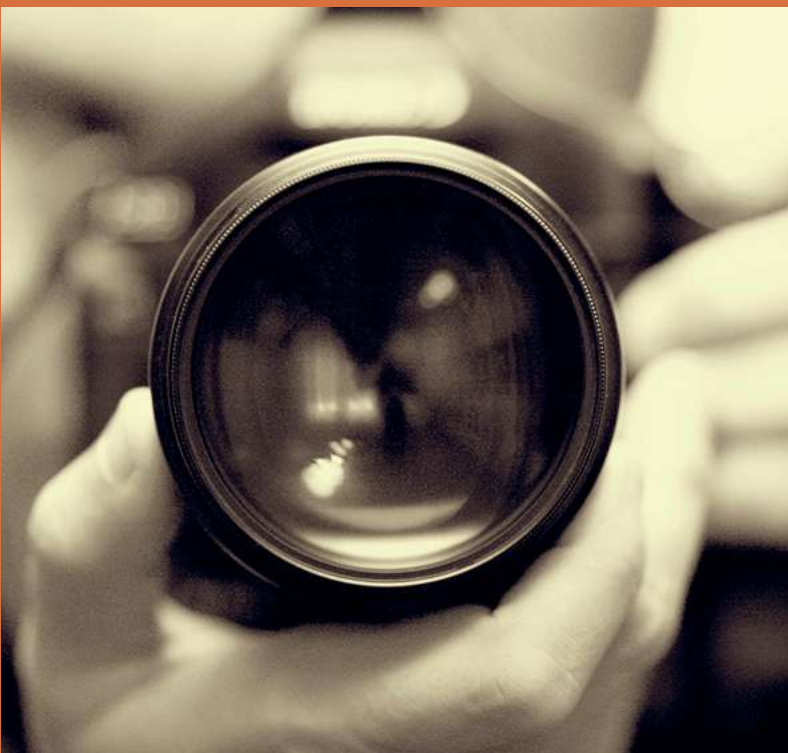


INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making



Council of Europe report
DGI(2017)09

Claire Wardle, PhD
Hossein Derakhshan

Information Disorder

Toward an interdisciplinary framework for research and policymaking

By Claire Wardle, PhD and Hossein Derakhshan

With research support from Anne Burns and Nic Dias

September 27, 2017

The opinions expressed in this work are the responsibility of the authors and do not necessarily reflect the official policy of the Council of Europe.

All rights reserved.

No part of this publication may be translated, reproduced or transmitted in any form or by any means without the prior permission in writing from the Directorate of Communications (F-67075 Strasbourg Cedex or publishing@coe.int).

Photos © Council of Europe

Published by the Council of Europe

F-67075 Strasbourg Cedex

www.coe.int

© Council of Europe, August 2018, 2nd revised edition

Table of content

Authors' Biographies	3
Executive Summary	4
Introduction	10
Part 1: Conceptual Framework	20
The Three Types of Information Disorder	20
The <i>Phases</i> and <i>Elements</i> of Information Disorder	22
The Three Phases of Information Disorder	24
The Three Elements of Information Disorder	25
1) The Agents: Who are they and what motivates them?	30
2) The Messages: What format do they take?	39
3) Interpreters: How do they make sense of the messages?	42
Part 2: Challenges of filter bubbles and echo chambers	49
Part 3: Attempts at solutions	59
Part 4: Future trends	77
Part 5: Conclusions	79
Part 6: Recommendations	82
Appendix: European Fact-checking and Debunking Initiatives	88
References	93

Authors' Biographies

Claire Wardle, PhD

Claire Wardle is the Executive Director of First Draft, which is dedicated to finding solutions to the challenges associated with trust and truth in the digital age. Claire is also a Research Fellow at the Shorenstein Center on Media, Politics and Public Policy at the Harvard Kennedy School where First Draft is now a project. She sits on the World Economic Forum's Global Agenda Council on the Future of Information and Entertainment. She was previously the Research Director at the Tow Center for Digital Journalism at Columbia Journalism School, head of social media for the UN Refugee Agency and Director of News Services for Storyful. She is one of the world's experts on user generated content, and has led two substantial research projects investigating how it is handled by news organizations. In 2009 she was asked by the BBC to develop a comprehensive social media training curriculum for the organization and she led a team of 19 staff delivering training around the world. She holds a PhD in Communication and an MA in Political Science from the University of Pennsylvania, where she won a prestigious Thouron Scholarship. She started her career as a professor at Cardiff University's School of Journalism, Media and Cultural Studies.

Hossein Derakhshan

Hossein Derakhshan is an Iranian-Canadian writer and researcher. The pioneer of blogging in Iran in the early 2000s, he later spent six years in prison there over his writings and web activities. He is the author of *'The Web We Have to Save'* (Matter, July 2015), which was widely translated and published around the world. His current research is focused on the theory and socio-political implications of digital and social media. His writings have appeared in *Libération*, *Die Zeit*, the *New York Times*, *MIT Technology Review*, and *The Guardian*. He studied Sociology in Tehran and Media and Communication in London.

Executive Summary

This report is an attempt to comprehensively examine information disorder and its related challenges, such as filter bubbles and echo chambers. While the historical impact of rumours and fabricated content have been well documented, we argue that contemporary social technology means that we are witnessing something new: information pollution at a global scale; a complex web of motivations for creating, disseminating and consuming these ‘polluted’ messages; a myriad of content types and techniques for amplifying content; innumerable platforms hosting and reproducing this content; and breakneck speeds of communication between trusted peers.

The direct and indirect impacts of information pollution are difficult to quantify. We’re only at the earliest of stages of understanding their implications. Since the results of the ‘Brexit’ vote in the UK, Donald Trump’s victory in the US and Kenya’s recent decision to nullify its national election result, there has been much discussion of how information disorder is influencing democracies. More concerning, however, are the long-term implications of dis-information campaigns designed specifically to sow mistrust and confusion and to sharpen existing socio-cultural divisions using nationalistic, ethnic, racial and religious tensions.

So, how do we begin to address information pollution? To effectively tackle the problems of mis-, dis- and mal- information, we need to work together on the following fronts:

1. **Definitions.** Think more critically about the language we use so we can effectively capture the complexity of the phenomenon;
2. **Implications for democracy.** Properly investigate the implications for democracy when false or misleading information circulates online;
3. **Role of television.** Illuminate the power of the mainstream media, and in particular television, in the dissemination and amplification of poor-quality information that originates online;
4. **Implications of weakened local media.** Understand how the collapse of local journalism has enabled mis- and dis-information to take hold, and find ways to support local journalism;
5. **Micro-targeting.** Discern the scale and impact of campaigns that use demographic profiles and online behavior to micro-target fake or misleading information¹;
6. **Computational amplification.** Investigate the extent to which influence is bought through digital ‘astroturfing’—the use of bots and cyborgs to manipulate the outcome of online petitions, change search engine results and boost certain messages on social media;

