about the health behaviours of the population. To know the scope of the fake-news problem in Brazilian public health, it was necessary to seek official information from the responsible public institution. For this purpose, an interview was conducted with Gabriela Rocha (2019), coordinator of social media, Communication Department advisor of the Brazilian Ministry of Health, to understand what measures have been taken to combat fake news. Within the interview results, it was observed that fake news could directly affect the health of the Brazilian population. According to Rocha, Brazil's low vaccination rate has become a very serious problem in recent years, and there are several factors related to it. Among them, the spread of fake news on social media has contributed to increasing insecurity among the population, and thus the resurgence of diseases eradicated by massive vaccination in the past, such as measles, is now causing people to die. The Ministry of Health has investigated the sources and purpose of the agents or companies disseminating these articles. However, according to the Rocha, these investigations are still ongoing, and it is currently not possible to determine whether it is the trolls, bots, users, or companies that disseminate the most fake news. Rocha reiterates that, context of posts about nutrition, the messages are related to the cure of diseases through miraculous diets. Many people believe in these recipes and end up quitting traditional treatment. She claims, for example, there is a fake-news stream that erroneously states that cancerous tumours are fuelled by sugar, encouraging people to eliminate sugar from their dietary intake. If they do not follow medical guidelines, patients may

Table 4.1. The Ministry of Health in the Campaign Health Without Fake News

Date	*Fake News in Brazilian Health Between: 24/8/2018 to 10/7/2019*	*Replica from the Health Ministry of Brazil*	*Type Information Misuse:*	*Medical Analysis from Dr Marina Caparell Question: What is the potential impact of the misconception present in the fake news below for the health of the Brazilian population?*
24/8/2018	60 American doctors tell the world 'don't take the poison shot of death by yellow fever' Source: Twitter	Yellow fever vaccines are safe and effective when administered in accordance with the National Immunization Program regulations. However, like any immunobiological agent, it has contraindications and precautions. Vaccination is contraindicated for children under 6 months of age and women who are breastfeeding infants under 6 months of age.	Mal-information	1) Yellow fever is a serious mosquito-borne disease present in many states of Brazil. Although the urban yellow fever form is controlled in the country, the Sylvatic (jungle) form is still a great concern. The vaccine is safe, and complications are rarely reported. This kind of fake news could easily lead foreign travellers and many Brazilian communities to avoid the health system campaigns and therefore contribute to an outbreak, highly increasing the chances of introducing the virus to domestic mosquitoes in areas with high human population.
28/8/2018	Brazilian Federal Public Ministry prohibits HPV vaccine Source: Twitter	There is no—and there has never been—a ban on HPV (Human Papillomavirus) vaccine. Adverse events following immunization (AEFI) are mild (application-site pain, edema, and erythema with mild intensity) and may also cause systemic manifestations (fever in 4 per cent to 9 per cent of vaccinated population, headache and gastroenteritis). In Brazil, there are an estimated 16,000 cases of cervical cancer a year and 5,000 deaths of women due to illness.	Mal-information	2) There are pervasive disparities in morbidity and mortality rates of HPV-related cancers for Latin America individuals in comparison to Europe and the United States. The HPV vaccine is an important tool to prevent this morbidity and it represents an effective achievement of the Brazilian government in health policies. Claiming a prohibition would discourage people to look for vaccination programs.

(*continued*)

Table 4.1. ***Continued***

Date	*Fake News in Brazilian Health Between: 24/8/2018 to 10/7/2019*	*Replica from the Health Ministry of Brazil*	*Type Information Misuse:*	*Medical Analysis from Dr Marina Caparell Question: What is the potential impact of the misconception present in the fake news below for the health of the Brazilian population?*
28/8/2018	Japan: Vaccine against HPV under judgement due to horrifying side effects. Source: Twitter	The HPV vaccine is used in over 80 countries, and no evidence in the literature relates the use of HPV vaccine to severe neurological conditions. The World Health Organization (WHO) Vaccine Safety Commission in 2017 released a document reaffirming the security of this vaccine.	Mal-information	3) The human papillomavirus (HPV) vaccination coverage rate had fallen sharply in Japan since 2013 when newspapers began covering negative campaigns against the vaccination. The same could happen in Brazil, where the population is more vulnerable and underinformed. Journalists should strive for impartial coverage so readers can make more informed decisions. Health professionals should be expected to work with journalists to help improve impartiality in news coverage.
30/8/2018	Anticancer vaccine Source: Twitter	There is no anticancer vaccine. The Sírio-Libanês Hospital quoted in this message denied it; this is a piece of false news that has been circulating since 2008. The Health Surveillance Secretariat of the Ministry of Health also testified.	Misinformation	4) There are up to the present time many different therapeutic strategies focused on targeting tumour-associated cells, but there is no 'anticancer' vaccine so far. This information misleads patients to pursue a dramatic quests for a cancer panacea, which creates false hope and disappointment.
3/9/2018	Compulsory vaccines: What's behind it? Are they reliable? Source: Twitter	Many common diseases in Brazil and the world have ceased to be a public-health problem since massive population vaccination. With technical support from specialized teams, the Ministry of Health ensures that vaccination is safe.	Disinformation	5) Vaccination programs are a worldwide recognized health-protection strategy and a mandatory measure of health protection. Immunization-focused educational messages may be enough to improve vaccination coverage, and, to a small degree, enhance knowledge, particularly where awareness is identified as a barrier to vaccination, such as occurs in Brazil.

24/9/2018	The new vaccines still cause autism and governments know it. Source: Twitter	A study presented in 1998, which raised concerns about a possible link between the measles, mumps, and rubella (MMR) vaccine and autism, was later considered seriously flawed and the article was withdrawn by the journal that published it. Unfortunately, its publication triggered a panic that led to a drop in vaccination coverage and subsequent outbreaks of these diseases.	Mal-information	6) The original research conducted in 1998 claiming that MMR vaccine increased the incidence of autism was proven not only wrong but also a serious ambush with many illegal issues. Vaccines are not associated with autism. There are several well-conducted studies up to the present date showing no link between the MMR vaccine or any other vaccine with this cognitive impairment. This kind of statement is very dangerous, especially for developing countries such as Brazil, where most people still don't have access to scientific information. This misleads parents into avoiding vaccinating their children, and therefore allows infectious diseases to increase.
30/1/2019	Vaccines are harmful. Source: Twitter	Only the vaccines with safety and efficacy guaranteed enter into the national calendar. In addition to the safety control for vaccine registration, sample lots of these immunobiological tests undergo quality testing at the National Institute of Health Quality Control (INCQS).	Mal-information	7) Early childhood vaccination is an essential global public-health practice that saves two to three million lives each year, but many children do not receive all the recommended vaccines, mainly because vaccination programs rely on the public's acceptance of vaccines. This is especially true in Brazil where most people don't have access to a scientific information database and an argument such as this one could easily lead to misunderstanding the importance of vaccination.
31/1/2019	New fatal flu and fennel tea. Source: Twitter	The Hospital das Clínicas de São Paulo explained that it did not warn the population about a supposed influenza virus spread. It is false information that fennel tea has the same substance as the drug Tamiflu.	Disinformation	8) Brazilian people often consider flu and cold as the same entities, which they are obviously not. Alkaline diets or drinks might have antiviral properties as in vitro studies demonstrated inactivation effect of alkaline medium on the respiratory virus; however, it is a mistake to say that herbal teas have the same compounds of antiviral drugs. Statements such as this one underrate the importance of drugs in the treatment of a serious disease such as influenza.

(*continued*)

Table 4.1. ***Continued***

Date	*Fake News in Brazilian Health Between: 24/8/2018 to 10/7/2019*	*Replica from the Health Ministry of Brazil*	*Type Information Misuse:*	*Medical Analysis from Dr Marina Caparell Question: What is the potential impact of the misconception present in the fake news below for the health of the Brazilian population?*
2/8/2019	Measles vaccination D-Day on 16 February. Source: Twitter	The Ministry of Health warns that the image circulating on social media announcing a National D-Day vaccination on 16 February is false! The poster refers to vaccination taking place only in the municipality of Timóteo, Minas Gerais, Brazil. Measles (triple viral and tetra viral) vaccines are offered throughout the year at the immunization routine for age groups indicated on the National Vaccination Calendar.	Disinformation	9) The 2019 flu vaccination campaign had its 'D-Day' on Saturday, 4 May. On that date, health centres opened exceptionally throughout Brazil to immunize all priority groups. The national immunization program in Brazil is available in all the states during weekdays, all you must do is go to the basic health unity with the vaccination chart in your neighbourhood. A nurse will attend and explain which vaccines are recommended for your child.
2/5/2019	10 reasons why you should not vaccinate your child. Source: Twitter	The Ministry of Health clarifies that all vaccines offered by the National Immunization Program are safe and have the appropriate records in the National Agency of Sanitary Surveillance (in Portuguese, Agência Nacional de Vigilância Sanitária, Anvisa). The vaccines are submitted to a rigorous process of quality evaluation by the INCQS of the Oswaldo Cruz Foundation.	Mal-information	10) The Brazilian immunization program was created by pediatricians and immunologists among many other specialists considering safety and efficacy. There is no reason to consider the vaccines a threat to a child; however, the diseases that the vaccines cover are life-threatening for children and adults. There should be a systematic effort to combat this kind of fake news and to create a long basis informative and educational program for Brazilian communities relating to vaccines.

From the data of table 4.1, it was verified that in the spectrum of the information disorder, 60 per cent mal-information, 30 per cent misinformation, and 10 per cent of disinformation occurred. If we establish a link between the percentage of fake-news results exposed, we can conclude that 90 per cent of false information is concentrated in the area of potential harm to the health of the Brazilian population. Another relevant fact, which corroborates Rocha's comments, is that 100 per cent of fake news in the area of vaccination was propagated via Twitter (Rocha 2019), despite the Ministry of Health's creation of a channel of communication with society to resolve misleading information. The radius and action of each of them are immeasurable, causing harmful and irreversible effects to the health of the Brazilian population, such as the return of measles and yellow fever outbreaks in Brazil in previous years.

RESULTS AND DISCUSSION

The analysis of the fake news about selected vaccines on the website of the Brazilian Ministry of Health, and categorized from the perspective of the conceptual framework of information disorder (Wardle and Derakshan 2017), drew a worrisome scenario of dissemination that had previously been eradicated and the great potential risk of death. The propagation and sharing of this type of information by social-media mechanisms have a negative impact on underdeveloped populations where there is a lack of access to scientific information on how to properly treat certain diseases. However, such fake news does not only plague adults or young people of school age. It also affects defenceless children whose parents and guardians are well-educated people yet refuse to vaccinate their children. This is the case of antivaccine groups, which influence many people in Brazil. Even though we only identified ten fakes, the damage to public health is enormous, considering the multiple, instantaneous instances of sharing promoted by the users and the possibility of publishing the same material in different networks. According to Dizikes (2018), people are primarily responsible for spreading fake news. The interview with Rocha revealed that the government has developed various technological and media resources to combat misleading information that causes damage to public health. The Brazil ~~Without~~ Vaccine Movement [AQ] and Health ~~Without~~ Fake News Campaign are actions developed to prevent the drop in vaccination rates recorded in recent years. However, according to the Ministry of Health data, from August 2018 to July 2019, the institution received 10,034 instances of fake news in all areas of health. That is an average of just over twenty-seven pieces of fake news a day. That is alarming data,

AQ: Earlier in this chapter, this was called the Brazil Vaccine Movement; please clarify.

even considering the 9,102 responses sent to the population within just eleven months of the Health Channel Without Fake News. The Ministry of Health's approach in using social media like Twitter and YouTube failed to consider that closed social media groups, such as WhatsApp's encrypted network, are also where the wide circulation of fake news occurs. These initiatives are the counterpart to minimizing the potential harm to public health categorized in the spectrum of information disorder—disinformation and mal-information. Furthermore, the analysis of coauthor Marina Caparelli, from the General Clinic, brought clarification from the medical area about the contagion of diseases, the risks of nonimmunization of children, the severity of every disease that affects people of greater educational and economic vulnerability. The population belonging to this social sphere does not have access to Brazil's private and high-quality health system. These people are assisted by the Public Health System (SUS),[6] which in recent years has been subjected to funding cuts, and that for several reasons cannot meet the demand for the diseases of this social group. This problem is serious because medical information is not properly transmitted to members of the needy public who are bombarded by fake news.

CONCLUSION AND FUTURE WORK

This study conducted an analysis of fake news in the subject area of health in Brazil, observed from the information disorder scenario (Wardle and Derakshan 2017), through a sampling of fake news selected from the website of the Brazilian Public Ministry, between 24 August 2018 and 10 July 2019. We conclude that the importance of the role of government agencies responsible for Brazilian public health through media mechanisms and the monitoring of open social media is a first step in addressing the reality of disinformation and mal-information which is harmful to the health of the population. From the perspective of the public services rendered, these are pertinent, necessary, and urgent actions. Nevertheless, it is necessary to reach populations on the margins of large urban centres, but with access to the Internet and social media. In the context of information disorder, media literacy (Mallon 2018; Leetaru 2019) is one of the valid tools for users to recognize and remain critical of the flow of fake news. According to Mallon (2018, 329), analysing news sources in their varied forms is more important and challenging than ever; from 'misinformation' to misleading headlines, researchers are bombarded with such a high amount of content that it can be a struggle to know what is factual and what is not. Critical thinking and close analysis of news and multimedia sources is a skill that is required to be an informed learner

and citizen at any age—making it extremely relevant for not only students, but also librarians, faculty, staff, and virtually all members of a university community and beyond (Mallon 2018, 329). The solution to the problem of fake news is in media literacy and not in technology, according to Leetaru (2019). Media literacy, or information literacy, is one of the elements that can be included in school and play activities for children and young people, and in curricular disciplines that are not in the area of communication. Creating free software or applications to track fake news, in addition to media resources, and extending media literacy to public and private schools in Brazil, are concerns for research.

NOTES

1. Department of Computer Science, Brunel University London, Uxbridge, UB8 3PH, United Kingdom.
2. Department of Journalism, Brasília University, Brazil. Campus Darcy Ribeiro, Brasília, Distrito Federal, Brazil.
3. Brazilian Public Health System at a Psychosocial Center, Brasília, Distrito Federal, Brazil.
4. FAPESP–Fundação de Amparo à Pesquisa do Estado de São Paulo is a local research support public foundation in São Paulo state.
5. Health Campaign Without Fake News. Ministry of Health. http://www.saude.gov.br/fakenews.
6. Sistema Único de Saúde (SUS) is one of the largest and most complex public-health systems in the world, providing simple blood-pressure assessment through Basic Care to organ transplantation, guaranteeing universal and free access to the entire population of the country. http://www.saude.gov.br/sistema-unico-de-saude.

REFERENCES

Brasil. 2019. *Boatos e notícias falsas prejudicam campanhas de vacinação*. http://www.brasil.gov.br/noticias/saude/2018/09/boatos-e-noticias-falsas-prejudicam-campanhas-de-vacinacao.

Cambricoli, F., and Palhares, I. 2017. Grupos contrários à vacinação avançam no País e preocupam Ministério da Saúde. *O Estado de São Paulo. São Paulo*, 21.

Dizikes, P. 2018. 'Study: On Twitter, False News Travels Faster Than True Stories'. *MIT News*, 8 March. http://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308.

Fallis, D. 2015. 'What Is Disinformation?' *Library Trends* 63 (3): 40–26.

Farkas, J., and Schou, J. 2018. 'Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood'. *Javnost-The Public* 25 (3): 298–314.

Ferrari, C., and Figueiredo, P. 2018. 'Correntes de WhatsApp espalham informações falsas sobre febre amarela'. *Mensagens trazem dados errados ou exageros sobre os efeitos da vacina e a incidência da forma grave da doença.* https://apublica.org/2018/02/truco-correntes-de-whatsapp-espalham-informacoes-falsas-sobre-febre-amarela/.

IBGE–Instituto Brasileiro de Geografia e Estatística. 2010. 'Alfabetização e classes de rendimento'. *Censo demográfico do IBGE, 2010.* https://sidra.ibge.gov.br/tabela/1390.

INCA. n.d. *Instituto Nacional do Câncer.* https://www.inca.gov.br/.

Karlova, N. A., and Fisher, K. E. 2013. 'A Social Diffusion Model of Misinformation and Disinformation for Understanding Human Information Behaviour'. *Information Research* 18 (1) paper 573.

Keshavarz, H. 2014. 'How Credible Is Information on the Web: Reflections on Misinformation and Disinformation'. *Infopreneurship Journal* 1 (2): 1–17.

Kumar, S., West, R., and Leskovec, J. 2016. 'Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes'. In *Proceedings of the 25th International Conference on World Wide Web*, 591–602.

Leetaru, K. 2019. 'A Reminder That "Fake News" Is an Information Literacy Problem—Not a Technology Problem'. Forbes.com. https://www.forbes.com/sites/kalevleetaru/2019/07/07/a-reminder-that-fake-news-is-an-information-literacy-problem-not-a-technology-problem/#64ac93c66a6f.

Mallon, M. 2018. 'Media Literacy'. *Public Services Quarterly* 14 (4): 329–41.

Revista, V. 2018. 'Falha em vacinação ameaça erradicação de doença no Brasil'. https://veja.abril.com.br/saude/falha-em-vacinacao-ameaca-erradicacao-dedoencas-no-brasil/.

Rocha, G. 2019. Personal interview, 8 July. Brasília, Ministério da Saúde.

Wardle, C. 2019. '7 Types of Mis- and Dis-information'. First Draft. https://firstdraftnews.org/fake-news-complicated/.

Wardle, C., and Derakhshan, H. 2017. 'Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making'. *Council of Europe Report*, 27.

———. 2018. 'Thinking about 'Information Disorder': Formats of Misinformation, Disinformation, and Mal-Information'. In *Journalism, Fake News and Disinformation.* https://en.unesco.org/fightfakenews.

Whewell, T., Shypko, K., and Kuryshko, D. 2019. 'The Medical Student Who Died of Measles'. *BBC World Service.* https://www.bbc.com/ portuguese/geral-48880367.

5

The Psychology of Fake News

Dr. Seth M. Porter

In chapter 1, we established that the term *fake news* 'should be reserved for cases of deliberate presentation of false or misleading claims as news, where these are misleading by design' (Gelfert 2018, 1) with an emphasis on the term *by design*. The argument is that this form of news is linked to channels intent on the manipulation of their audience's cognitive processes. This is misinformation designed to take advantage of psychological errors in human processing. This influences individual and collective cognitive processing on many different levels. For example, fake news and misinformation affects memory recall, decision making, belief in misinformation, and even affects collective memory and cultural norms. The literature in psychology and neuroscience indicate that the cause is fundamental human processing errors. Throughout this chapter, the principal errors that affect the belief in and sharing of fake news are illustrated in the following order. Section 1 covers memory and how fake news impact human memory. This section includes the 'illusory truth affect', 'collective memory' and 'cross-cuing'. Section 2 covers the principle of cognitive biases that affect the belief in and sharing of fake news. This section includes 'What You See Is All There Is' (WYSIATI) and cognitive dissonance, confirmation bias, in-group bias, and choice-supportive bias. In addition, this chapter explores how this all cascades into motivated reasoning. Section 2 also investigates how these affects are more extreme in delusion-prone individuals and religious fundamentalists. Additionally, it explores the literature on how simple laziness affects human processing and fake news. While there is still much research to be done on the human psyche, this chapter illustrates the key inferences on fake news and our propensity to believe it.

MEMORY AND FAKE NEWS

Memory plays an essential part in assessing the validity of information, including the belief in and sharing of fake news. The cognitive psychology behind this memory processing is the 'illusory truth effect'. The illusory truth effect is the propensity to believe the validity and accuracy of information based on the repetition and continual exposure to information (Hasher, Goldstein, and Toppino 1977). As individuals listen to previously encoded information, this consistently reinforces their recall of this encoded information, and it becomes frictionless and their trust in the information increases (Bronstein et al. 2018).

Essentially, the more misinformation is repeated, the more likely individuals will trust in the validity and accuracy of the information. As individuals critically asses information, they cognitively relay how familiar it feels, and this can greatly affect the spread of and belief in fake news. Research supports this assertion, and political candidates and adherents of misinformation understand this memory-processing error. Polage (2012) illustrates this effect on fake news through a cleverly designed experiment. This experiment exposed the treatment group to false news stories and misinformation that was portrayed as true. The control group was not exposed to these false news stories and misinformation. Both groups were reexamined after a five-week delay, and the treatment group rated the false news stories and misinformation as valid and more accurate than the control group.

Furthermore, the treatment group members were much more likely to believe they had encountered the fake news and misinformation outside of the boundaries of the experiment, which indicates the creation of false memories. Moreover, there is further research that supports this assertion and even indicates that a single exposure to fake news can affect the formation of false memories and the belief in misinformation (Polage 2012; Dreyfuss 2017; Bronstein et al. 2018). Furthermore, misinformation and fake news influence belief in and sharing of fake news, and even after it has been proven untrue memories continue to influence this belief (Singh et al. 2018).

These alarming findings reveal a memory-processing error on the individual level that contributes to the propagation of fake news and leaves an opening for unethical providers of misinformation. Furthermore, there is a growing body of literature that suggests there is a macro effect of fake news and memory, through a similar processing error—collective memory.

COLLECTIVE MEMORY

Collective memory denotes how groups of individuals remember the past (McDougal 1920; Halbwachs 1980; Roediger and Desoto 2016). This process

a shared Arab–Palestinian identity. Similarly, divergent truths emerged after the erection of the Berlin Wall.

This is just one example of the real-world effect of intergenerational-shared memory and how cross-cueing can adversely affect group relations. Regardless of the accuracy of the information that is cued from one individual to another and within a group, human memory and cognition are predisposed to have memories evolve and change based on this memory input (Spinney 2017). Additionally, as discussed previously in regard to the illusory truth effect, if these two issues compound, collective memory is used to propagate the belief in and sharing of misinformation and fake news (Spinney 2017). And unfortunately, these are far from the only cognitive-processing errors that affect the propensity to do so.

COGNITIVE BIASES

Memory is a fundamental concept to understanding the foundational concepts within psychology and fake news. However, it is far from the only one. Cascading downstream from memory and collective memory is the principle issue in cognitive psychology and fake news. This is the simple fact that as human beings, we are fundamental, error-prone, and predisposed to cognitive biases (Kahneman and Tversky 1979). These biases come in many forms and guide much of our decision-making ability subconsciously. This can lead to many undesirable outcomes, including the current pestilence of fake news and the distinct effect it has had on individuals and the greater macro-civil society.

This section illustrates the core concepts of cognitive biases and how it affects fake news. While this will be far from complete in cataloguing the full-range of cognitive biases, it describes the essential cognitive functions that are of significant concern. The principle biases that will be covered are What You See Is All There Is (WYSIATI) and cognitive dissonance, confirmation bias, in-group bias, and choice-supportive bias. Furthermore, this section, in conjunction with the previous section on memory, is the basis of the rest of the cognitive psychology on fake news and the corresponding literature.

WYSIATI is a heuristic for the human tendency of pattern recognition. To fundamentally understand this, one must look to cognitive psychology and the effects of fake news. It is essential to understand System 1 and System 2 processing. System 1 is the fast, automatic, unconscious thinking that we use every day—however, it is error-prone. System 2 is the slow, conscious, effortful, and reliable type needed for complex decision making (Kahneman 2012). Both of these processes work together, and we would not be able to function without this parallel processing. To illustrate this, think of driving a

can occur at the national, subnational, and even as local as tribal or family unit. Collective memory is an intergeneration process on how information and beliefs are shared and constructed, to assist in the contribution to a shared belief of the past and how specific events unfolded (Svob et al. 2016).

Collective memory can be shared memories about historical events, or something as simple as a family vacation. Much of the historical research focuses on how cultures and the memories of countries are affected by the transmission of memories from the previous generation (Svob et al. 2016). This literature indicates that the second generation often collectively remembers an event or a set of events similarly based on information they were given, regardless of the accuracy of the information. This can affect entire countries' understanding of events and how they will interact with future related information—and even fake news.

While collective memory has been a field of historians and humanists for most of the twentieth century, it has been recently an emerging field in experimental cognitive psychology. Coman and Hirst (2012) illustrate this in a research they conducted on pairs of people discussing past shows where a speaker is able to reinforce aspects of an event by selectively reiterating them. Their research indicates that collective memory is a multifaceted individual and societal phenomena which includes the illusory truth effect at a macro scale (Rubin 1996; Wang 2008; Spinney 2017; Hirst, Yamshiro, and Coman 2018). This literature shows that while memories are formed individually, there is also a large effect on the formation of the memories by shared experience and the selective information inputs of an individual's environment. Moreover, cross-cueing, a central processing function in collective memory, greatly affects the potential for distribution of misinformation from individuals to the collective. Cross-cueing occurs when information is exchanged among group members that progresses to become a collective understanding of the topic of the information that an individual would have never understood individually. Basically, if an individual shares information with another individual, this can cue memories, and even form false memories, in a second individual (Rajaram and Pereira-Pasarin 2010). This can even lead to forgetting and an increase in memory errors, which can cue a similar forgetting and memory error in other individuals (Rajaram and Pereira-Pasarin 2010; Spinney 2017). Because of this error in our processing, if an individual is caught in a reinforcing memory bubble, his memories will actually change, and this will propagate to the greater group, and depending on an individual's information inputs, differing truths for national, subnational, and even as local as tribal or family units can emerge. Spinney (2017) illustrates the real-world impact of this:

> This effect is evident in real-world situations. Palestinians living in Israel and those in the West Bank, who were separated by force during the Arab–Israeli wars of 1948 and 1967, have gravitated to different versions of their past, despite

car. When you first learn to drive, you are using System 2. You are thoughtful, conscious, and slow. It is exhausting and stressful. As you become an experienced driver, this task becomes the realm of System 1 thinking, or automaticity. System 1 is a master of pattern recognition and unconscious decision making (Kahneman 2012).

The duel processing is vital to decision making; however, human beings are inclined to overuse System 1 processing and decision making where they are not domain experts. This inclination to rely on System 1 and the innate pattern recognition of this processing can lead to the assumption that all the information needed to make a sound, analytical decision is present—or WYSIATI (Kahneman 2012).

Individuals make decisions and create narratives that do not reflect the empirical truth but do reflect the individual's truth, based on their assumption that the patterns they recognize as the truth. This theory states that a mind will see a story or a set of facts and the mind will convince itself that it is true. Essentially, the mind will create a narrative of the knowns and ignore unknowns that do not fit the pattern or narrative that it desires (Kahneman 2012; Givens et al. 2018). This idea incorporates the fundamental human processing error of *cognitive dissonance*. (Festinger 1962/1975). Cognitive dissonance is a cognitive bias that states that if a set of ideas is not consistent with another, human processing will automatically construct a narrative that will make the ideas more consistent (Festinger 1962/1975; Tsipursky, Votta, and Roose 2018). This is also consistent with the idea of WYSIATI and the central human processing errors that lead individuals to create coherent patterns out of isolated bits of information. This cognitive dissonance motivates individuals to adjust their behaviour to avoid distress (Festinger 1962). The idea of WYSIATI and its association with cognitive dissonance is an unpleasant cognitive error that directly affects the sharing of and belief in fake news. If individuals are naturally wired to create narratives, and even have an incentive to minimize the cognitive discomfort by changing their behaviour and beliefs when confronted with a set of ideas that is individually inconsistent, this will inherently lead them to create a narrative that is cognitively agreeable (Festinger 1962/1975; Kahneman 2012; Tsipursky, Votta, and Roose 2018). WYSIATI is a central cognitive bias that effects the psychology of fake news. Although confirmation bias, an associated error, is of similar significance.

CONFIRMATION BIAS

Confirmation bias is the most recognized of cognitive errors. This common cognitive error is the propensity for human beings to process information through interpretation and recall that is reaffirming and consistent with existing beliefs and opinions (Kahneman and Tversky 1979; Plous 1993;

Nickerson 1998; Kahneman 2012). This tendency extends to information-seeking behavior. Noticeably, for this discussion, confirmation bias is closely related to the belief in and sharing of fake news.

Moreover, this error of inductive reasoning is even stronger when associated with emotionally charged beliefs such as religion or partisan fundamentalism (Pennycook et al. 2018). This research demonstrates that individuals consciously and unconsciously prefer, and even seek, information that is desirable and confirms to their preexisting beliefs. Information that confirms to preexisting and emotionally charged beliefs is perceived as more accurate, regardless of the accuracy of the information (Lazer et al. 2018; Pennycook et al. 2014). Clearly, confirmation bias is a systematic error of human reasoning and one of the causal motives for the belief in and sharing of fake news. Specifically, individuals seek out information that supports their preexisting beliefs, and this reinforces their prior convictions (Nickerson 1998; Tsipursky, Votta, and Roose 2018). They are also more likely to spend less cognitive energy critically thinking about the factual accuracy of the information once found and will ignore data that is outside of their preconceived scope of factual accuracy. Moreover, as the research has shown, when the information is of an emotionally charged nature, this systematic error is deepened. Fake news is designed to take advantage of this cognitive error and is common concerning political, religious, cultural, and tribal topics. Because of this, the impact of confirmation bias compounds and individuals are exponential, more likely to believe, share, and support information regardless of the accuracy or clear evidence, if it supports their initial belief (Hirt and Markman 1995; Nickerson 1998; Kray and Galinsky 2003; Lilienfeld, Ammirati, and Landfield 2009; Givens et al. 2018; Tsipursky, Votta, and Roose 2018).

While the cognitive psychology literature clearly supports this causal association between confirmation bias and information seeking, and the lack of analytical rigour in individuals' cognitive assessment, this error in reasoning can be clearly illustrated by much early psychology and philosophy, in the following quote, 'When men wish to construct or support a theory, how they torture facts into their service!' (Mackay 2012, 552). Much of confirmation bias is unconscious information seeking and processing. However, this cognitive error is also closely associated with motivated reasoning—a conscious cognitive error that is directly related to the belief and support of fake news, which will be covered at length subsequently.

IN-GROUP BIAS

In-group bias, similar to collective memory, is concurrently an individual as well as a societal phenomenon. In-group bias is the propensity of individual

behavioural ranges to affiliate with a similar range of other individuals that are observed or understood to be in their own in group or tribe (Mullen, Brown, and Smith 1992; Verkuyten and Nekuee 1999; Tsipursky, Votta, and Roose 2018). Additionally, when behaviour or thought is perceived to be consistent with the behaviour or thought of their out-group, individuals will adjust their behaviour and thought accordingly. Moreover, individuals are more likely to lie if it supports their in-group (Kahneman and Tversky 1979; Mazar, Amir, and Ariely 2006). All of this behaviour is consistent with online behaviour, and this causes an outsized influence on belief and the sharing of fake news. Individually, this propensity to believe, and share information and news without skeptical thought or analysis, is problematic. However, when this content and belief progresses it on a wide-range scale, as promotion of questionable content, favorable to one's in-group, indicates shared values, is likely to result in positive consequences for behavior that supports the mutual cause, and increases group social reinforcement.' (p. 53). **[AQ]**Fake news is a symptom of this cognitive error, regardless of the social impact.