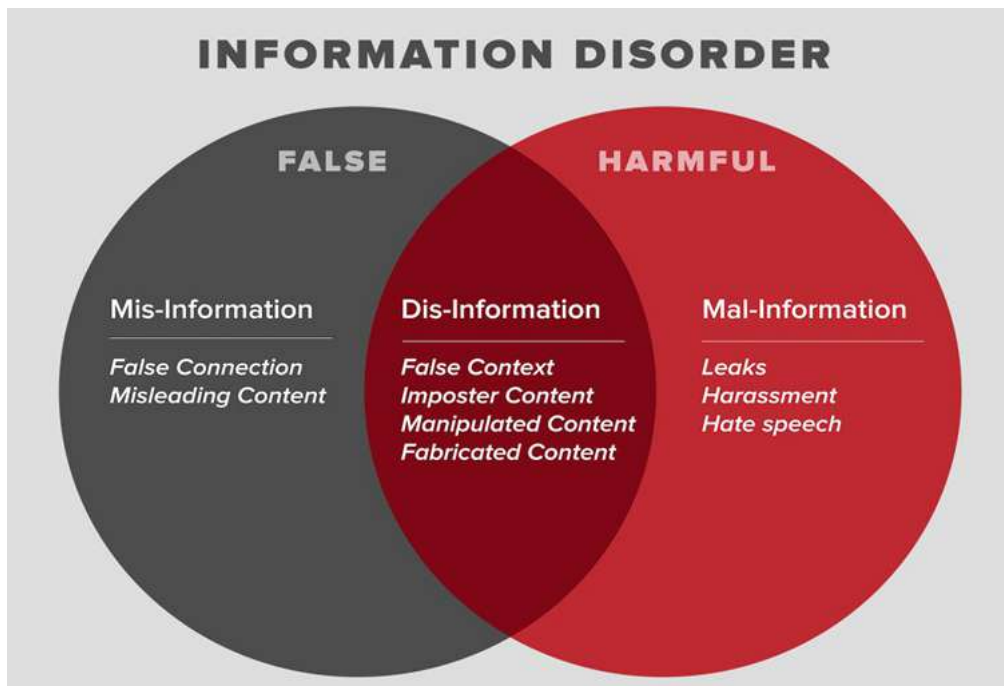
¹ Hendrix, J. and Carroll, D. (2017) *Confronting a Nightmare for Democracy*. Available at: <https://medium.com/@profcarroll/confronting-a-nightmare-for-democracy-5333181ca675>

7. **Filter bubbles and echo chambers.** Consider the implications of the filter bubbles and echo chambers that have emerged because of media fragmentation, both offline (mediated via partisan talk radio and cable news) and online (mediated via hyper-partisan websites, algorithmically derived feeds on social networks and radical communities on WhatsApp, Reddit and 4chan.)²
8. **Declining trust in evidence.** Understand the implications of different communities failing to share a sense of reality based on facts and expertise.

In this report, we refrain from using the term ‘fake news’, for two reasons. First, it is woefully inadequate to describe the complex phenomena of information pollution. The term has also begun to be appropriated by politicians around the world to describe news organisations whose coverage they find disagreeable. In this way, it’s becoming a mechanism by which the powerful can clamp down upon, restrict, undermine and circumvent the free press.

We therefore introduce a new conceptual framework for examining information disorder, identifying the three different types: mis-, dis- and mal-information. Using the dimensions of harm and falseness, we describe the differences between these three types of information:

- Mis-information is when false information is shared, but no harm is meant.
- Dis-information is when false information is knowingly shared to cause harm.
- Mal-information is when genuine information is shared to cause harm, often by moving information designed to stay private into the public sphere.



² Beran, Dale. “4chan: *The Skeleton Key to the Rise of Trump.*” Medium.com, February 14, 2017. <https://medium.com/@DaleBeran/4chan-the-skeleton-key-to-the-rise-of-trump-624e7cb798cb>

We also argue that we need to separately examine the ‘elements’ (the agent, messages and interpreters) of information disorder. In this matrix we pose questions that need to be asked of each element.

Agent	Actor Type: Level of Organisation: Type of Motivation: Level of Automation: Intended Audience: Intent to Harm: Intent to Mislead:	Official / Unofficial None / Loose / Tight / Networked Financial / Political / Social / Psychological Human / Cyborg / Bot Members / Social Groups / Entire Societies Yes / No Yes / No
Message	Duration: Accuracy: Legality: Imposter Type: Message Target:	Long term / Short-term / Event-based Misleading/ Manipulated / Fabricated Legal / Illegal No / Brand / Individual Individual / Organisation / Social Group / Entire Society
Interpreter	Message reading: Action taken:	Hegemonic / Oppositional / Negotiated Ignored / Shared in support / Shared in opposition

We also emphasise the need to consider the three different ‘phases’ (creation, production, distribution) of information disorder.



As we explain, the ‘agent’ who creates a fabricated message might be different to the agent who produces that message—who might also be different from the ‘agent’ who distributes the message. Similarly, we need a thorough understanding of who these agents are and what motivates them.

We must also understand the different types of messages being distributed by agents, so that we can start estimating the scale of each and addressing them. (The debate to date has been overwhelmingly focused on fabricated text news sites, when visual content is just as widespread and much harder to identify and debunk.)

Finally, we need to examine how mis-, dis- and mal-information are being consumed, interpreted and acted upon. Are they being re-shared as the original agent intended? Or are they being re-shared with an oppositional message attached? Are these rumours continuing to travel online, or do they move offline into personal conversations, which are difficult to capture?

A key argument within this report, which draws from the work of the scholar James Carey, is that we need to understand the ritualistic function of communication. Rather than simply thinking about communication as the transmission of information from one person to another, we must recognize that communication plays a fundamental role in representing shared beliefs. It is not just information, but drama — “a portrayal of the contending forces in the world.”³

The most ‘successful’ of problematic content is that which plays on people’s emotions, encouraging feelings of superiority, anger or fear. That’s because these factors drive re-sharing among people who want to connect with their online communities and ‘tribes’. When most social platforms are engineered for people to publicly ‘perform’ through likes, comments or shares, it’s easy to understand why emotional content travels so quickly and widely, even as we see an explosion in fact-checking and debunking organizations.

In addition to our conceptual framework, we provide a round-up of related research, reports and practical initiatives connected to the topic of information disorder, as well as filter bubbles and echo chambers. We examine solutions that have been rolled out by the social networks and consider ideas for strengthening existing media, news literacy projects and regulation. We also introduce some key future trends, particularly in terms of the rise of closed messaging apps and the implications of artificial intelligence technology for manufacturing as well as detecting dis-information.

³ Carey, J. (1989), *Communication as Culture: Essays on Media and Society*, London: Routledge. p.16

The report ends with an explanation of thirty-four recommendations, targeted at technology companies, national governments, media organisations, civil society, education ministries and funding bodies. They are explained in detail after the report's conclusions.

What could *technology companies* do?

1. Create an international advisory council.
2. Provide researchers with the data related to initiatives aimed at improving public discourse.
3. Provide transparent criteria for any algorithmic changes that down-rank content.
4. Work collaboratively.
5. Highlight contextual details and build visual indicators.
6. Eliminate financial incentives.
7. Crack down on computational amplification.
8. Adequately moderate non-English content.
9. Pay attention to audio/visual forms of mis- and dis-information.
10. Provide metadata to trusted partners.
11. Build fact-checking and verification tools.
12. Build 'authenticity engines'.
13. Work on solutions specifically aimed at minimising the impact of filter bubbles:
 - a. Let users customize feed and search algorithms.
 - b. Diversify exposure to different people and views.
 - c. Allow users to consume information privately.
 - d. Change the terminology used by the social networks.

What could *national governments* do?