AQ: Please indicate with an opening quotation mark where this quote begins.

Done

CHOICE-SUPPORTIVE BIAS

While these previous cognitive biases and human processing errors are the central cognitive biases that affect the belief in and sharing of fake news, they are not the only factors. Another singular influence is *choice-supportive bias*. Choice-supportive bias is an interpretation of our choices to favour the choice previously made (Correia 2014; Tsipursky, Votta, and Roose 2018). Human cognitive processing attempts to minimize adverse emotional reactions and conditions by focusing on aspects of the choice that favour the decision made (Correia 2014; Tsipursky, Votta, and Roose 2018). This is a fundamental processing error in human cognition that preexisting beliefs and decisions made. An example of this is illustrated by Tsipursky, Votta, and Roose (2018, 59):

> If one has a choice between buying a Ford or a Honda and buys the Ford, attention to stimuli that support the value of the Ford will be enhanced to reinforce the decision, while attention to stimuli that support the value of the Honda will be suppressed to avoid the aversive condition of making the wrong decision. This relates to fake news as people may be more likely to attend to and share news that supports their previous choices (e.g., the selection of a political party or candidate).

These cognitive biases, with the effects of the previous cognitive errors, has a clear effect on the belief in and sharing of fake news. As an individual makes an emotionally charged decision, based on preexisting beliefs that lead them to be more susceptible to believing in and sharing fake news

based on confirmation bias, their cognitive errors are compound with choice-supportive bias, which leads individuals to retroactively support a potentially incorrect decision (Correia 2014; Tsipursky, Votta, and Roose 2018; Lazer et. al 2018). As stated above, this is very common in politics where individuals make a decision and share beliefs, information, and fake news that support their prior choice regardless of accuracy and truthfulness. This becomes potentially more disturbing with motivated reasoning.

ANALYSIS AND REFLECTION

Much like WYSIATI, motivated reasoning is almost a compilation of the human processing errors previously discussed that leads to a tendency to a belief in and sharing of fake news. It is best to understand motivated reasoning not as a singular cause of the spread of fake news, but a compilation of process that results in the belief and sharing of fake news (Kahan 2013). This is a heuristic for understanding the complex interplay of multiple cognitive processes that lead to individuals making processing errors regarding fake news. Motivated reasoning is an unconscious and, sometimes, almost a conscious cognitive process that motivates individuals to process information that fits into a preexisting belief (Kunda 1990). The evidence indicates that through a processing strategy, much like confirmation bias, individuals are prone to believe the information they want to believe. This evidence also indicates that individuals are constrained by their mental capacity, critical-thinking skills, and overall cognitive ability to construct rational justifications for these conclusions (Kunda 1990; Kahan 2013; Bronstein et al. 2018). Kahan (2013) illustrates this through a memorable example:

> In the 1950s, psychologists asked experimental subjects, students from two Ivy League colleges, to watch a film that featured a set of controversial officiating calls made during a football game between teams from their respective schools. The students from each school were more likely to see the referees' calls as correct when it favoured their school than when it favoured their rival. The researchers concluded that the emotional stake the students had in affirming their loyalty to their respective institutions shaped what they saw on the tape.

The students truly believed that the referees were correct when favourable and incorrect when unfavourable. They were motivated to believe this way based on their preexisting beliefs and tribal identify; however, they were not consciously aware that they believed what they believed because of their identity as a member of the group (Kahan 2013; Kahan et al. 2017).

As a proud Philadelphian, I have seen this from a personal vantage point. Even though I often know that empirically, the referees' decisions are not consistently screwing the Philadelphia Eagles, I and the rest of the city truly think this when the Eagles lose. Unfortunately, motivated reasoning and its underlying cognitive processes do not confine themselves to athletic competitions. This explanation of the proliferation of fake news is known as Motivated System 2 Reasoning (MS2R) (Kahan 2013). MS2R contends that individuals, when confronted with information, will give preferential belief to the information that aligns with their political and personal identity (Kahan 2013; Pennycook and Rand 2017). Motivated reasoning implies that belief in fake news is primarily tribal and partisan (Mercier and Sperber 2011; Haidt 2012). And, because of the innate political nature of fake news, motivated reasoning leads individuals to make fundamental processing errors that lead to an inclination in believing and sharing fake news that support their preferred policy, or political party, or candidate, even when presented with factually grounded negative information (Redlawsk, Civettini, and Emmerson 2010). This problem, combined with other cognitive errors, implies that individuals are much more susceptible when the content of fake news is aligned with political or cultural ideology (Kahan et al. 2017; Van Bavel and Pereira 2018). In addition, much like the previous cognitive errors discussed, this is not solely an individual phenomena (Kahan 2013). People's motivated reasoning, information seeking, and the propensity to believe in and share fake news is centred on the conscious and unconscious motivation in maintaining a valued identity in their tribal groupings. Common sense would dictate that motivated reasoning, confirmation bias, and related cognitive errors would employ these mechanisms consistently and to their economic benefit, when possible—this is not the case (Khaneman 2012). Kahan (2013) illustrates that the empirical evidence indicates that individuals[AQ] use these mechanisms to lobby politically for their economic self-interest. Often, individuals lobby against potentially self-serving policies like social welfare spending and tax policy. While this behaviour is inconsistent with rational choice theory, it is consistent with cognitive psychology and information-seeking behaviour (Kahan 2013). The individuals are using information, and often inaccurate and fake information, to protect their identity and perceived status within a group.

AQ: Should the words "do not" be inserted here?

Furthermore, the evidence seems to indicate that the propensity of this mechanism is more robust with delusion-prone individuals, dogmatism, and religious fundamentalism (Bronstein et al. 2018). The stated rationale for this mechanism is a deficiency of analytical reasoning and open-minded reasoning (Evans and Stanovich 2013). Additionally, this relates to the propensity of individuals with less analytical reasoning skills to also have lower working

memories, which makes an impact on continuing to believe fake news after it has been corrected or proven inaccurate (Singh et al. 2018). The evidence appears to indicate that dogmatic individuals and religious fundamentalist engage in less analytical thinking and open-minded analysis. Bronstein et al. (2018) show a strong correlation between dogmatism, religious fundamentalism, and the propensity to believe in and share fake news because of the reduction of analytical thinking and open-minded analysis (Martin 2008; Pennycook et al. 2014). This evidence is aligned with the previous literature and illustrates the analytical, cognitive, and memory-related functions that affect the belief in and sharing of fake news.

In spite of the almost overwhelming evidence, there is a developing field of research that argues these cognitive processes are less to blame for believing and sharing misinformation than just pure analytical laziness. Pennycook and Rand's (2017) findings claim that by using the cognitive reflection test (CRT) as a measure to analyse cognitive and analytical reasoning, it may illustrate laziness. Their findings were negatively correlated with motivated reasoning and even for information that aligned with the individual's partisan beliefs. Essentially, their findings indicate that analytical thinking, based off the CRT assessment, shows that the belief in fake news and misinformation is driven by analytical laziness instead of memory, cognitive biases, and motivated reasoning. If true, this could potentially be a positive aspect in assessing and combating fake news and misinformation (Pennycook and Rand 2017). Principally, individuals would need just to slow down and think critically.

Pennycook and Rand (2017, 6) ask for caution and contend that 'another possibility is that explicitly asking participants to assess fake and real news led them to be more analytic and objective than they otherwise might be.' Generalizing the findings of one counterintuitive study to minimize the effect of the entire body of the empirical literature is unwise; therefore, these findings should be approached with restraint and studied further. As of now, the findings are more likely to be explained by motivated reasoning and the participants in the study being primed for analysis. Individuals and groups outside of the lab are more likely to interact with information with System 1 processes.

CONCLUSION

Throughout this review of the literature, the psychology behind fake news has been explored at length. While this has been an abridged review of the literature, it has illustrated the key principles behind the human susceptibility to believe in fake news and misinformation. This includes weaknesses in the human memory. As a reminder, these are the illusory truth affect, collective memory, and cross-cuing. These fundamental errors in human memory are a

key component of the human psyche and add to the liability in believing in fake news and misinformation. Additionally, the human psyche is error prone to cognitive biases. As a reminder, these biases were WYSIATI and cognitive dissonance, confirmation bias, in-group bias, and choice-supportive bias, and all of these compound into motivated reasoning. This is even more acute in delusion-prone individuals and religious fundamentalists. And of course, laziness plays its part. The human psyche is almost perfectly designed to fall for fake news and misinformation. Knowing what the processing errors are will assist in fighting fake news. In addition, understanding our nature through empirical research is the first step, but far from the last.

REFERENCES

Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., and Cannon, T. D. 2018. 'Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic Thinking'. *Journal of Applied Research in Memory and Cognition* 8 (1): 108–17.

Coman, A., and Hirst, W. 2012. Cognition through a Social Network: The Propagation of Induced Forgetting and Practice Effects. *Journal of Experimental Psychology: General* 141 (2): 321–36.

Correia, V. 2014. Biased Argumentation and Critical Thinking. In *Rhétorique et cognition Rhetoric and Cognition*. Bern, Switzerland. https://www.peterlang.com/view/9783035195415/09_ch03.html.

Dreyfuss, E. 2017. 'Want to Make a Lie Seem True? Say It Again. And Again. And Again'. *Wired*, 11 February. https://www.wired.com/2017/02/dont-believe-lies-just-people-repeat/.

Festinger, L. 1962. 'Cognitive Dissonance'. *Scientific American* 207 (4): 93–106.

———. 1975. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Givens, M., Porter, S., Viars, K., and Holdsworth, L. 2018. *Fighting Fake News and Biases with Cognitive Psychology*. Georgia Southern University. https://digitalcommons.georgiasouthern.edu/cgi/viewcontent.cgi?article=2084andcontext=gaintlit.

Haidt, J. 2012. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Paragon.

Halbwachs, M. 1980. *The Collective Memory*. Translated by F. J. Ditter, Jr. and V. Y. Ditter. New York: Harper Row. (Original work published 1950).

Hasher, L., Goldstein, D., and Toppino, T. 1977. 'Frequency and the Conference of Referential Validity'. *Journal of Verbal Learning and Verbal Behavior* 16 (1): 107–12.

Hirst, W., Yamashiro, J. K., and Coman, A. 2018. 'Collective Memory from a Psychological Perspective'. *Trends in Cognitive Sciences* 22 (5): 438–51.

Hirt, E. R., and Markman, K. D. 1995. 'Multiple Explanation: A Consider-an-Alternative Strategy for Debiasing Judgments'. *Journal of Personality and Social Psychology* 69 (6): 1069–86.

Kahan, D. M. 2013. 'Ideology, Motivated Reasoning, and Cognitive Reflection'. *Judgment and Decision* Making 8 (4): 407–24.

Kahan, D. M., Peters, E., Dawson, E. C., and Slovic, P. 2017. Motivated Numeracy and Enlightened Self-Government. *Behavioural Public Policy* 1 (1): 54–86.

Kahneman, D. 2012. *Thinking, Fast and Slow*. New York: Macmillan. https://us.macmillan.com/thinkingfastandslow/danielkahneman/9780374533557.

Kahneman, D., and Tversky, A. 1979. 'Prospect Theory: An Analysis of Decision Under Risk'. *Econometrica* 47 (2): 263–91.

Kahneman, D., Slovic, S. P., Slovic, P., and Tversky, A., eds. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kray, L. J., and Galinsky, A. D. 2003. 'The Debiasing Effect of Counterfactual Mind-Sets: Increasing the Search for Disconfirmatory Information in Group Decisions'. *Organizational Behavior and Human Decision Processes* 91 (1): 69–81.

Kunda, Z. 1990. 'The Case for Motivated Reasoning'. *Psychological Bulletin* 108 (3): 480–98.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Miriam J. Metzger, M., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S., Sunstein, C., Thorson, E., Watts, D., and Zittrain, J. L. 2018. 'The Science of Fake News'. *Science* 359 (6380): 1094–96.

Lilienfeld, S. O., Ammirati, R., and Landfield, K. 2009. 'Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare?' *Perspectives on Psychological Science* 4 (4): 390–98.

Mackay, C. 2012. *Extraordinary Popular Delusions and the Madness of Crowds*. New York: Simon and Schuster.

Martin, N. 2008. *Examination of the Belief Bias Effect Across Two Domains of Reasoning*. UWSpace.com. https://uwspace.uwaterloo.ca/handle/10012/3751.

Mazar, N., Amir, O and Ariely, D. 2008. 'The Dishonesty of Honest People: A Theory of Self-Concept Maintenance'. *Journal of Marketing Research* 45 (6): 633–44.

McDougal, W. 1920. *The Group Mind*. Cambridge: Cambridge University Press.

Mercier, H., and Sperber, D. 2011. 'Why Do Humans Reason? Arguments for an Argumentative Theory'. *Behavioral and Brain Sciences* 34 (2): 57–74.

Mullen, B., Brown, R., and Smith, C. 1992. 'Ingroup Bias as a Function of Salience, Relevance, and Status: An Integration'. *European Journal of Social Psychology* 22 (2): 103–22.

Nickerson, R. S. 1998. 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises'. *Review of General Psychology* 2 (2): 175–220.

Pennycook, G., Cannon, T. D., and Rand, D. G. 2018. 'Prior Exposure Increases Perceived Accuracy of Fake News'. *Journal of Experimental Psychology: General* 147 (12): 1865–80.

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., and Fugelsang, J. A. 2014. Cognitive Style and Religiosity: The Role of Conflict Detection. *Memory and Cognition* 42 (1): 1–10.

Pennycook, G., and Rand, D. G. 2018. 'Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning Than by Motivated Reasoning'. *Cognition* 188:39–50.

Polage, D. C. 2012. 'Making Up History: False Memories of Fake News Stories'. *Europe's Journal of Psychology* 8 (2): 245–50.

Rajaram, S., and Pereira-Pasarin, L. P. 2010. 'Collaborative Memory: Cognitive Research and Theory'. *Perspectives on Psychological Science* 5 (6): 649–63.

Redlawsk, D. P., Civettini, A. J. W., and Emmerson, K. M. 2010. 'The Affective Tipping Point: Do Motivated Reasoners Ever "Get It"?' *Political Psychology* 31 (4): 563–93.

Roediger III, H. L., and DeSoto, K. A. 2016. 'Recognizing the Presidents: Was Alexander Hamilton President?' *Psychological Science* 27 (5): 644–50.

Rubin, D. C., ed. 1996. *Remembering Our Past: Studies in Autobiographical Memory*. Cambridge: Cambridge University Press.

Singh, K. A., Gignac, G. E., Brydges, C. R., and Ecker, U. K. H. 2018. 'Working Memory Capacity Mediates the Relationship Between Removal and Fluid Intelligence'. *Journal of Memory and Language* 101:18–36.

Spinney, L. 2017. 'How Facebook, Fake News and Friends Are Warping Your Memory'. *Nature News* 543 (7644): 168.