1. Commission research to map information disorder.
2. Regulate ad networks.
3. Require transparency around Facebook ads.
4. Support public service media organisations and local news outlets.
5. Roll out advanced cybersecurity training.
6. Enforce minimum levels of public service news on to the platforms.

What could *media organisations* do?

1. Collaborate
2. Agree policies on strategic silence.
3. Ensure strong ethical standards across all media.
4. Debunk sources as well as content.
5. Produce more news literacy segments and features.
6. Tell stories about the scale and threat posed by information disorder.
7. Focus on improving the quality of headlines.
8. Don't disseminate fabricated content.

What could ***civil society*** do?

1. Educate the public about the threat of information disorder.
2. Act as honest brokers.

What could ***education ministries*** do?

1. Work internationally to create a standardized news literacy curriculum.
2. Work with libraries.
3. Update journalism school curricula.

What could ***funding bodies*** do?

1. Provide support for testing solutions.
2. Support technological solutions.
3. Support programs teaching people critical research and information skills.

Introduction

Rumours, conspiracy theories and fabricated information are far from new.⁴ Politicians have forever made unrealistic promises during election campaigns. Corporations have always nudged people away from thinking about issues in particular ways. And the media has long disseminated misleading stories for their shock value. However, the complexity and scale of information pollution in our digitally-connected world presents an unprecedented challenge. While it is easy to dismiss the sudden focus on this issue because of the long and varied history of mis- and dis-information⁵, we argue that there is an immediate need to seek workable solutions for the polluted information streams that are now characteristic of our modern, networked and increasingly polarised world.

It is also important to underline from the outset that, while much of the contemporary furor about mis-information has focused on its political varieties, ‘information pollution’⁶ contaminates public discourse on a range of issues. For example, medical mis-information has always posed a worldwide threat to health, and research has demonstrated how incorrect treatment advice is perpetuated through spoken rumours⁷, tweets⁸, Google results⁹ and Pinterest boards¹⁰. Furthermore, in the realm of climate change, a recent study examined the impact of exposure to climate-related conspiracy theories. It found that exposure to such theories created a sense of powerlessness, resulting in disengagement from politics and a reduced likelihood of people to make small changes that would reduce their carbon footprint.¹¹

In this report, we hope to provide a framework for policy-makers, legislators, researchers, technologists and practitioners working on challenges related to mis-, dis- and mal-information—which together we call information disorder.

⁴ Sunstein, Cass R., and Adrian Vermeule. “Conspiracy Theories: Causes and Cures”. *Journal of Political Philosophy* 17, no. 2 (2009): 202–227.

⁵ Uberti, D. (2016) *The Real History of Fake News*, *Columbia Journalism Review*, December 15, 2016. https://www.cjr.org/special_report/fake_news_history.php

⁶ The term information pollution was first used by Jakob Nielsen in 2003 as a way to describe irrelevant, redundant, unsolicited and low-value information.

⁷ Smith, L. C., Lucas, K. J., & Latkin, C. (1999). *Rumor and gossip: Social discourse on HIV and AIDS*. *Anthropology & Medicine*, 6(1), 121-131.

⁸ Oyeyemi, S. et al., (14 Oct., 2014), *Ebola, Twitter, and mis-information: a dangerous combination*, *British Medical Journal*, 349

⁹ Venkatraman A. et al., (2016) *Zika virus mis-information on the internet*, *Travel Medicine and Infectious Disease*, Vol 14: 4, pp 421-422

¹⁰ Guidry, J. et al., (2015) *On pins and needles: How vaccines are portrayed on Pinterest*, *Vaccines*, Vol 33 (39), pp.5051-5056

¹¹ Jolley, D. and K. Douglas, (2014) *The Effects of Anti-Vaccine Conspiracy Theories on Vaccination Intentions*, *PLOS ONE* 9(2)

But first, how did we get to this point? Certainly, the 2016 US Presidential election led to an immediate search for answers from those who had not considered the possibility of a Trump victory—namely the major news outlets, pundits and pollsters. And while the US election result was caused by an incredibly complex set of factors – socio-economic, cultural, political and technological – there was a desire for simple explanations, and the idea that fabricated news sites could provide those explanations drove a frenzied period of reporting, conferences and workshops.¹²

Reporting by BuzzFeed News’ Craig Silverman provided an empirical framework for these discussions, offering evidence that the most popular of these fabricated stories were shared more widely than the most popular stories from the mainstream media: “In the final three months of the US presidential campaign, 20 top-performing false election stories from hoax sites and hyper-partisan blogs generated 8,711,000 shares, reactions, and comments on Facebook. Within the same time period, the 20 best-performing election stories from 19 major news websites generated a total of 7,367,000 shares, reactions, and comments on Facebook.”¹³

In addition, research on referral data shows that “fake news” stories relied heavily on social media for traffic during the election¹⁴. Only 10.1% of traffic to the top news sites came from social media, compared with 41.8% for ‘fake news sites’. (Other traffic referral types were direct browsing, other links and search engines.)

While we know that mis-information is not new, the emergence of the internet and social technology have brought about fundamental changes to the way information is produced, communicated and distributed. Other characteristics of the modern information environment include:

- a) Widely accessible, cheap and sophisticated editing and publishing technology has made it easier than ever for anyone to create and distribute content;

¹² See write-ups of some of the most important events here: Shorenstein Center’s ‘Fake News Agenda for Research and Action: (<https://shorensteincenter.org/combating-fake-news-agenda-for-research/>); and Yale University’s Information Society Project’s Fighting Fake News Workshop https://law.yale.edu/system/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf and the Westminster Media Forum Keynote Seminar: Fake news - scope, public trust and options for policy, <http://blogs.lse.ac.uk/mediapolicyproject/2017/08/10/the-evolving-conversation-around-fake-news-and-potential-solutions/>

¹³ Silverman, C. (2016b) *This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook*, BuzzFeed News, November 16, 2016. <https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>

¹⁴ Allcott, H., & Gentzkow, M. (2017). *Social Media and Fake News in the 2016 Election*. National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w23089>

- b) Information consumption, which was once private, has become public because of social media;
- c) The speed at which information is disseminated has been supercharged by an accelerated news cycle and mobile handsets;
- d) Information is passed in real-time between trusted peers, and any piece of information is far less likely to be challenged.

As Frederic Filloux explained: “What we see unfolding right before our eyes is nothing less than Moore’s Law applied to the distribution of mis-information: an exponential growth of available technology coupled with a rapid collapse of costs.”¹⁵

A study conducted in eighteen countries by the BBC World Service in September 2017 found that 79% of respondents said they worried about what was fake and what was real on the internet.¹⁶ Brazilians were most troubled, with 92% of respondents from that country expressing some concern about the issue. The least concerned were Germans, where 51% of respondents indicated that they were worried. Unfortunately, we don’t have similar data from previous years to understand whether concern has increased in light of recent discussions about the phenomenon. But one thing to bear in mind is that when the purpose of dis-information campaigns is to sow mistrust and confusion about what sources of information are authentic, it is important that we continue to track attitudes about the information people source from the internet.

Another critical point is that popular social networks make it difficult for people to judge the credibility of any message, because posts from publications as unlike as the *New York Times* and a conspiracy site look nearly identical. This means that people are increasingly reliant on friends and family members to guide them through the information ecosystem. As Messing and Westwood have argued, “social media has had two effects: by collating stories from multiple sources, the focus is on the story, and not on the source; secondly, endorsements and social recommendations guide readership”¹⁷ rather than traditional gatekeepers or ingrained reading habits.