Svob, C., Brown, N. R., Takšić, V., Katulić, K., and Žauhar, V. 2016. 'Intergenerational Transmission of Historical Memories and Social-Distance Attitudes in Post-War Second-Generation Croatians'. *Memory and Cognition* 44 (6): 846–55.

Tsipursky, G., Votta, F., and Roose, K. M. 2018. 'Fighting Fake News and Post-Truth Politics with Behavioral Science: The Pro-Truth Pledge'. *Behavior and Social Issues* 27:47–70.

Van Bavel, J. J., and Pereira, A. 2018. 'The Partisan Brain: An Identity-Based Model of Political Belief'. *Trends in Cognitive Sciences* 22 (3): 213–24 .

Verkuyten, M., and Nekuee, S. 1999. 'Ingroup Bias: The Effect of Self-Stereotyping, Identification and Group Threat'. *European Journal of Social Psychology* 29 (2–3): 411–18.

Wang, Q. 2008. 'On the Cultural Constitution of Collective Memory'. *Memory* 16 (3): 305–17.

6

Modelling the Propagation of Fake News in Social Media

Dr. Pardis Pourghomi

While sociograms, the science of modelling human social interactions, has been evolving for the last eighty years, the challenges of modelling digital social-media interactions provided new avenues to explore (Correa 2014). Researchers have attempted to model the spread of information as well as misinformation in various ways. In the early stages, these attempts revolved around modelling several nodes (users) communicating to a group of users; the model then evolved into a two-dimensional network, and finally, it ended up with a three-dimensional model. The main idea of this chapter is to present the latest findings and to set out the premise that the ways in which individuals react to false news is comparable to epidemiology, or the science of spreading diseases. Some individuals may remain naturally immune, while some may be vulnerable. Some would be afflicted and the rest may not. This chapter demonstrates four different scenarios in which the authors have modelled the spread of misinformation empirically, mathematically, or through the lab, using simulations to showcase their results. To fight the spread of false information, scholars and industries need a deeper understanding of the growth of misinformation on social-media channels. In this light, the present findings give an insight into the source of fake-news stories and the patterns behind them, both being vital to enrich the quality of information flow in cyberspace.

Although previous research has raised doubts about the role of fact-checking efforts and crowdsourcing corrections (Flynn, Nyhan, and Reifler 2017), it is nevertheless essential to recognize the socio-technical factors embedded in a networked system to design interventions and algorithms that interrupt the information flow from noncredible sources. For instance, if the information from noncredible sources or related bot accounts is identified, computer

scientists can generate algorithms that would reduce the appearance of such information. Such a methodology would be mainly effective if a sufficient number of sources are connected to the fake-information point of origin. Lastly, in parallel to the algorithmic approach, educational efforts such as literacy interventions on information and digital media can notify the digital public about such issues and diminish related misperceptions (Mele et al. 2017). The understanding from the primary studies has pertained to the basic interactions that users undergo. In most cases, users are described as nodes on a network. Scientists have analysed human behavioural aspects while dealing with the fake news that is online, with several perspectives such as the authentication process, correlation to disease spread, and lab tests involving small- and large-scale data simulations.

MODELLING USERS' AUTHENTICATION PROCESS

A study of relevant literature and open-ended survey with a sample size of 2,501 responses from Singaporeans undertaken by Tandoc et al. (2018) suggested a conceptual framework on how individuals authenticate the information they come across on social-media platforms. Broadly, the study suggests that individuals depend on both their own reasoning and the source of the data. On the other hand, when this does not adequately provide a concrete answer, they look for external resources to verify and authenticate news items, as presented in figure 6.1.

The conceptual framework, which is referred to as audience's acts of authentication (three As), argues that individuals are inclined to participate in

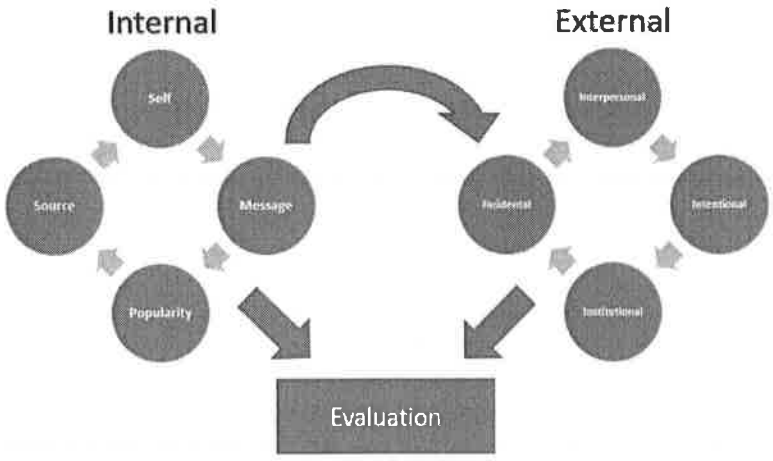

Figure 6.1. Audience acts of authentication.

authentication, users are capable of turning into nonbelievers and spread the same manner [AQ] with a mechanism similar to that of hoax-spreaders. As time goes by, both the believers and nonbelievers can return to a vulnerable state. The characteristics of this model are the following parameters: spread rate, truthfulness, the likelihood of verifying a hoax, and forgetting one's present belief. The authors seek a diversified network model with a variety of parameters to come up with a threshold; this limit looks at the fact-checking prospect, which promises the complete deletion of fake new from a certain network. The writers are certain that field approximation is the variable that depends on truthfulness only and forgets that probability to be the key factor.

AQ: It is not clear what "spread the same manner" means; please clarify.

First, the writers considered how long a hoax survives in a network, that is, the length of time in which the users believe the fake news is true. Majority of the nodes are *susceptible* (S) by definition. Naturally, hoaxes are very comparable to common viruses. That is, individuals on social media, as nodes, can become *infected* (I) and trust the fake news once they are in contact with other infected nodes. On the contrary, they can also *recover* (R) simply through a fact-checking action. Hence, the writers consider compartmental epidemic models, like SIR (Susceptible-Infected-Recovered) and SIS (Susceptible-Infected-Susceptible) (Bailey 1975), where nodes are described by several manners symbolized by states, and the dynamical development of the system is run by conversion rates among the states. They suggested a stochastic epidemic model to define the parallel diffusion of a hoax and its relative debunking: it can be seen as an SIS model in which the *infected* status is split into two subcompartments, *believers* (B) and *fact-checkers* (F), and the transition I → S can be understood as a forgetting process. Additionally, they have the transition B → F with a fixed probability p_{verify} which specifies the segment of infected users that check the trustworthiness of the received information, revealing the hoax.

There was a simultaneous simulation of the hoax spread, as well as its debunking, taking into consideration some users believe the fake news while others do not; and that is due to the verification of information, since the user already knows the information is not accurate. Hence, a model by Trpevski, Tang, and Kocarev (2010) for the competitive spread of two rumours was built to define the opposition between the believers and fact-checkers. Then, they continued the model by announcing verifying and forgetting processes. They took into consideration a network demonstrated by a graph G = (V, E). Each node, i, is linked to a triple of binary indicators, demonstrating its state at the time, t, which can assume one of the three possible values (see figure 6.2).

$$\forall i \in V \qquad s_i(t) = [s_i^B(t), s_i^F(t), s_i^S(t)] = \begin{cases} [1,0,0] \\ [0,1,0] \\ [0,0,1] \end{cases}$$

Figure 6.2. Representation of spread of fake news.

Table 6.1. Audiences' Acts of (External) Authentication

	Incidental	*Intentional*
Institutional	Stumbling upon news posted by other media	Searching on Google Verifying with mainstream sites
Interpersonal	Seeing more friends in social media sharing it	Asking friends, family members, and experts

a two-step authentication process: first internal and then external (see figure 6.1). The internal act of authentication pertains to the individual's preliminary encounter with social-media news. In the primary encounter, individuals depend on three main authentication framings: (1) the self, (2) the source, and (3) the message. If in the initial stage the individual is pleased with the information's legitimacy, then the process is ended, and the information is accepted as authentic. The conceptual framework here is a two-step approach to verification concerning internal primarily and external activity consequently. Each approach relates to both the news item and the individual. Hence, individuals at first look at the trustworthiness indicators within the story (the message, source of the message, and the message style) and within themselves (previous information awareness of the message subject and the source, which is an inborn reaction to the news item based on prior knowledge). Since the individual himself and the story can be subject to trust, it challenges the belief that trust lies with the third party. Hence, the relationship between the two types of internal trust, which are the individual and the news item, gives this conceptual framework the right to question the role of trust in how the individual endorses news items. On the contrary, if the individual is left sceptical about the information's authenticity, then he or she carries on to the next stage, which is the external acts of authentication. In this phase, the external authentication strategies are employed, which either can be planned or be accidental, by relying on interpersonal and institutional resources. That is, individuals consciously look for ways to authenticate news items using their personal contacts or seek authentication in formalized sources (see table 6.1).

MODELLING FAKE NEWS AS RESEMBLING DISEASE

One of the early modelling frameworks pertaining to circulation of fake news in the form of online hoaxes was presented by Tambuscio et al. (2015). Typically performed in mathematical modelling of information-diffusion processes, Tambuscio et al. considered hoaxes as viruses. The writers of this study suggested a stochastic epidemic model to define the spread of a hoax in a social network. In such a situation, users can catch the 'disease' once they are exposed to fake news and consequently pass on the disease. Upon

Correlating with the three potential behaviours for agent i. The authors drew the following three phenomena:

spreading [S → B, S → F]: each agent adapts ~~with some likelihood~~ its state [AQ] considering the points of view (states) of its neighbours;
verifying [B → F]: each agent can fact-check the hoax with the fixed probability p_{verify};
forgetting [B → S, F → S]: each agent, irrespective of belief state, overlooks the news with a fixed probability p_{forget}.

AQ: The meaning of the phrase each agent adapts with some likelihood its state" is not clear; is there a word missing?

Thus, Tambuscio et al. (2015) demonstrated a simple model of the various states a node takes place while operating online (see figure 6.3).

Studying this model and the formula facilitated the recognition of some theoretical findings. The writers found that the dynamics of believers and fact-checkers are responsible for the victory of one of the two behaviours. Increasing fact-checking has the power to eliminate the hoax; nevertheless, the whole infection stays active in the network. The study reveals that a somewhat minor fact-checking activity can stop the hoax, even while the user is motivated to believe the information with high possibility. Based on the analytics, they came up with a theoretical threshold for p_{verify} to assure the elimination of fake content. They implemented and tested the model on heterogeneous (scale-free), homogeneous (random), and real networks, varying parameters and topology. They focused on analysing the crucial role of fact-checking activity, ruled by a verifying probability. Analytically, they found a threshold for this probability, a sufficient condition that assures the hoax will be removed.

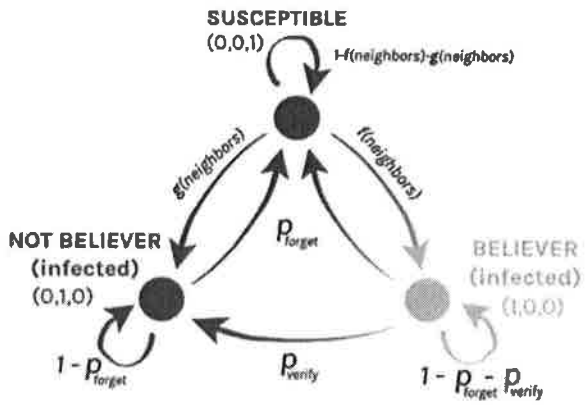

Figure 6.3. States and transitions of the model (Tamusico et al., 2015, p. 978).

SMALL-SCALE SIMULATIONS

Budak, Agrawal, and Abbadi (2011) present the first attempt to model the travel of information or misinformation online (see figure 6.4). In this model, there are multiple starting nodes: 0, 4, and 7, suggesting that misinformation can spread from a variety of sources beyond social media. Misinformation can crisscross and travel in a variety of ways. The paper uses network algorithm to test two competing campaigns as a means of testing the accuracy of the information. In such a scenario, Budak et al. suggested the need for 'influential' people to counter the 'bad' campaign and limit misinformation as a means to fight misinformation. Nowadays, with a large volume of data scattering on social media, finding a true and reliable piece of information requires sifting-out different types of misinformation, which is computationally a difficult task.

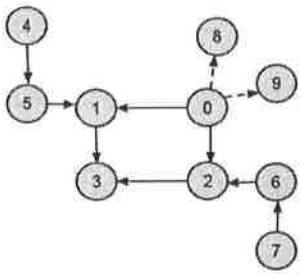

Figure 6.4. Interaction between nodes (Budak, Agrawal, and Abbadi, 2011, p. 669).

Yu and Kak (2012) looked at predictor characteristics of a message and the social network that could determine if a message goes viral or not. It also predicts how far and for how long the message could travel. Few key findings could be used, including an attempt to model the social-media network (see figure 6.5).

First, they explain why automated prediction has a greater impact than human prediction. They argue that automated prediction has no involvement of human desire in making decisions, it comes at a much lower cost, and it could process a lot more volume of data than humans can do. They categorized the prediction metrics in two: (1) message characteristics and (2) social-media characteristics. The former focuses on the message itself by considering its sentiment—statistical features such as happiness and anxiety on a case-by-case basis—and time stamp series that refers to the speed and process of message generation. The latter, however, focuses on measuring structure features such as terminologies used in social media, for example, follower, followee, node, and so on; degree: in-degree means the number of directed lines to this node, out-degree means the number of its tail endpoints

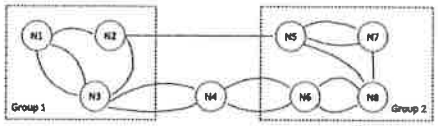

Figure 6.5. An example of a directed social network.

(followees), and total degree is the sum of in-degree and out-degree; density of followers and followees; betweenness centrality that measures the control of a node on the communication between other nodes; closeness centrality, which refers to the distance of two nodes within their shortest path; group degree centrality where the extent the degree from the node level to the group level is measured; and effective size and the efficiency. Furthermore, they present several methods for prediction in social media. Table 6.2 illustrates the summary of these methods.

Dordevic et al. (2016) presented a two-dimensional simulation that provided bases for a proof-of-concept and identification of key variables. The authors set out to demonstrate proof-of-concept using 2D modelling and identified the variables involved in the travel of information. They identified eight key variables and applied theoretical values to demonstrate their applicability. These variables are: i as the first vertex and jn is the last vertex of the given simulation, $V1$ representing the first phase of the spread of misinformation and l representing the maximum possible reach of information through the network. They concluded that combating misinformation online is also influenced by the following variables: rate the of authentication A,

Table 6.2. Prediction Methods in Social Media

Method	*Description*
Regression method	Analyses relationships between the dependent variable, prediction result, and one or more independent variables, such as the social-network characteristics.
Bayes classifier	Bayes classifier is a probabilistic classifier using Bayes' theorem (Sharda and Delen 2006).
K-nearest neighbour classifier	It is one of the simplest machine-learning algorithms, which clusters the objects according to their distance to others.
Artificial neural network	A computational model for human brain simulation (Kak, Chen, and Wang 2010).
Decision tree	A visual technique in data mining by which travelling from the root node to leaf enables one entity to get the prediction results.
Model-based prediction	Requires mathematical modelling to undertake the prediction. It remains an open and challenging topic.

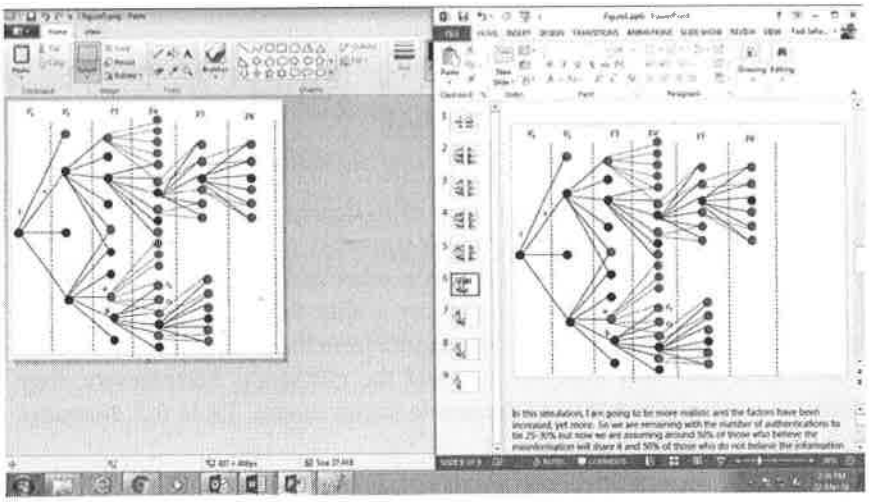

Figure 6.6. Passing on rate and cross-wire rate simulation (Dordevic et al., 2016, p. 450).

rate of sharing *S*, passing on information rate *P*, average cross-wire rate *Cw*, success rate of the same-level communication rate *Sl*, and reverse validation rate *Rv*. They demonstrated near elimination of spread of misinformation online whereby the red nodes represented individuals who do not longer believe the misinformation. This demonstrated that providing easily accessible tools that would allow users to authenticate images and text could effectively cascade the process back to the source or at least to the layer immediate to the source (see figure 6.6).

BIG DATA SIMULATION

One of the first attempts to model social-media networks was done by Hewlett-Packard, which uses a software called GUESS to model the spread of information inside a company social network as shown in figure 6.7 (Adar 2007). The significance of this simulation is that it considered that the spread of information always starts from a source and then propagates across a spatial space where nodes interlink and interconnect, passing individuals and groups, and reaching some point its population saturation where it slows down and stops.

In another paper, Pourghomi, Dordevic, and Safieddine (2018) used big data model Biolayout Express three-dimensional modelling to expand their simulations of different scenarios. Biolayout is a tool that has been used in the past to model epidemiology events and thus seems to fit the needs to

Figure 6.7. An example of the social network from GUESS (Adar, 2007).

model how information and misinformation travel on social media. Using the given variables and values, this paper presents a better understanding of how misinformation travels in the spatial space of social media. Using a population of one hundred nodes, the simulation assumed no authentication tool or effort to authenticate by users, rendering figure 6.8 where all nodes are coloured in blue.

The authors conducted several simulations that looked at variables linked to users' behaviour online, such as sharing of a post, liking, commenting, and reposting of articles. The authors also looked at the impact of validation in slowing or stopping the spread of fake news. Using several permutations of variables, the authors suggested that having an accessible tool to validate fake news could stop the propagation of fake news where 30 per cent of the

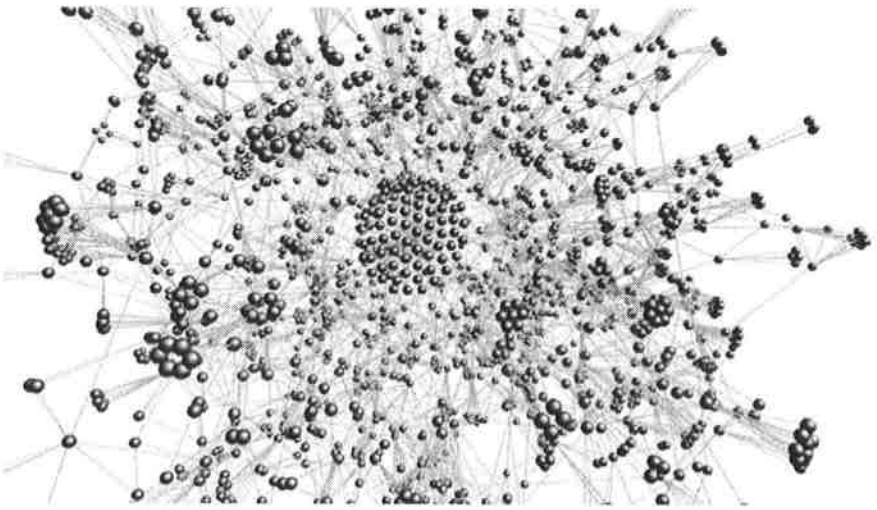

Figure 6.8. Three-dimensional simulation with no authentication.

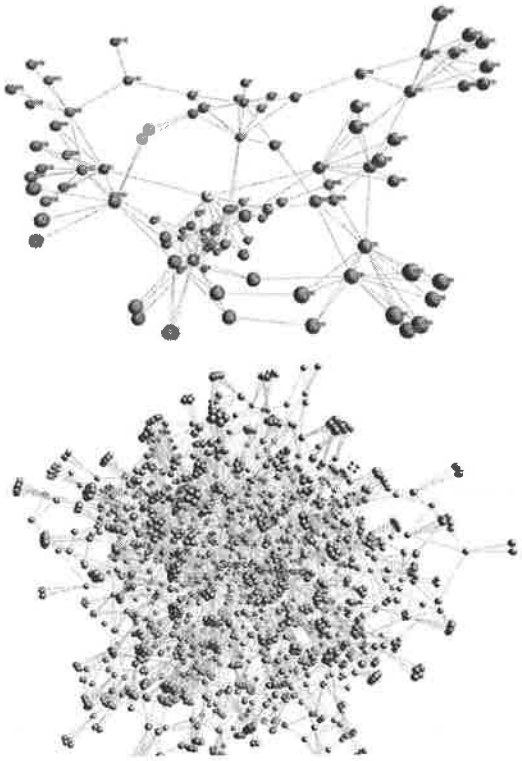

Figure 6.9. Three-dimensional simulation with and without influencers.

population takes an active interest in validating and taking any number of actions to respond to fake-news posts. This was demonstrated in figure 6.9, where the red nodes show infected nodes that have recovered following 30 per cent of the population validating given posts. But where we have influencers, that rate goes up to 54 per cent.

Simulations further confirmed that the number of users becomes irrelevant as the extremities of the 'tree' are eliminated, and that misinformation will be confined to the first source and first-line, beyond which misinformation will not be able to travel easily.

TREE MODELLING METHOD

A study conducted by a team of social scientists and computer science researchers (Jang et al. 2018) proposed a novel framework for analysing,

identifying, and predicting the flow of misinformation on digital platforms. Researchers examined the methods by which misinformation is initiated and its transmission occurs. They also studied the evolvement of misinformation in the online news system and looked at how it is managed. Predominantly, they investigated the sources and authors by whom fake news is generated and identified fake news' antecedent tweets. Their study revealed that different evolution patterns are generated through the transmission of real and fake news. Researchers obtained Twitter data provided by Crimson Hexagon. Crimson Hexagon is a private corporation that collects data such as blog entries, Facebook posts, and tweets from social-network sources. In their study, researchers considered the propagated real and fake news on Twitter. They selected sixty news stories, of which thirty were fake and thirty were real. To further justify why they used sixty news stories provided by Crimson Hexagon, they explain that they looked into several news sources such as newspapers, political blogs, and online news. Then, they collected one hundred news stories and checked the popularity of that news through Crimson Hexagon's database according to their media coverage and rankings. Next, they chose their sixty news stories based the number of times that news was tweeted from the database. Finally, to ensure the accuracy of real and fake news within those sixty news topics, they used two fact-checking websites (factcheck.org. and snopes.com).

The approach they used to analyse the flow of misinformation is similar to biological phylogeny. A phylogeny tree is a diagram that includes branches and trees that can help demonstrate evolutionary relationships between different entities based on similarities and differences in characteristics (Robinson and Foulds 1981).**[AQ]** is done on the text. Text phylogeny is conducted to create an evolution tree and illustrate the evolutionary process of a set of tweet texts on the same topic (Jang et al. 2018). Researchers first calculated the string distance between every pair of tweets and combined them into a matrix to build the evolution tree. The method by which the algorithm works is that it takes the matrix as an input and then it outputs a minimum spanning tree that shows all the string distances. As part of the process, priority queues, distance matrix, cleaning the shared tweets, and other mechanisms were also included for building the phylogenetic tree to show a substantial difference between how fake news travels and how factual news travels, as shown in figures 6.10 and 6.11, respectively.

AQ: There seems to be missing text here.

Figure 6.10 represents the propagation of a fake-news tweet that allegedly suggested that Donald Trump was born in Pakistan, whereas figure 6.11 represents the propagation of factual-news tweet that suggested Barak Obama emailed Hilary Chilton on her private server. Analysis' results demonstrated that real and fake news have significant differences in network character-

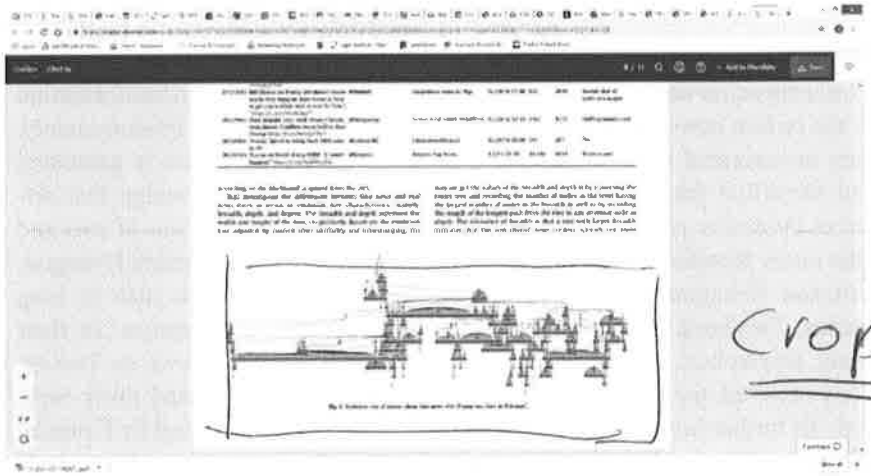

Figure 6.10. Fake news propagation using Evolution tree (Jang et al., 2018, p. 110).

istics. For example, the analysis revealed that on average, fake-news trees showed greater depth, whereas real news displayed greater breadth. Changes in the content are normally shown in the depth of the evolution tree analysis, and wider dispersion is an indication of greater breadth. What is understood from their analysis is that real-news stories have been shared and retweeted many times without being modified, but fake-news stories were modified several times as they were being disseminated. Their interpretation from the analysis is that when real-news topics are shared and tweeted, they are not

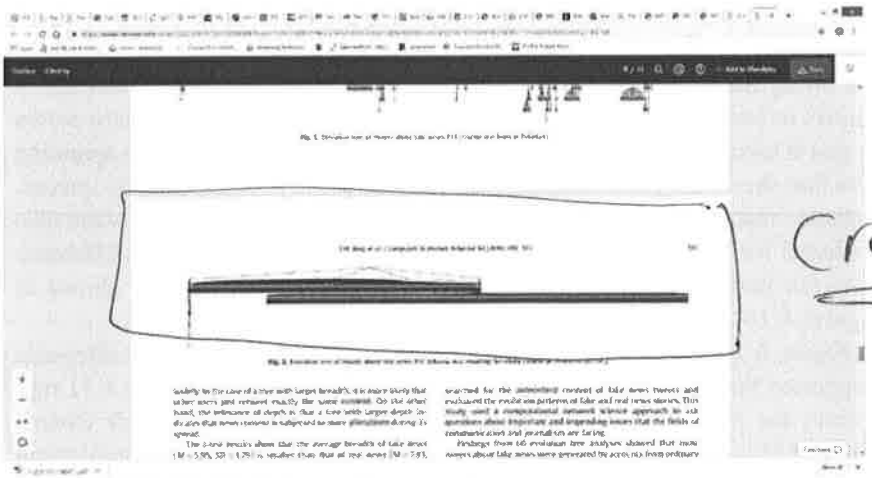

Figure 6.11. Factual news propagation using Evolution tree (Jang et al., 2018, p. 110).

being commented upon, or their content is not modified as opposed to the fake news, which presented the opposite characteristics. Moreover, they claim that factual news stories are those that are shared as soon as the news is published for the first time, that is, by a reliable news agency or website, and this happens in a very short period. Great breadth and shorter depth of these news patterns are the main characteristics of factual news. In contrast, fake news takes a longer time to be shared across social media, and it tends to attract more comments from those users who read and/or share the news across the platform. This shows that those who tweet fake news are more cautious and at the end, they either share it with major modifications to the news itself or they do not share it at all.

The researchers' analysis thus provides some insights about the initiation of fake news. For instance, they stated that those ordinary users who were not known as journalists, celebrities, or politicians tweeted 87 per cent of fake-news stories—that is twenty-six out of thirty fake-news items. Out of thirty fake-news stories, only two nonaverage users initiated the sharing of fake news and the other two accounts that also shared the fake news had already been identified and suspended by Twitter. Based on their analysis, the researchers estimate that the 43 per cent of fake-news stories—that is thirteen out of thirty—were first posted on Twitter with a link to noncredible news agencies such as those who mimic the logo of credible news agencies. On the other hand, 43 per cent of factual news—again, thirteen out of thirty—were first triggered by news media, journalists, or political groups and the other seventeen factual news stories were shared by ordinary users. These thirteen real news stories included certain news links; among them, ten real-news stories delivered news links to mainstream media such as nytimes.com and online news media such as huffingtonpost.com.

CONCLUDING REMARKS

This chapter discussed the methods used by different authors to model the spread of misinformation, and how they can be identified, evaluated, and potentially stopped. Further details were provided to describe the statistics and conclusions that were made as outcomes of explained research studies. There are clear parallels in the research approaches. Researchers have attempted to draw from real-world human interactions and translate them to the virtual world. Thus, in the process of conceptualizing fake news, researchers have linked fake news to epidemiology, the science of disease spread in an epidemic. At each level, we see researchers examining the changing status of an individual, the way an illness changes one' status. We also see how

small-group interactions changed the dynamics of peer-to-peer conversations. Finally and as we move up to larger and big-data analysis, we see how fake news explodes in the form of an unstoppable disease that is only slowed by either the size of the population or a remedy some researchers referred to as the ability to validate the news. All these contribute to a better understanding of how interactions develop online and fake news spreads. It is clear that much more research into this area is needed, not least because the landscape is constantly evolving. We hope such further research will provide practical and policy-making responses to problems of online disinformation with a more robust base of evidence.

REFERENCES

Bailey, N. T. 1975. *The Mathematical Theory of Infectious Diseases and Its Applications*. Second edition. London: Griffin.

Budak, C., Agrawal, D., and El Abbadi, A. 2011. 'Limiting the Spread of Misinformation in Social Networks'. In *Proceedings of the 20th International Conference on World Wide Web* (pp. 665–74).

Correa, C. D. 2014. 'History and Evolution of Social Network Visualization'. *Encyclopedia of Social Network Analysis and Mining* (pp. 677–95).

Dordevic, M., Safieddine, F., Masri, W., and Pourghomi, P. 2016. 'Combating Misinformation Online: Identification of Variables and Proof-of-Concept Study'. In *Conference on e-Business, e-Services and e-Society* (pp. 442–54). New York: Springer.