¹⁵ Filloux, F. (2017) *You can’t sell news for what it costs to make*, The Walkley Magazine on Medium, <https://medium.com/the-walkley-magazine/you-cant-sell-news-for-what-it-costs-to-make-7a4def964ffa>

¹⁶ Cellan-Jones, R. (Sept 22, 2017) *Fake news worries 'are growing' suggests BBC poll*, BBC News, <http://www.bbc.com/news/technology-41319683>

¹⁷ Messing, S., & Westwood, S. J. (2014). *Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online*. *Communication Research*, 41(8), 1042-1063.

Daily, we spend twice as much time online compared with 2008. During that protracted amount time, we consume incredible amounts of information¹⁸ and inevitably make mistakes. Recent research by Filippo Menczer and colleagues shows we are so utterly inundated that we share untruths. Parsing information and judging the credibility of sources on Facebook or other social platforms will require our brains to adapt with new cognitive strategies for processing information. But Facebook is only 13 years old.¹⁹

Social networks are driven by the sharing of emotional content. The architecture of these sites is designed such that every time a user posts content—and it is liked, commented upon or shared further—their brain releases a tiny hit of dopamine. As social beings, we intuit the types of posts that will conform best to the prevailing attitudes of our social circle.²⁰ And so, on this issue of information disorder, this performative aspect of how people use social networks is critical to understanding how mis- and dis-information spreads.

However, we must also recognize the role of television in spreading dis-information.²¹ While much has been written about the growing influence of *Sputnik* and *Russia Today*²², as well as its new youth channel, *In the Now*, the unintentional amplification of dis-information by the mainstream media across the world needs to be acknowledged. From the *New York Times*' inaccurate reporting on Iraq's weapons of mass destruction, to the wall-to-wall coverage of Hillary Clinton's leaked emails (now known to be carried out by Russian hackers), or the almost daily amplification of Trump's tweets (some including information from conspiracy sites²³), getting the mainstream media to amplify rumour and dis-information is the ultimate goal of those who seek to manipulate. Without amplification, dis-information goes nowhere.

It is within this context that we have to study information disorder. These technology platforms are not neutral communication pipelines. They cannot be, as they are inherently social, driven by billions of humans sharing words, images, videos and memes that affirm their positions in their own real-life social networks.

¹⁸ Meeker, M (2017) *Internet Trends, 2017*. P.9. Available at:
<http://dq756f9pzlyr3.cloudfront.net/file/Internet+Trends+2017+Report.pdf>

¹⁹ Qiu, X. et al. (2017) *Limited individual attention and online virality of low-quality information*, *Nature Human Behaviour*, Vol 1

²⁰ Derakhshan blames web 2.0 for starting this 'tyranny of the novel and the popular' which then moved on to social platforms. See Derakhshan H. (July 14, 2015) *The Web We Have to Save*. Matter.
<https://medium.com/matter/the-web-we-have-to-save-2eb1fe15a426>

²¹ Derakhshan H. (Nov 29, 2016) *Social Media Is Killing Discourse Because It's Too Much Like TV*, *MIT Technology Review*. <https://www.technologyreview.com/s/602981/social-media-is-killing-discourse-because-its-too-much-like-tv/>

²² Rutenberg, J. (Sept. 13, 2017) *RT, Sputnik and Russia's New Theory of War*, *New York Times*,
<https://www.nytimes.com/2017/09/13/magazine/rt-sputnik-and-russias-new-theory-of-war.html?>

²³ Benkler Y. et al (March 3, 2017) *Study: Breitbart-led right-wing media ecosystem altered broader media agenda*. *Columbia Journalism Review*. <https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>

The shock of the Brexit referendum, the US election, Le Pen reaching the run-off vote in the French election and the overturning of the Kenyan election have been used as examples of the potential power of systematic dis-information campaigns. However, empirical data about the exact influence of such campaigns does not exist.

As danah boyd argues about recent responses to fears about mis- and dis-information, “It’s part of a long and complicated history, and it sheds light on a variety of social, economic, cultural, technological, and political dynamics that will not be addressed through simplistic solutions.”²⁴ Certainly, we have to look for explanations for how societies, particularly in the West, have become so segregated in terms of terms of age, race, religion, class and politics.²⁵ Recognizing the impact of factors such as the collapse of the welfare state, the failure of democratic institutions to provide public services, climate change and miscalculated foreign interventions are required. We cannot see the phenomenon of mis- and dis-information in isolation, but must consider its impact amid the new-media ecosystem. This ecosystem is dominated by increasingly partisan radio, television and social media; exaggerated emotional articulations of the world; quick delivery via algorithmically derived feeds on smartphones and audiences that skim headlines to cope with the floods of information before them. Making sense of mis-, dis- and mal-information as a type of information disorder, and learning how it works, is a necessity for open democracies. Likewise, neglecting to understand the structural reasons for its effectiveness is a grave mistake.

Communication as Ritual

One of the most important communication theorists, James Carey, compared two ways of viewing communication - transmission and ritual - in his book *Communication as Culture: Essays on Media and Society*.²⁶

Carey wrote, “The transmission view of communication is the commonest in our culture—perhaps in all industrial cultures... It is defined by terms such as ‘imparting,’ ‘sending,’ ‘transmitting,’ or ‘giving information to others.’”²⁷ The ‘ritual view of communication’, by contrast, is not about “the act of imparting information but the representation of shared beliefs.”

²⁴ boyd, d. (March 27,2017) “Google and Facebook can’t just make fake news disappear”, Backchannel, <https://medium.com/backchannel/google-and-facebook-cant-just-make-fake-news-disappear-48f4b4e5fbe8>

²⁵ Iyengar S. and S. J. Westwood (2015) *Fear and Loathing across Party Lines: New Evidence on Group Polarization*, American Journal of Political Science, Vol. 59, No. 3 (July 2015), pp. 690-707

²⁶ Carey, J. (1989) *Communication as Culture: Essays on Media and Society*, London: Routledge.

²⁷ Carey, (1989), p.18

Under a transmission view of communication, one sees the newspaper as an instrument for disseminating knowledge. Questions arise as to its effects on audiences—as enlightening or obscuring reality, as changing or hardening attitudes or as breeding credibility or doubt. However, a ritual view of communication does not consider the act of reading a newspaper to be driven by the need for new information. Rather, it likens it to attending a church service. It's a performance in which nothing is learned, but a particular view of the world is portrayed and confirmed. In this way, news reading and writing is a ritualistic and dramatic act.²⁸

In this report, we pay close attention to social and psychological theories that help to make sense of why certain types of dis-information are widely consumed and shared. Considering information consumption and dissemination from merely the transmission view is unhelpful as we try and understand information disorder.

Four Key Points

The term 'fake news' and the need for definitional rigour

Before we continue, a note on terminology. One depressing aspect of the past few months is that, while it has resulted in an astonishing number of reports, books, conferences and events, it has produced little other than funding opportunities for research and the development of tools. One key reason for this stagnation, we argue, is an absence of definitional rigour, which has resulted in a failure to recognize the diversity of mis- and dis-information, whether of form, motivation or dissemination.

As researchers like Claire Wardle²⁹, Ethan Zuckerman³⁰, danah boyd³¹ and Caroline Jack³² and journalists like the Washington Post's Margaret Sullivan³³ have argued, the term 'fake news' is woefully inadequate to describe the complex phenomena of mis- and dis-information. As Zuckerman states, "It's a vague and ambiguous term that spans everything from false balance (actual news that doesn't deserve our attention), propaganda (weaponized speech designed

²⁸ Carey (1989), pp.20-21

²⁹ Wardle, C. (Feb 16, 2017) *Fake News. It's Complicated*, First Draft, <https://firstdraftnews.com/fake-news-complicated/>

³⁰ Zuckerman, E. (Jan 30, 2017) *Stop Saying Fake News, It's not Helping*, My Heart is in Accra, <http://www.ethanzuckerman.com/blog/2017/01/30/stop-saying-fake-news-its-not-helping/>

³¹ boyd, d. (March 27, 2017) *Google and Facebook can't just make Fake News Disappear*, Wired, <https://www.wired.com/2017/03/google-and-facebook-cant-just-make-fake-news-disappear/>

³² Jack, C. (2017) *Lexicon of Lies*, *Data & Society*, https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf

³³ Sullivan, M (Jan 6, 2017,) *It's Time To Retire the Tainted Term Fake News*, Washington Post, https://www.washingtonpost.com/lifestyle/style/its-time-to-retire-the-tainted-term-fake-news/2017/01/06/a5a7516c-d375-11e6-945a-76f69a399dd5_story.html

to support one party over another) and *disinformatzya* (information designed to sow doubt and increase mistrust in institutions).”³⁴

A study by Tandoc *et al.*, published in August 2017, examined 34 academic articles that used the term ‘fake news’ between 2003 and 2017.³⁵ The authors noted that the term has been used to describe a number of different phenomena over the past 15 years: news satire, news parody, fabrication, manipulation, advertising and propaganda. Indeed, this term has a long history, long predating President Trump’s recent obsession with the phrase.

The term “fake news” has also begun to be appropriated by politicians around the world to describe news organisations whose coverage they find disagreeable. In this way, it’s becoming a mechanism by which the powerful can clamp down upon, restrict, undermine and circumvent the free press. It’s also worth noting that the term and its visual derivatives (e.g., the red ‘FAKE’ stamp) have been even more widely appropriated by websites, organisations and political figures identified as untrustworthy by fact-checkers to undermine opposing reporting and news organizations.³⁶ We therefore do not use the term in this report and argue that the term should not be used to describe this phenomenon.

Many have offered new definitional frameworks in attempts to better reflect the complexities of mis- and dis-information. Facebook defined a few helpful terms in their paper on information operations:

1. **Information (or Influence) Operations.** Actions taken by governments or organized non-state actors to distort domestic or foreign political sentiment, most frequently to achieve a strategic and/or geopolitical outcome. These operations can use a combination of methods, such as false news, dis-information or networks of fake accounts aimed at manipulating public opinion (false amplifiers).
2. **False News.** News articles that purport to be factual, but contain intentional misstatements of fact to arouse passions, attract viewership or deceive.
3. **False Amplifiers.** Coordinated activity by inauthentic accounts that has the intent of manipulating political discussion (e.g., by discouraging specific parties from participating in discussion or amplifying sensationalistic voices over others).

In ‘Fake News. It’s Complicated’, Wardle outlines seven types of mis- and dis-information, revealing the wide spectrum of problematic content online, from satire and parody (which,

³⁴ Zuckerman, (2017).

³⁵ Tandoc, Jr., E. C, Lim, Z. W., and Ling, R. (Aug. 2017) *Defining ‘Fake News’: A Typology of Scholarly Definitions*, *Digital Journalism*, 5 (7): 1-17

³⁶ Haigh et al, (2017) *Stopping Fake News: The work practices of peer-to-peer counter propaganda*. *Journalism Studies*, 1-26.

while a form of art, can become mis-information when audiences misinterpret the message) to full-blown fabricated content.

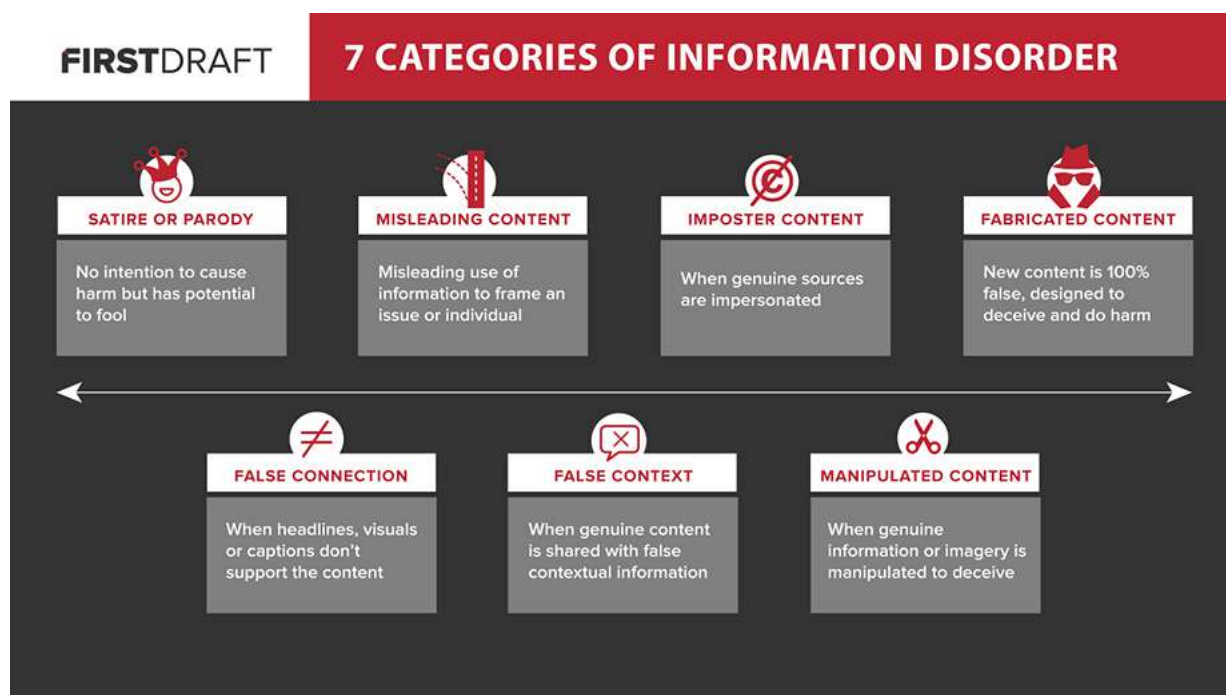


Figure 1: 7 Types of Mis- and Dis-information (Credit: Claire Wardle, First Draft)

While these seven classifications are helpful in encouraging people to see beyond the infamous ‘Pope endorses Trump’-type news sites that received so much attention after the US election, the phenomenon requires an even more nuanced conceptual framework—particularly one that highlights the impact of visuals in perpetuating dis-information. We have therefore created such a framework, and we will use as the organizing structure for the report.