Flynn, D. J., Nyhan, B., and Reifler, J. 2017. 'The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics'. *Political Psychology 38*:127–50.

Hewlett-Packard. 2007. *GUESS, the Graph Exploration System*. http://graphexploration.cond.org/.

Jang, S. M., Geng, T., Li, J. Y. Q., Xia, R., Huang, C. T., Kim, H., and Tang, J. 2018. 'A Computational Approach for Examining the Roots and Spreading Patterns of Fake News: Evolution Tree Analysis'. *Computers in Human Behavior 84*:103–13.

Jin, L., Chen, Y., Wang, T., Hui, P., and Vasilakos, A. V. 2013. 'Understanding User Behavior in Online Social Networks: A Survey'. *IEEE Communications Magazine 51* (9): 144–50.

Kak, S. C., Chen, Y., and Wang, L. 2010. 'Data Mining Using Surface and Deep Agents Based on Neural Networks'. In *AMCIS* (August): 16.

Mele, N., Lazer, D., Baum, M., Grinberg, N., Friedland, L., Joseph, K., Hobbs, W., and Mattsson, C. 2017. *Combating Fake News: An Agenda for Research and Action*. Harvard Kennedy School. https://shorensteincenter.org/wp-content/uploads/2017/05/Combating-Fake-NewsAgenda-for-Research-1.pdf.

Pourghomi, P., Dordevic, M., and Safieddine, F. 2018. 'The Spreading of Misinformation online: 3D Simulation'. In *2018 IEEE 5th International Conference*

on Information Technology, Computer, and Electrical Engineering (ICITACEE) (September): 299–304.

Pourghomi, P., Safieddine, F., Masri, W., and Dordevic, M. 2017. 'How to Stop Spread of Misinformation on Social Media: Facebook Plans vs. Right-Click Authenticate Approach'. In *2017 IEEE International Conference on Engineering and MIS (ICEMIS)* (May): 1–8.

Robinson, D. F., and Foulds, L. R. 1981. 'Comparison of Phylogenetic Trees'. *Mathematical Biosciences* 53 (1–2): 131–47.

Sharda, R., and Delen, D. 2006. 'Predicting Box-Office Success of Motion Pictures with Neural Networks'. *Expert Systems with Applications* 30 (2): 243–54.

Tambuscio, M., Ruffo, G., Flammini, A., and Menczer, F. 2015. Fact-Checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks. In *Proceedings of the 24th International Conference on World Wide Web* (May): 977–82.

Tandoc Jr, E. C., Ling, R., Westlund, O., Duffy, A., Goh, D., and Zheng Wei, L. 2018. 'Audiences' Acts of Authentication in the Age of Fake News: A Conceptual Framework'. *New Media and Society* 20 (8): 2745–63.

Trpevski, D., Tang, W. K., and Kocarev, L. 2010. Model for Rumor Spreading over Networks. *Physical Review E* 81 (5): 056102.

Yu, S., and Kak, S. 2012. 'A Survey of Prediction Using Social Media'. *arXiv preprint arXiv:1203.1647*.

7

Corporate Attempts to Combat Fake News

Dr. Milan Djordjevic

This chapter looks at the reported attempts to combat fake news by different social media and search engines from Google's Fact-Checkers website links to Facebook *Report Fake News* and *War Room* strategies. This chapter draws parallels between the spread of computer viruses in the late 1990s and early 2000s, providing evidence that while using antivirus software had reduced the spread of computer viruses, the real impact of reduction only came about when big email providers included built-in filters that stopped their spread. In this chapter, we present the work done by social-media companies in combatting fake news. These papers have explored proposed approaches and techniques that are based on platform policies, digital algorithms, filter, and artificial intelligence (AI) in the process of detecting fake news, as shown in figure 7.1.

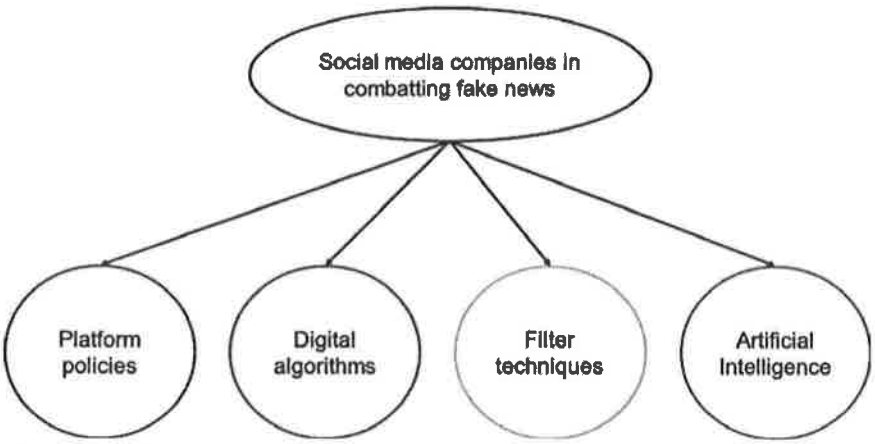

Figure 7.1. The work done by social media companies in combatting fake news online.

Drawing from these examples, the chapter will conclude that the key tool to combating misinformation remains in the hands of big corporations in providing tools that support users in validating information sources online.

One of the main features of social-media networks is that individuals can share information, and this information may be shared and propagated relatively easily to a large number of users. The quality of information shared on the network is a concern as more users rely on social-media platforms as their main source of news (Halimeh, Pourghomi, and Safieddine 2017).

PLATFORM POLICIES

Role of Social-Media Platforms

Research by Verstraete, Bambauer, and Bambauer (2017) categorized important problems social-media platforms have in combatting fake news. The researchers described a set of intervention models, grouped under four categories. First, the paper looks at the ease with which fake-news generators can provide online content that appears to be authentic. Second, the financial stakes that platforms have in sharing fake material online is considered. Third, the principles of the *filter bubble* seem to influence much of the findings and confirm our assessment of the impact of filter bubbles. Thus, clients of fake news have a strong disposition to accept the news, and thus limited incentive to participate in challenging and validating its content. Finally, a key component in making fake news believable is to introduce fibres of truth to make it believable. Thus, the research suggests some potential solutions such as changes to codes, norms, laws, and markets. Respectively, all of them have advantages and disadvantages. Some legal solutions to fabricated news are arguably going to conflict with the constitutional and statutory protections for freedom of speech. Many of these solutions are likely only to reach a subset of fabricated news. For example, code solutions may be constrained by the challenging judgments essential to separate satire from other types of fabricated news. But it is the combination of all four, and the input from social-media platforms, that will likely make the leap in handling fake news.

Lazer et al. (2019) present an investigation on a platform-based policy in managing and combatting fake news. The work looks at both self-regulated social-media platforms or those subjected to government intervention. Direct government intervention may be perceived as curtailing and biased as such government-based policies may impact freedom of speech. The authors instead suggest that the US government facilitate an approach dubbed *tort lawsuits alleging* where policies would be shaped by lawsuits initiated by people affected or harmed by the impact of fake news. To materialize this approach,

social networks could provide users with indications of source quality, which can be merged into the algorithmic rankings of news. Additionally, these platforms can reduce the personalization of political information content and allow alternative reporting of news and views—thus removing users from filter-imposed bubbles. Methods that at present underline trending news could be redesigned to allow a variety of news.

Closing Down Fake Profiles

The big corporations have attempted several preventative measures to a varying degree of success. In 2018, Lazer et al. reported several actions taken by social-media platforms to address the spread of fake news. Twitter blocked some accounts connected to Russian users and other users connected to those accounts. Facebook announced a determination to shift its algorithm to check for quality in its content-curation process. Nevertheless, the social networks have not opened up about their techniques for evaluation by the research community nor shared their results for peer review, which can support their efforts. There exists a huge need for social networks to work with academic researchers in assessing the scope of the fake-news problem and the design and implementation of actions against the sharing of fake news.

Mark Zuckerberg announced after the 2016 US election that 'Of all the content on Facebook, more than 99 per cent of what people see is authentic. Only a very small content is fake news and hoaxes. Overall, this makes it extremely unlikely hoaxes changed the outcome of this election in one direction or the other' (cf. Kokalitcheva 2016). But by mid-2019, Facebook claimed to have closed millions of fake accounts, with some reports suggesting the platform closed as much as three billion accounts (Segarra 2019). Evidently, at the time, even heads of social-media giants were not well informed of the challenges they are facing.

Third-Party Checkers

Third-party checkers are defined as an independent entity, usually a collection of volunteering journalists, who take it upon themselves to validate or rebuke trending stories on social media. In late 2016 and following the US elections, Facebook announced means to report fake news that supposedly makes use of independent third-party checkers to verify news' authenticity (Mosseri 2016). Facebook seemed to react to a public backlash that the platform had failed to respond to constant reports suggesting that its platform is encouraging the propagation of fake news. The Facebook approach was designed to help users decide for themselves what to trust and what to

share. The Facebook authentication process is associated with third-party fact-checking societies that are signatories of Poynter's International Fact-Checking Code of Principles (Poynter Inc. 2016).

When users report stories to be fake, these reports are sent to third-party checkers societies. If many individuals report a post as fake, the platform puts an indication on the post that the post is questionable news, and the report is forwarded to third-party checkers for verification. Should the fact-checking organizations categorize a story as fake, the post would get flagged as disputed. A follow-up message would pop up to notify the users who reported the posts as confirmation of the decision by the third-party checkers. Pourghomi et al. (2017), pointed out that the process has a major problem in the fact that it is slow. What is more, the process is platform-based and thus would not stop the news spreading in another platform. Finally, at best, the post is flagged as 'disputed' and reasoning behind that decision is not provided.

Other Approaches

There is little in the literature about policy-based approaches being taken by platforms to combat fake news. But a few activities have been observed and reported in the news. Google News attempted to provide 'fact check' links to news articles on its trending news, but that proved to be a pointless exercise as changes to the way the algorithms works meant that much of the news trending on the Google News page was now linked to reliable news sources. Google eventually relegated the fact-checking websites and their links to the right-hand side column as they posted reviews of current nonfactual fake news. The impact of this approach remains limited, as this is far beyond the limits of social-media platforms. The approach has the same limitation reported in the Facebook approach, namely a significant delay in processing articles and their authenticity.

In mid-2019, several media outlets reported that Facebook set up a *War Room* in Ireland ahead of Europe's Parliamentary elections to monitor for misinformation, fake accounts, and election interference that violates the site's rules. This follows similar War Rooms set up ahead of the 2018 mid-term elections in the United States, Brazil's presidential elections, Spanish elections, and Indian elections. The room is said to be staffed by forty people who are fluent in all twenty-four EU languages. The team reviews materials flagged by their system algorithms. Where violations are detected, accounts could be closed, and pages would be unpublished. There is still no comprehensive assessment on how successful this or any of the approaches have had on the propagation of fake news.

DIGITAL ALGORITHMS

An Automated System Based on Datasets and Crowdsourcing

Buntain and Golbeck (2017) claim to have developed a process for automating fake-news detection on Twitter. The authors claim to have achieved detection of fake news by learning how to forecast precision assessments in two dedicated Twitter datasets. The first data set is a *credbank*, a crowdsourced dataset of precision assessments for events on Twitter. The second one is called *Pheme*. *Pheme* represents a dataset of potential gossip on Twitter and newspapers' assessments of their truthfulness. When implemented in a computerized system, the results are said to have been able to detect fake news in popular Twitter threads. As part of the process, the model includes the use of crowdsourced workers in updating and maintaining the datasets. In the same experiment, the authors found that crowdsourced workers outperform both newspapers' assessment and model-trained workers. The argument being that crowdsourced workers provide an easy way to categorize true and false stories on Twitter compared to newspapers. This model, while promising, continues to have two key risks associated with building trust in both the technological element and the human element. Analysis of findings suggests another significant result. The limited predictive feature overlap in *pheme* and *credbank* will require regular crowdsourcing input to evaluate precision in social media differently. As such, there is a risk of overrelying on crowdsourcing, and the authors proposed a solution. Namely, when a thread in Twitter starts with a story link and the story is defined by journalists as factual, then there is no need to command crowdsourced workers' assessment of that thread. Rather, it is this perception of correctness that commands how the story is shared. A potential limitation in this approach is that it focuses on popular Twitter threads only. The approach identifies highly retweeted threads of discussion and use the topographies of these threads to classify stories. This can potentially bound applicability of the approach to only the collection of popular tweets. The other concern comes from relying on crowdsourcing. There is no indication of how immune the system is from abuse by troll armies and fake profiles that could compromise the reliability of the crowdsourcing factor.

Labelling Posts as a Way to Reduce the Spread

Clayton et al. (2019) looked at trust in rating fake news to be fake and factual using labelling posts to be explicitly factual or otherwise general labellings. The authors suggested that combatting fake news online can have some degree of success when appropriate warning tags are used. The result of their

study found that both *Disputed* and *Rated False* tags modestly decrease trust in sharing fake news; an issue also confirmed in the studies of Ecker, Lewandowsky, and Tang (2010) and Bolsen and Druckman (2015). However, their results demonstrated that *Rated False* tags, which exactly articulate to users that the news is not true, appear to be more effective at decreasing trust in misinformation than the *Disputed* tags when used on Facebook. When compared to all-purpose notices that are not specific, it was also found to reduce trust in false headlines, but by a small margin. The authors have listed several limitations. For example, the sample they used was more educated and politically active than the general population. More importantly, the study did not consider the impact on *tribalists*, the hardcore supporters of a given view, as identified in chapters 1 and 8,. Next, their study did not examine effects over time and of course, as in any experimental study, they couldn't completely rule out the chance of demand effects. But despite these limitations, their study showed important perceptions into how efforts to stop trust in misinformation on social media could be more effective.

The Role of Bots and Cyborgs

Clark et al. (2016) investigated automated behaviour of bots on Twitter. Using a flexible and transparent classification scheme, the authors showed the potential of using linguistic features as a means of classifying automated bot activity on Twitter. As these features do not use the metadata provided by Twitter, the classification scheme is also applicable to other social-media platforms. The paper suggests that current algorithms-based approaches by social-media giants rely on metadata sets to find automated bot accounts by using variables such as the time between tweets, number of followers, and other variables. Clark et al. (2016), however, demonstrated a classification scheme that uses the natural language text from nonbot users to deliver a standard for detecting accounts that post automated bot posts. Researchers concentrated on three different classes of automated tweeting: *Spammers*, *Cyborgs*, and *Robots*. *Spammers* are semigenuine user accounts that take advantage of an automated program to post an eruption of tweets. *Spammers* regularly find ways to bypass social-media spam-detection protocols. In the context of this research, cyborgs are consumers who automatically share news from a set of sources, sometimes without even reading them at all (Shu 2017). The researchers call *Cyborg* accounts puppets because they masquerade as genuine users pretending to be another genuine user. One very interesting category of automated users are *Robots*, sometimes abbreviated as *Bots*. *Robots* can autogenerate messages but have a limitation that is characterized by having a limited vocabulary. Their posts follow a very structured

design. Robots have been used legally in business applications such as online assistants. There exists a possibility in the future for these different types of automated programs to hybridize together. Thus, we can expect in the future classes of algorithms like a robotic spammer, cyborg robot, or spamming cyborg. In a similar study, Varol et al. (2017) present a framework for the detection of bots on Twitter. The system is a machine-learning algorithm that can extract features divided into six different classes. The first and second classes are users' and friends' meta-data. Third and fourth classes are network patterns and activity time series. The last two classes are tweet content and sentiment. The system evaluates patterns, thus developing a framework which is trained on a dataset of bots. The results suggest an accuracy of 0.95 Area Under Curve (AUC); thus a good rate of precision but a margin of error that will need enhancing. One of the challenges of the system is that it cannot distinguish between bots and humans excessively using social media in a format that may mimic bots. Sometimes the border between these two groups is not obvious as some accounts show the characteristics of both.