While we work through terms and descriptions, it’s important that we recognise the importance of shared definitions. As Caroline Jack argued in the introduction to her recent report, *Lexicon of Lies*, for Data & Society:

“Journalists, commentators, policymakers, and scholars have a variety of words at their disposal — propaganda, dis-information, mis-information, and so on — to describe the accuracy and relevance of media content. These terms can carry a lot of baggage. They have each accrued different cultural associations and historical meanings, and they can take on different shades of meaning in different contexts. These differences may seem small, but they matter. The words we choose to describe media manipulation can lead to assumptions about how information spreads, who

spreads it, and who receives it. These assumptions can shape what kinds of interventions or solutions seem desirable, appropriate, or even possible.”³⁷

Visuals, Visuals, Visuals

As well as the other problematic aspects of the popular term ‘fake news’ outlined above, it has also allowed the debate to be framed as a textual problem. The focus on fabricated news ‘sites’ means the implications of misleading, manipulated or fabricated visual content, whether that’s an image, a visualization, a graphic, or a video are rarely considered. The solutions by the technology companies have been aimed squarely at articles, and while admittedly that is because natural language processing is more advanced, and therefore text is easier to analyse computationally, the framing of the debate as ‘fake news’, has not helped.

As we describe in this report, visuals can be far more persuasive than other forms of communication³⁸, which can make them much more powerful vehicles for mis- and dis-information. In addition, over the past couple of months, we’ve been confronted with the technological implications whereby relatively limited audio or video clips of someone can act as very powerful ‘training data’ allowing for the creation of completely fabricated audio or video files, making it appear that someone has said something that they have not.³⁹

Source-Checking vs Fact-Checking

There is much discussion of fact-checking in this report. There has been an explosion of projects and initiatives around the world, and this emphasis on providing additional context to public statements is a very positive development. Many of these organisations are focused on authenticating official sources: politicians, reports by think tanks or news reports (a list of European fact-checking organization are listed in Appendix A), but in this age of dis-information where we are increasingly seeing information created by unofficial sources (from social media accounts we don’t know, or websites which have only recently appeared), we argue that we need to be doing source-checking as well as fact-checking.

Increasingly, when assessing the credibility of a piece of information, the source who originally created the content or first shared it, can provide the strongest evidence about whether something is accurate. Newsrooms, and people relying on social media for information, need to be investigating the source, almost before they look at the content itself. For example, routinely people should be researching the date and location embedded in

³⁷ Jack, C. (August 2017) *Lexicon of Lies*, Data & Society, https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf

³⁸ Birdsell, D. S., & Groarke, L. (1996). *Toward a theory of visual argument*. *Argumentation and Advocacy*, 33(1), 1-10

³⁹ WNYC Radio Lab (July 27, 2017) *Breaking News*, <http://www.radiolab.org/story/breaking-news/>

domain registration information of a supposed 'news site' to seeing whether it was created two weeks ago in Macedonia. Similarly, people should be instinctively checking whether a particular tweeted message has appeared elsewhere, as it could be that the same message was tweeted out by ten different accounts at exactly the same time, and six of them were located in other countries. Newsrooms in particular need more powerful tools to be able to visually map online networks and connections to understand how dis-information is being created, spread and amplified.

Strategic Silence

Newsrooms also need more powerful tools to help them understand how dis-information is moving across communities. In election monitoring projects First Draft has been involved with in France, the UK and Germany, Newswhip (a platform which helps newsrooms discover content before it goes viral) was used as a way of monitoring whether a piece of misleading, manipulated or fabricated content was predicted to be shared widely. Newswhip has a prediction algorithm which allows the user to see how many social interactions a piece of content has received at any given moment and to offer a prediction about how many interactions it would have twenty-four hours later. First Draft used this technology to inform decisions about what stories to debunk and which ones to ignore. If certain stories, rumours or visual content, however problematic, were not gaining traction, a decision was made not to provide additional oxygen to that information. The media needs to consider that publishing debunks can cause more harm than good, especially as agents behind dis-information campaigns see media amplification as a key technique for success. Debunks themselves can be considered a form of engagement. The news industry needs to come together to think about the implications of this type of reporting and the philosophical and practical aspects of incorporating these ideas related to strategic silence.

The Report

The report starts with a new conceptual framework for talking about information disorder, including three types, three phases and three elements. We then consider the specific challenges of filter bubbles and echo chambers, before moving on to examine the solutions that have been put in place to date (including those by the technology companies, education initiatives, the media and regulatory bodies). We end the report with a look at future trends, before wrapping up with some conclusions, and additional details about the thirty-four recommendations we are proposing.

Part 1: Conceptual Framework

Our conceptual framework has three components, each of which is also broken down into three parts:

1. The Three Types of Information Disorder: Dis-information, Mis-information and Mal-information
2. The Three Phases of Information Disorder: Creation, Production and Distribution
3. The Three Elements of Information Disorder: Agent, Message and Interpreter

The Three Types of Information Disorder

Much of the discourse on ‘fake news’ conflates three notions: mis-information, dis-information and mal-information. But it’s important to distinguish messages that are true from those that are false, and messages that are created, produced or distributed by “agents” who intend to do harm from those that are not:

- **Dis-information.** Information that is false and deliberately created to harm a person, social group, organization or country.
- **Mis-information.** Information that is false, but not created with the intention of causing harm.
- **Mal-information.** Information that is based on reality, used to inflict harm on a person, organization or country.

TYPES OF INFORMATION DISORDER

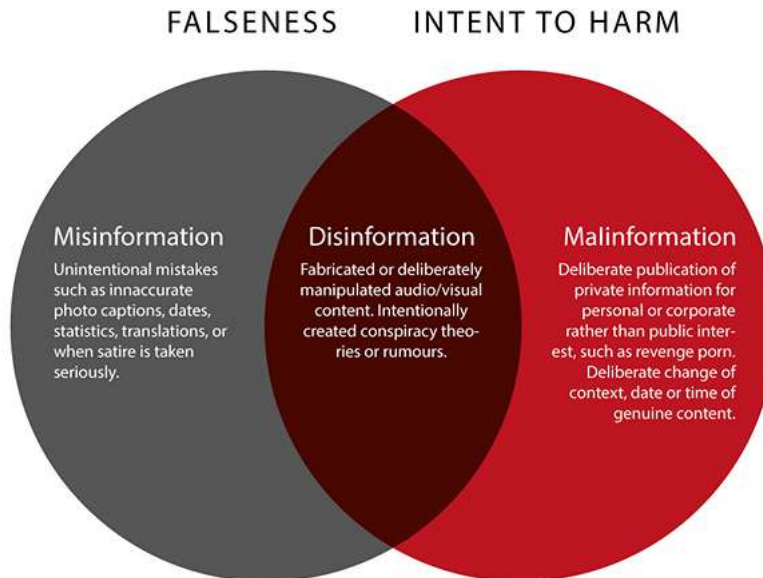


Figure 2: Examining how mis-, dis- and mal-information intersect around the concepts of falseness and harm. We include some types of hate speech and harassment under the mal-information category, as people are often targeted because of their personal history or affiliations. While the information can sometimes be based on reality (for example targeting someone based on their religion) the information is being used strategically to cause harm.

The 2017 French Presidential election⁴⁰ provides examples that illustrate all three types of information disorder.

1) Examples of dis-information:

One of the most high profile hoaxes of the campaign, was the creation of a sophisticated duplicate version of the Belgian newspaper *Le Soir*, with a false article claiming that Macron was being funded by Saudi Arabia.⁴¹ Another example was the circulation of documents online claiming falsely that Macron had opened an offshore bank account in the Bahamas.⁴² And finally, dis-information circulated via 'Twitter raids' in which loosely connected networks

⁴⁰ For an in depth analysis of Dis-information and the French President Election see also: Bakamo (2017a) *The Role and Impact of Non-Traditional Publishers in the 2017 French Presidential Election*. Available at: <https://www.bakamosocial.com/frenchelection/> and

Bakamo (2017b) *Patterns of Dis-information in the 2017 French Presidential Election*. Available at: <https://www.bakamosocial.com/frenchelection/>

⁴¹ CrossCheck, (March 2, 2017) *Was Macron's Campaign for the French Presidency Funded by Saudi Arabia?* CrossCheck, <https://crosscheck.firstdraftnews.com/checked-french/macrons-campaign-french-presidency-financed-saudi-arabia/>

⁴² CrossCheck (May 5, 2017) *Did Emmanuel Macron Open an Offshore Account?* CrossCheck, <https://crosscheck.firstdraftnews.com/checked-french/emmanuel-macron-open-offshore-account/>

of individuals simultaneously took to Twitter with identical hashtags and messages to spread rumours about Macron (e.g., that he was in a relationship with his step-daughter).

2) Examples of mis-information:

The attack on the Champs Elysees on 20 April 2017 inspired a great deal of mis-information⁴³, as is the case in almost all breaking news situations. Individuals on social media unwittingly published a number of rumours, for example the news that a second policeman had been killed. The people sharing this type of content are rarely doing so to cause harm. Rather, they are caught up in the moment, trying to be helpful, and fail to adequately inspect the information they are sharing.

3) Examples of mal-information:

One striking example of mal-information occurred when Emmanuel Macron's emails were leaked the Friday before the run-off vote on 7 May. The information contained in the emails was real, although Macron's campaign allegedly included false information to diminish the impact of any potential leak.⁴⁴ However, by releasing private information into the public sphere minutes before the media blackout in France, the leak was designed to cause maximum harm to the Macron campaign.

In this report, our primary focus is mis- and dis-information, as we are most concerned about false information and content spreading. However, we believe it's important to consider this third type of information disorder and think about how it relates to the other two categories. However, hate speech, harassment and leaks raise a significant number of distinct issues, and there is not space in this report to consider those as well. The research institute Data & Society is doing particularly good work on mal-information and we would recommend reading their report *Media Manipulation and Disinformation Online*.⁴⁵

The *Phases and Elements* of Information Disorder

In trying to understand any example of information disorder, it is useful to consider it in three elements:

- 1) **Agent.** Who were the 'agents' that created, produced and distributed the example, and what was their motivation?

⁴³ One example was the rumour that London Muslims were celebrating the attack on the Champs Elysee, which was debunked by the CrossCheck project: CrossCheck, (April 22, 2017) *Did London Muslims 'celebrate' a terrorist attack on the Champs-Elysees?* CrossCheck, <https://crosscheck.firstdraftnews.com/checked-french/london-muslims-celebrate-terrorist-attack-champs-elysees/>

⁴⁴ Rachel Donadio, (May 8, 2017) *Why the Macron Hacking Attack Landed with a Thud in France*, The New York Times, <https://www.nytimes.com/2017/05/08/world/europe/macron-hacking-attack-france.html>

⁴⁵ Marwick, A and R. Lewis (May 2017) *Media Manipulation and Dis-information Online*, Data & Society, <https://datasociety.net/output/media-manipulation-and-disinfo-online/>,

- 2) **Message.** What type of message was it? What format did it take? What were the characteristics?
- 3) **Interpreter.** When the message was received by someone, how did they interpret the message? What action, if any, did they take?



Figure 3: The Three Elements of Information Disorder

We argue that it is also productive to consider the life of an example of information disorder as having three phases:

1. **Creation.** The message is created.
2. **Production.** The message is turned into a media product.
3. **Distribution.** The message is distributed or made public.

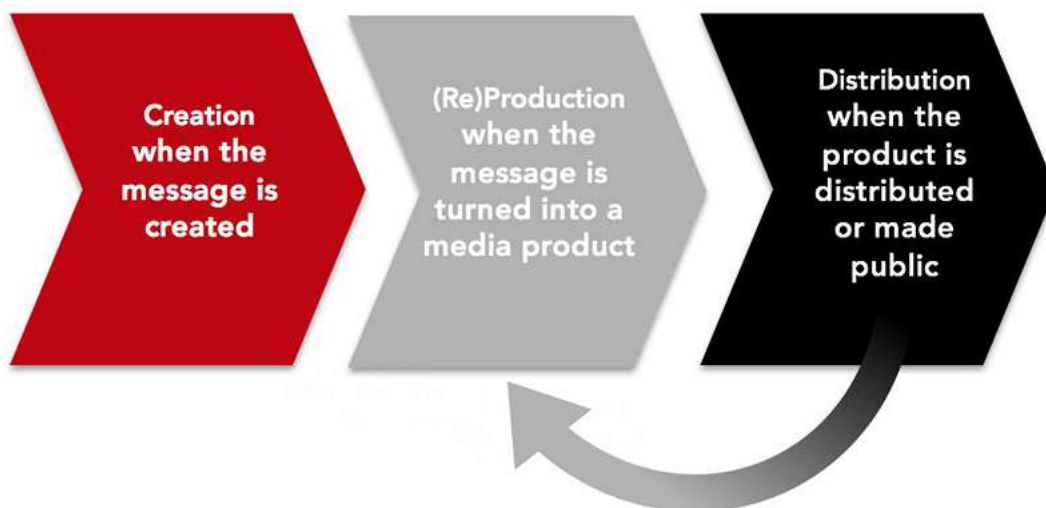


Figure 4: The Three Phases of Information Disorder

In particular, it's important to consider the different phases of an instance of information disorder alongside its elements, because the agent that creates the content is often fundamentally different from the agent who produces it. For example, the motivations of the

mastermind who ‘creates’ a state-sponsored dis-information campaign are very different from those of the low-paid ‘trolls’ tasked with turning the campaign’s themes into specific posts. And once a message has been distributed, it can be reproduced and redistributed endlessly, by many different agents, all with different motivations. For example, a social media post can be distributed by several communities, leading its message to be picked up and reproduced by the mainstream media and further distributed to still other communities. Only by dissecting information disorder in this manner can we begin to understand these nuances.

In the next two sections, we will examine these elements and phases of information disorder in more detail.

The Three Phases of Information Disorder

To examine how the phases of creation, production and distribution help us understand information disorder, let’s use the example of the article ‘Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement’ published on the self-proclaimed fantasy news site WTOE 5 in July 2016. For an in-depth analysis of this article and the network of sites connected to it, we would recommend reading ‘The True Story Behind The Biggest Fake News Hit Of The Election’ from BuzzFeed.⁴⁶



Figure 5: Screenshot of the fabricated news article published in July 2016 on WTOE5News.com (The site no longer exists).

⁴⁶ Craig Silverman (Dec 2016) *The True Story Behind The Biggest Fake News Hit Of The Election*, BuzzFeed, https://www.buzzfeed.com/craigsilverman/the-strangest-fake-news-empire?utm_term=.yrzPyEpLXq#.nqXD9N7opO

If we think about the three phases in this example, we can see how different agents were involved in creating the impact of this content.