The researchers urged social-media platforms to cooperate with academics on combating fake news online and to work together on design and implementations of interventions. There is little research focused on misinformation and no wide-ranging data-collection system to make available a dynamic understanding of how misinformation is growing. In general, academics must perform severe and dynamic research on how the major social networks filter information. Social networks could control the automated spread of news content by bots and cyborgs. Social-media cyborgs can spread fake news in a method that combines human input with automated activities. Typically, cyborgs are shown as humans in a disguise and set automated actions to do actions in social networks. The functionalities between human and bot can be switched very easily, which give cyborgs a great chance to share misinformation online. In general, good misinformation-intervention methods will try to remove cyborg accounts that share fake news or convince the users with true news to change their beliefs.

FILTER TECHNIQUES AND FILTER BUBBLES

Another approach to combating social media is presented by DiFranzo and Gloria (2017), where the authors attribute the spread of fake news to what they call 'filter bubbles'. Filter bubbles are the digital echo chambers where users see content and posts that agree with their preexisting beliefs. The social-media platform algorithms have been designed to show popular influencers, pages, groups, and posts based on a person's preexisting

set of choices. For example, where a person opens a new social-media account and joins social groups supportive of a particular football team, the algorithm is programmed to push forward fan groups, support groups, merchandise, influencers, current and past players, and so on. Relatively quickly, this social-media user finds themselves integrated into a subculture of fans and supporters of that group. This same principle applies to when a user joins and curiously enquires about antivaccination groups. By doing so, the platform inadvertently starts bombarding this user with posts and groups that promote antivaccination beliefs. While social-media platforms aim to encourage users with the same interests to share ideas and stay online longer, the unintended consequence is that these users end up living in an information bubble. By default, users find themselves in a filter bubble where they only see the one-sided view of an argument. In such an environment, fake news can go unchallenged and be further reinforced by the agreeability infested in being shared in a filter bubble. Personal recommendation systems, or systems that learn and react to individual users, have been identified as one cause of filter bubbles.

An investigation by DiFranzo and Gloria (2017) concluded that these practices implicated many social-media platforms in inadvertently playing a role in exacerbating the spread of fake news. Their work concludes that the dissemination of fake news may, at times, not break out of the filter bubble at its point of origin. In fact, the work of Jin et al. (2013) suggests that the dissemination of fake news is similar to the epidemic and that such stories may typically stay inside a closed group. More so, the authors provide a mathematical analysis formula that proves the similarity between epidemics and rumours. The formula uses compartmental population models to quantify the broadcast of news and rumours on Twitter, focusing on the susceptible-infected-susceptible model (SIS) as the primary model. Then they looked at the susceptible-exposed-infected-sceptic (SEIZ) model. Together, these two models are widely used in epidemiological modelling. The SIS model is common in infectious-disease modelling, where individuals can transition back and forth between susceptible (S) and infected (I) states that resemble viruses, allergies, and flu. On the other side, the SEIZ model takes the approach of introducing an exposed state (E). Users in such a state take some time before they get infected, actually before they begin to believe in fake news. Their results showed that the SEIZ model could suit the extended period of adoption with reasonable inaccuracy, but their analysis does not reveal whether this model can be functional on the large scale datasets. Moreover, it does not show if it can be applied to big social-media platforms, where the news discloses in real-time.

ARTIFICIAL INTELLIGENCE AND MACHINE LANGUAGE

Despite the many advances in algorithm-based detection of fake news, it seemed inevitable that Artificial Intelligence (AI) is the natural partner in constructing the knowledge that would adapt to variations of fake newsmakers. Cybenko and Cybenko (2018) showed an important application of AI in identifying fake news online. In a network vocabulary, AI would use *latency* in delays before a transfer of data begins. Then it would follow an instruction for its transfer to perform analysis. The latency that human fact-checkers bring into fake-news sharing seems to take too long to reduce the harm done by fake news. Thus, Cybenko and Cybenko suggest faster, reliable, and more stable AI methods that are needed in fighting misinformation online. The aim is to have AI methods that can categorize news as either genuine or fake based on two approaches. The first approach uses linguistic and semantic analysis of written content to make their identification. The second approach uses sharing patterns and rates to categorize news into rumours and nonrumours. Thus, it is claimed that a combination of both helps AI identify posts that are factual or fake.

Moreover, the researchers provided a summary of four key cognitive protections counter to accepting fake news, together with a selection of AI tools for overcoming and protecting those protections. The four key cognitive protections are described as follows. First is the content of the news being incompatible with the existing worldview. Second is the news storyline presenting an inconsistently coherent story. Third is news that does not come from what is considered a reliable source. And finally, the news is not rated as believable by social-media readers. One of the critical facts contributing to pressure on coming up with AI tools that would automate fake news is that so far, human-checker processes are proving to be slow and less efficient in responding to the challenges. Their response is deemed as late to mitigate the damage already done by the impact of fake news.

FUTURE RESEARCH DIRECTION

Combating fake news on social media is a newly developing research area, and as the work shows there are still several opportunities to explore for companies as well as for academic researchers who are willing to contribute to this field. The goal is to point out encouraging directions and open problems. In figure 7.2, we organize the research directions in three classes: data science, application-driven, and feature-driven open problems.

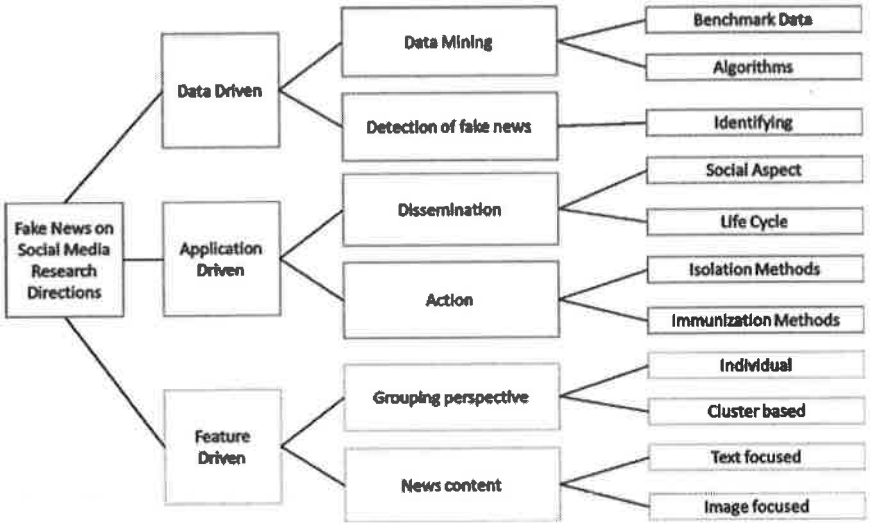

Figure 7.2. Open problems and future challenges.

Data-driven fake-news research is focusing on two kinds of data characteristics: data mining and detection of fake news. The data-mining approach will focus on establishing and analysing benchmark data that includes resources to extract appropriate structures. A promising direction is to generate a complete and large-scale fake-news benchmark data, which then can be used by researchers to enable further research in this area. The possibility is of testing different algorithms and methods in fighting fake news online. There are interesting issues related to data-driven open problems with regard to determining how to perform early fake-news detection. The aim is to give early warnings of fake news during the sharing process. Identifying and labelling fake news early can help prevent additional dissemination on social media. Researchers urge that social-media platforms need to cooperate with academics on combating fake news online, and that they work together to design and implement appropriate interventions. There is little research concentrated on misinformation and no wide-ranging data-collection system to make available a dynamic understanding of how misinformation is growing.

In future, we can expect that different types of automated programs will hybridize together. Future classes of algorithms that large-scale social-media platforms can use to fight bot-spammers, cyborg-bots, spamming-cyborgs, and cyborg-bot spammers (CBS), as illustrated in the Venn diagram in figure 7.3. Bot and spam detection are therefore going to be an important area of future research. Many social-media platforms use a range of tools, from machine learning, artificial intelligence, and data mining to perform social-

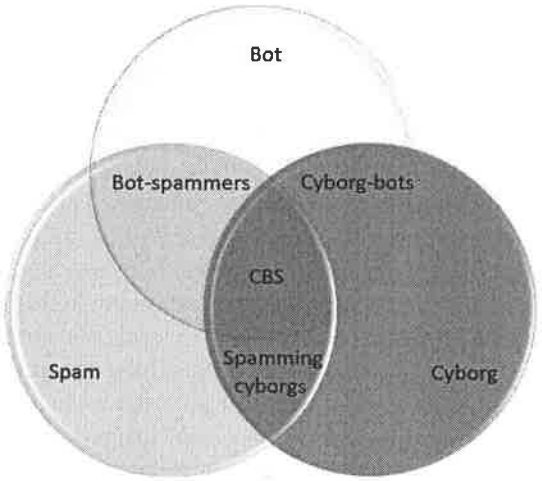

Figure 7.3. New hybridized automated programs for fake news dissemination.

network analysis with the aim of detecting and stopping bots. Independent groups and researchers have also developed mechanisms to detect bots.

Application-driven fake-news research focuses on open problems that go in other areas outside of detection of fake news. These two directions are *dissemination* and *action*. Dissemination of fake news characterizes the paths and patterns of fake news on social media. As discussed previously, true news and fake news follow different patterns when disseminated on social media. The dissemination of fake news on social media shows its own features that will need additional research, such as *life cycle* and *social aspect*. Social aspect refers to the fake-news diversity inside various social groups. Evidently, breaking news exhibits unusual life cycles in social networks. Studying the life cycle of fake news will provide a more comprehensive understanding of how some news goes viral from the normal sharing of information. Additionally, recognizing key propagators of fake news is critical in combating fake news on social media. By propagators, we identify humans, cyborgs, bots, and a variety of their permutations. The particular area of open problems is in building better and faster detectors of these fake profiles.

Next, we suggest further research into fake-news action. The goal of this action is to research ways to minimize the effects of fake news by building proactive action methods to combat the spread of fake news after fake news become viral. The process of immunization should be performed by exposing sources of news and thereby identifying less-reliable news. The isolation segment is focused on research about identifying and putting the spreaders of fake news into quarantine. For example, the goal would be to provide easily

accessible tools that allow users to validate a given post or news against reliable information sources. Feature-driven fake news aims to determine effective structures for detecting fake news from multiple data sources. Two proposed directions are grouping perspective and news content. From a grouping context perspective, we suggest individual-based and web-based features. Individual-based research features should focus on identifying different account types and extracting some user-specific features. This can be achieved by using neural networks and AI methods to detect users opinions and replies to fake news. Cluster-based features are to represent how different types of networks in social media are constructed. It is important, in this part of the research, to explore how clusters are constructed in terms of connections among users and posts. Furthermore, the other advanced methods of user clustering and user networking are needed as well. From a news-content perspective, we propose two subdirections: text-focused and image-focused techniques. Embedding methods, such as word embedding and deep neural networks, are attracting research interest in the way they extract content on social media. The graphical features extracted from images are also presented to be significant pointers for fake news. Yet, partial research has been done in the area of recognizing programs that can manipulate video and image footage; future studies should investigate peoples' long-term conviction in the veracity of fake news after preliminary exposure to it, and how our manipulations affect those opinions. Research should engage designs that provide more leverage for understanding the effects of warnings on belief in fake stories.

REFERENCES

Bolsen, T., and Druckman, J. N. 2015. 'Counteracting the Politicization of Science'. *Journal of Communication* 65(5): 745–69.

Buntain, C., and Golbeck, J. 2017. 'Automatically Identifying Fake News in Popular Twitter Threads'. In *2017 IEEE International Conference on Smart Cloud (SmartCloud)* (November): 208–15.

Clark, E. M., Williams, J. R., Jones, C. A., Galbraith, R. A., Danforth, C. M., and Dodds, P. S. 2016. 'Sifting Robotic from Organic Text: A Natural Language Approach for Detecting Automation on Twitter'. *Journal of Computational Science* 16:1–7.

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., and Nyhan, B. 2019. 'Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media'. *Political Behavior* 1:1–23.

Cybenko, A. K., and Cybenko, G. 2018. 'AI and Fake News'. *IEEE Intelligent Systems* 33 (5): 1–5.

DiFranzo, D., and Gloria, M. J. K. 2017. 'Filter Bubbles and Fake News'. *ACM Crossroads* 23 (3): 32–35.

Ecker, U. K., Lewandowsky, S., and Tang, D. T. 2010. 'Explicit Warnings Reduce but Do Not Eliminate the Continued Influence of Misinformation'. *Memory and Cognition* 38 (8): 1087–100.

Halimeh, A. A., Pourghomi, P. and Safieddine, F. 2017. 'The Impact of Facebook's News Fact-Checking on Information Quality (IQ) Shared on Social Media'. *MIT International Conference on Information Quality.*

Jin, F., Dougherty, E., Saraf, P., Cao, Y., and Ramakrishnan, N. 2013. 'Epidemiological Modeling of News and Rumors on Twitter'. In *Proceedings of the ACM 7th Workshop on Social Network Mining and Analysis* (p. 8).

Kokalitcheva, K. 2016. Mark Zuckerberg Says Fake News on Facebook Affecting the Election Is a 'Crazy Idea.' *Fortune*, 11 November. http://fortune.com/2016/11/11/facebook-election-fakenews-mark-zuckerberg/?iid=sr-link1.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Nyhan, B., Pennycook, G., Rothschild, D. and Schudson, M. 2018. 'The Science of Fake News'. *Science* 359 (6380): 1094–96.

Mosseri, A. 2016. *News Feed: Addressing Hoaxes and Fake News—Facebook News Room.* http://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/.

Pourghomi, P., Safieddine, F., Masri, W., and Dordevic, M. 2017. 'How to Stop Spread of Misinformation on Social Media: Facebook Plans vs. Right-Click Authenticate Approach'. *2017 IEEE International Conference on Engineering and MIS (ICEMIS)* (pp. 1–8).

Poynter Inc. 2016. *International Fact-Checking Network Fact-Checkers' Code of Principles*. http://www.poynter.org/fact-checkers-code-of-principles/.

Segarra, L. M. 2019. 'Facebook Removed 2.2 Billion Fake Accounts This Year. It Only Has 2.38 Billion Active Users'. *Fortune*, 23 May. http://fortune.com/2019/05/23/facebook-fake-accounts-transparency-report/.

Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H. 2017. 'Fake News Detection on Social Media: A Data Mining Perspective'. *ACM SIGKDD Explorations Newsletter* 19 (1): 22–36.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., and Flammini, A. 2017. 'Online human-bot Interactions: Detection, Estimation, and Characterization'. *Eleventh International AAAI Conference on Web and Social Media.*

Verstraete, M., Bambauer, D. E., and Yakowitz Bambauer, J. R. 2017. 'Identifying and Countering Fake News'. *Arizona Legal Studies Discussion Paper No. 17–15.*

8

Academic Research into Combating Fake News

Dr. Pardis Pourghomi

This chapter examines academic approaches to combating fake news. Researchers have looked at understanding the underlying incentives behind the creation and consumption of fake news, characteristics for its success, initiatives to educate audiences, filters for fake news, the use of algorithms, and the employment automated and semiautomated validation tools. Some of the research into these techniques is industry-led and as such, were covered in chapter 7. This chapter primarily focuses on academic research, highlighting the differences in academic approaches in comparison to that of the industry in the way researchers attempt to apply a level of pragmatism to gain an insight into human processes and how these can be paired with technological solutions to mediate behaviours around fake-news consumption. This chapter explores the distinct approaches while acknowledging that the challenges posed by fake news remain open to further introspection.

UNDERSTANDING SOCIAL-MEDIA ENVIRONMENT

One of the key contributions we see in academic publications is a drive to understand the architecture of the virtual environment social-media users encounter. The creation and consumption of fake news happen in a virtual world that resembles nothing humanity has witnessed before. It is a better understanding of this world that may enable us to tackle the challenges presented by fake news.

Fake-News Generator

What we know for a fact is that some news is fake and beyond that, users are left to fend for themselves to guess what is true. That is the true cost of fake news! Hal Berghel (2017) suggests that while the majority of people accept there is fake news and are maybe swayed where facts are presented, there are hardcore believers who simply cannot be swayed. Hence, the latter are termed as ideologists or *tribalists*. The tribalists strongly believe in the sentiment and the value of the fake news, which goes beyond the need to accept or reject the factuality of it. From a production perspective, Frankfurt (2009) delineates a clear distinction between liars and fake-news inventors, which the author colourfully labels as 'bullshitters' (BS). These BS speakers, unlike liars, have little or no concern if they are later held to account on their statements. With a liar, they would intentionally invent a story, fully aware it is not true. BS, pointedly, does not know the truth and does not seek it. The distinction we see here is that BS are inventors of fake news and tribalists are their diehard devotee consumers.

Berghel (2017) believes attempts at fact-checking to counter lies work for the general public, but countering BS inventors and their tribalist supporters through fact-checking can be futile as the intention behind the perpetrators of fake news is not only to depart from the truth, but to substitute it. In fact, one of the many characteristics of fake news is that 'it lacks the requisite semantics anchors to make it evaluable' (Berghel 2017, 41). For example, trying to fact check if Alexander the Great was buried in Alaska in 1985 is pointless because the claim is ludicrous and cannot be fact-checked. However, we can argue that such statements could be refuted through common sense and the common man on the street is able to discern these statements as nonsensical. According to researchers, fake-news generators use several methods when making a political or ideological proclamation that relies on performance rather than the reliability of the sources. Thus, statements tend to be focused on being assertive, directive, declarative, and authoritative. Such statements are evaluated not by their true value or facts but by the basis of *goodness-of-fit* with the ideology and serving of interest. An examples of such proclamations is: Mexico will pay for the wall. The suggestion from the literature is that fact-checking will only appeal to those who genuinely seek the truth and will have an impact on slowing the spread of fake news. However, fact-checking will have little to no effect on the hardcore supporters, that is, the tribalists. The suggestion here is that the best way to counter alt-truth is satire and ridicule. In an experiment involving staunch flat-earth supporters, Berghel noted that satirical memes had a better impact than facts in countering tribalists' responses. As it turns out, the use of memes and ridi-

cule is said to be the most efficient method to suppress the urge by tribalists to share what appears to be fake news.

The challenges at individual and small-group levels are further multiplied when big players take advantage of social-media open space. Haigh, Haigh, and Kozak (2018, 1) highlight the difficulties in challenging 'state-sponsored fake news campaign propagated over social media'. With the case study of a fact-checking a Ukrainian social-media group called StopFake, they illuminate the distinct approach to combating fake news in a so-called *Peer-to-Peer Propaganda* context. For example, PolitiFact ranks news as to how truthful it is. It does this by ranking articles after asking academic experts to research them and produce a ranking that can vary from *True*, *Mostly True*, *Mostly False*, *False*, and lastly *Pants on Fire* as a term to signify complete fabrication. StopFake does not follow this format. StopFake only publishes fake news, pointing out what is fake about it. In doing so, the approach of this website is to counter fake news in the hope that their evidence and posts can be used by other online users to respond to the given fake-news claims. The study of StopFake classification of fake news generated in Ukraine suggests the following categories of fakes: fake sources (39 per cent), fake images (35 per cent), misleading quotes (26 per cent), incorrect or fake individuals (15 per cent), government representative (14 per cent), fake expert opinion (11 per cent), photoshopped images (10 per cent), external debunked of the story (9 per cent), and government statistics (1 per cent) (Haigh et al. 2018, 17). In fact, 35 per cent of fake stories relied on finding the source of the fake images in that story. The problem with their approach is that where the claim is partly true, they would not take sides. In fact, a review of the literature presented in chapter 1 showed that some of the most challenging fake-news stories are those which are partially true. It is suggested that the reason for the difference in approaches between the American fact-checking and the Ukrainian counterpart is due to their varying agendas. With the American system, the aim is to keep politicians honest, whereas the Ukrainian system seeks to counter a systematic and coordinated state-sponsored propaganda machine. Moreover, the focus of the group is facts, not opinions. The *Peer-to-Peer Propaganda* approach discussed above is part of a covert operation where the Russian state, rather than relying on state-controlled media to sway public opinion, has resorted to troll farms and bots to disseminate misinformation. Troll farms, in this case, share views and posts that appear to be coming from ordinary people. The reality is that these seemingly ordinary people are in fact large numbers of either tribalist, paid, or robotic trolls who direct public opinion in certain ways. Beyond that point, fake news and views continue to propagate by individuals innocently and unaware of their role in this propaganda. Thus, this approach, dubbed *Pee-to-Peer Propaganda* (Haigh et al. 2018, 10), is a

much more powerful way of spreading propaganda where news is mixed with reactions and opinions of trusted friends. It is thus suggested that StopFake's approach is intended to address this kind of peer-to-peer dissemination of fake news, and Haigh et al. (2018) point out the challenges faced by fact-checkers. The human power to manage these operations can be overwhelming. At most, StopFake only posts once a day a post on a given subject. And at best they have a turnaround time of 1.6 days and at peak times, 5.4 days. Their averages during normal operations are 3.2 days.

Frequency of the Message

The frequency of the message is said to impact the users outside the tribalist sphere. In the case of fake news, the more the message is repeated, the more likely users are to believe the message. Likewise, when countering fake news, the more corrections users see, the less likely they are to believe that message. Polage (2012, 1) suggests that the key to making a story believable is in the repetition as 'information that is repeated is more likely to be rated as true than information that has not been heard before.' Thus, familiarity with a given news item could increase the rate of truthfulness and plausibility to such an extent that it would even become linked with false memories of having heard the story. In an interesting experiment, participants were presented with a fake story as a factual event, and after repeating that story five weeks later, the participants were more likely to believe the story as factual compared to those who heard it only the first time. More surprisingly, participants who had been introduced to the fake story a few weeks earlier are more likely to suggest they heard it outside the experiment. Thus, Polage (2012, 1) suggested that 'repeating false claims will not only increase their believability but may also result in source monitoring error', that is, false memories of the source of the news story. The authors suggest that this may well be one of the tactics used in elections to introduce ever so slightly key suggestions that otherwise, if presented outright, may not be believable. Exploring the other end of the spectrum, Porter, Wood, and Kirby (2018) contend that the frequency of exposure to corrective news can reduce believability in fake news. In an experiment involving exposing users to factual and fake news, as well as intervals of corrections to the fake news, users were then asked to rate the fake news. The more corrective posts the users were exposed to, the more likely they were to rate the fake news as fake. There were some variations based on the individuals' political and ideological disposition, but the results are consistent with an increase in identifying the fake news where initiatives were made to correct fake news.

Simulating Users' Behaviour

Safieddine, Masri, and Pourghomi (2016) argue that to prevent misinformation propagation effectively, it is essential to understand the processes of misinformation propagation in social media. By understanding the processes, researchers will be best placed to design operative techniques and strategies for controlling misinformation propagation in social media. This would constitute an important step in analysing and predicting the dynamic trend of misinformation propagation. Modelling and simulation of variables in such an ecosystem that describes the processes of misinformation propagation can provide an understanding of misinformation propagation and test the efficiency of a control strategy. Dordevic et al. (2016) set out to demonstrate proof-of-concept using 2D modelling and identified the variables involved in the travel of information (see figure 8.1). The source news in figure 8.1 sends fake news spreading, with some individuals believing and some not, some individuals sharing and some not.

The authors identified eight key variables and applied theoretical values to demonstrate their applicability. These variables are: *i* as the first vertex and *jn* is the last vertex of the given simulation. *V*1 represents the first phase of spread of misinformation and *l* represents the maximum possible reach of information through the network. Thus, it can be concluded that combating misinformation online is influenced by the following variables: rate the of authentication *A*, rate of sharing *S*, passing on information rate *P*, average cross-wire rate *Cw*, success the rate of Same Level communication rate *Sl*, and Reverse Validation rate *Rv*.

Critical to this simulation is the suggestion that if an authentication tool is developed, propagation of fake news can be halted under given scenarios. By

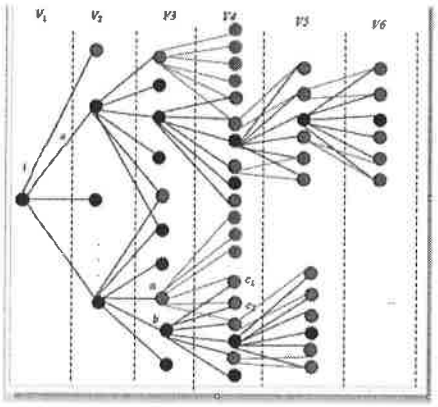

Figure 8.1. Passing on rate and cross-wire rate simulation.

Table 8.1. Critical Variables for Combating Information Online

Variables	*Notation*
Maximum population	(*I*)
Sharing	(*S*)
Passing on information	(*P*)
Authentication	(*A*)
Cross-wire	(*Cw*)
Same-level (cluster) communication	(*Sl*)
Reverse validation	(*Rv*)

simulating several scenarios, the authors proved near elimination of spread of misinformation online whereby the red nodes represented individuals who no longer believe the misinformation[1] (see figure 8.2). In order to curb the spread, there should be a critical mass of 30 per cent users either not believing the news or taking the time to authenticate it before sharing. This demonstrates that providing easily accessible tools would allow users to authenticate images and text and could effectively cascade the process back to the source or at least to the layer immediate to the source.

Several authors have argued that the spread of fake news resembles the spread of diseases in many ways. In a follow-up study, Pourghomi, Dordevic, and Safieddine (2018) replicated the study using big-data simulation software BioLayout Express, software that has been used to simulate the spread of diseases. Using the same variables as conducted using 2D simulations, the results

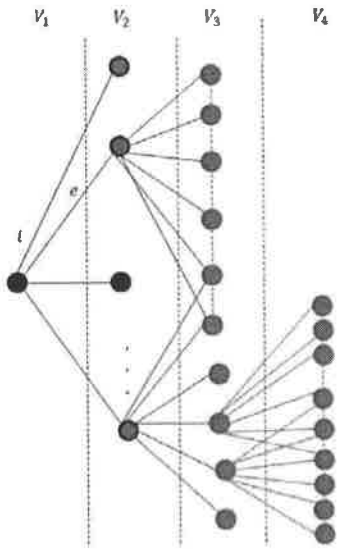

Figure 8.2. The outcome of the two-dimensional simulation.

Figure 8.3. Simulation with authenication at 30 percent shows stop in fake news propagation.

showed again that having 30 per cent of the population authenticate information is usually sufficient to stop fake news from propagating (see figure 8.3).

Finally, the simulation was able to identify the impact of influences, newsgroups, and automated trolls. Safieddine, Dordevic, and Pourghomi (2017) identified the challenges in authentication when fake news is precipitated by an influencer, newsgroup, or roll army with many followers. With all other variables fixed, when fake news is precipitated by an individual, 30 per cent of users are needed to authenticate and reject the news to stop it propagating. But when a newsgroup or individual with mass followers are concerned, the rate goes up to 54 per cent—thus making the process of stopping the spread of that fake post much more challenging. Part of the analysis also revealed that the initial cluster of users around the group facilitates the explosion of fake news on the social network, making it much more difficult to control, as shown in figure 8.4.

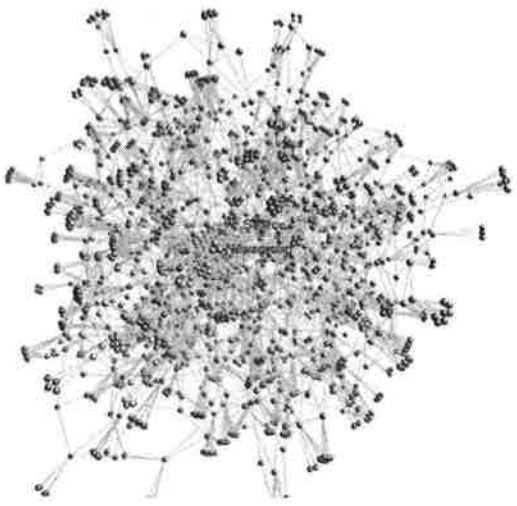

Figure 8.4. Impact of influencers with fake news with 30 percent authentication rate.

Another byproduct of that experiment is that the simulation shows how much power the source has in terminating the spread of fake news. Should the source delete or retract the post, the spread is halted with only 9 per cent of the population required to check its authenticity. Thus, should a mainstream media inadvertently share fake news and then retract it, the impact is far less than that of a source that creates fake news and refuses to retract it. This also shows the power social-media platforms can exert should they be active in deleting fake news.

USING ARTIFICIAL INTELLIGENCE AND MACHINE LANGUAGE

There has been a significant body of work done to promote and suggest artificial intelligence (AI) and machine language (ML) hold the answers to countering and halting the spread of fake news. There are some promising leads, but the suggestion that the technology is anywhere near to being ready to deal with the challenges remains to be seen. The following are some of the important leads identified from the literature. The way news is written tends to follow linguistic cues where a set of words grouped semantically refers to a specific subject. This is known as the lexical field of research. Choy and Chong (2018) argue that fake news has a lexical structure that is significantly different than factual news. Their research suggests that biased information is found to contain 'linguistic cues, active verbs, implications, hedges, and subjective intensifiers' (2018, 3). In fact, misleading headline news, also known as 'clickbait', is said to have a unique lexical style that uses vague pronouns, affective language, action words, dramatic language, and overuse of numbers (Chen, Conroy, and Rubin 2015). The other suggestion is that the creators of fake news tend to create fewer complex stories. The work of Richards and Gross (2000) suggests that lying requires additional cognitive resources, and as such, creators of fake news tend to create simple stories that have shorter sentences, more motion verbs, and less evaluation and judgement words, such as 'walking' and 'talking' as opposed to 'thinking' and 'believing'. Finally, on the emotional front, Chen, Conroy, and Rubin (2015, 4) point to several studies that show highly emotional stories that provoke feelings of anger and anxiety tend to attract readers and increase exposure. The authors then propose that any modelling system that aims to detect fake news needs to take into consideration these factors in what they termed as *LeSiE* framework, a term combining 'lexical', 'simplicity', and 'emotions'. In a lab test involving testing 11,523 headlines from the online fact-checking news agency *Politifact*, the researchers attempted to test how reliable the