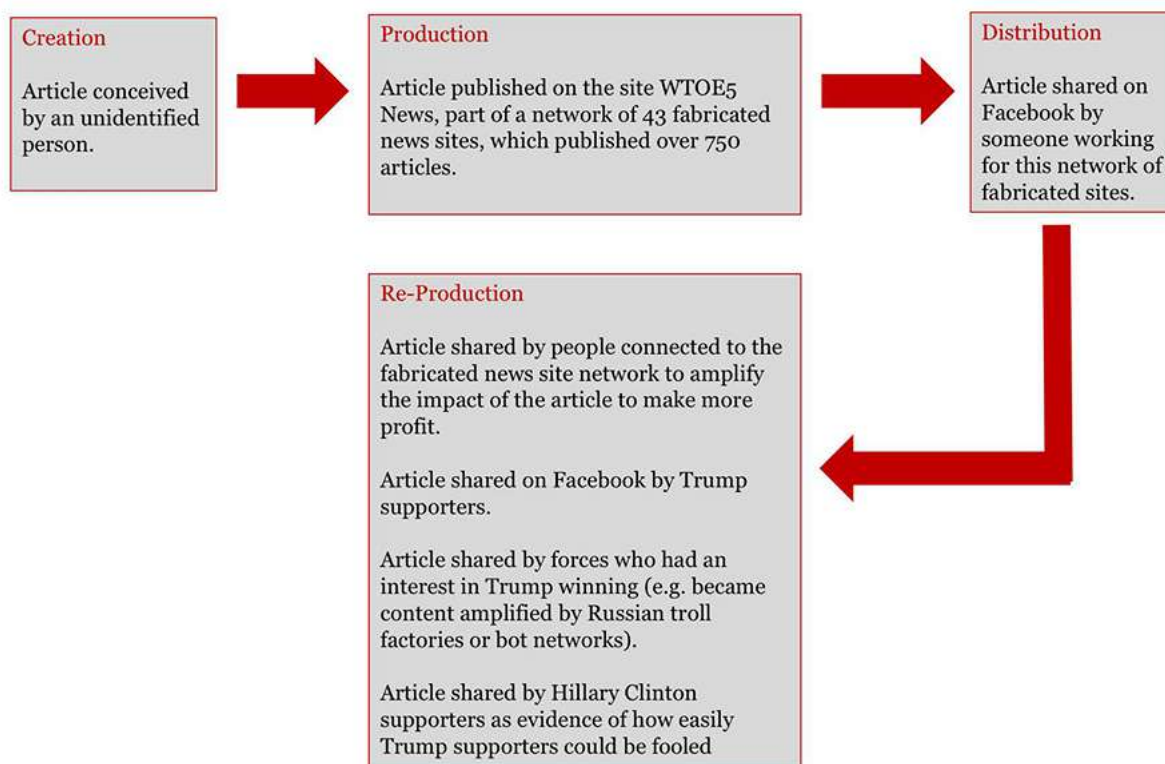


Figure 6: Using the example of the ‘Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement’ fabricated news articles to test the Three Phases of Information Disorder

The role of the mainstream media as agents in amplifying (intentionally or not) fabricated or misleading content is crucial to understanding information disorder. Fact-checking has always been fundamental to quality journalism, but the techniques used by hoaxers and those attempting to disseminate dis-information have never been this sophisticated. With newsrooms increasingly relying on the social web for story ideas and content, forensic verification skills and the ability to identify networks of fabricated news websites and bots is more important than ever before.

The Three Elements of Information Disorder

The Agent

Agents are involved in all three phases of the information chain – creation, production and distribution – and have various motivations. Importantly, the characteristics of agents can vary from phase to phase.

We suggest seven questions to ask about an agent:

1) What type of actor are they?

Agents can be official, like intelligence services, political parties, news organizations. They can also be unofficial, like groups of citizens that have become evangelized about an issue.

2) How organized are they?

Agents can work individually, in longstanding, tightly-organized organizations (e.g., PR firms or lobbying groups) or in impromptu groups organized around common interests.

3) What are their motivations?

There are four potential motivating factors: **Financial:** Profiting from information disorder through advertising; **Political:** Discrediting a political candidate in an election and other attempts to influence public opinion; **Social:** Connecting with a certain group online or off; and **Psychological:** Seeking prestige or reinforcement.

4) Which audiences to they intend to reach?

Different agents might have different audiences in mind. These audiences can vary from an organization's internal mailing lists or consumers, to social groups based on socioeconomic characteristics, to an entire society.

5) Is the agent using automated technology?

The ability to automate the creation and dissemination of messages online has become much easier and, crucially, cheaper. There is much discussion about how to define a bot. One popular definition from the Oxford Internet Institute is an account that posts more than 50 times a day, on average. Such accounts are often automated, but could conceivably be operated by people. Other accounts, known as cyborgs, are operated jointly by software and people.

6) Do they intend to mislead?

The agent may or may not intend to deliberately mislead the target audience.

7) Do they intend to harm?

The agent may or may not intend deliberately to cause harm

The Message

Messages can be communicated by agents in person (via gossip, speeches, etc.), in text (newspaper articles or pamphlets) or in audio/visual material (images, videos, motion-graphics, edited audio-clip, memes, etc.). While much of the current discussion about 'fake news' has focused on fabricated text articles, mis- and dis-information often appears in visual

formats. This is important, as technologies for automatically analysing text are significantly different from those for analysing still and moving imagery.

We offer five questions to ask about a message:

How durable is the message?

Some messages are designed to stay relevant and impactful for the long term (throughout an entire war or in perpetuity). Others are designed for the short term (during an election) or just one moment, as in the case of an individual message during a breaking news event.

1. How accurate is the message?

The accuracy of a message is also important to examine. As discussed earlier, mal-information is truthful information used to harm (either by moving private information into the public arena or using people's affiliations, like their religion, against them). For inaccurate information, there is a scale of accuracy from false connection (a clickbait headline that is mismatched with its article's content) to 100% fabricated information.

2. Is the message legal?

The message might be illegal, as in the cases of recognised hate speech, intellectual property violations, privacy infringements or harassment. Of course, what messages are legal differs by jurisdiction.

3. Is the message 'imposter content', i.e. posing as an official source?

The message may use official branding (e.g., logos) unofficially, or it may steal the name or image of an individual (e.g., a well-known journalist) in order to appear credible.

4. What is the message's intended target?

The agent has an intended audience in mind (the audience they want to influence) but this is different to the target of the message (those who are being discredited). The target can be an individual (a candidate or a political or business leader), an organisation (a private firm or a government agency), a social group (a race, ethnicity, the elite, etc.) or an entire society.

The Interpreter

Audiences are very rarely passive recipients of information. An 'audience' is made up of many individuals, each of which interprets information according to his or her own socio-cultural status, political positions and personal experiences.

As outlined earlier, understanding the ritualistic aspect of communication is critical for understanding how and why individuals react to messages in different ways. The types of information we consume, and the ways in which we make sense of them, are significantly impacted by our self-identity and the 'tribes' we associate with. And, in a world where what

we like, comment on and share is visible to our friends, family and colleagues, these 'social' and performative forces are more powerful than ever.

Having to accept information that challenges our sense of self can be jarring. Irrespective of how persuasive a message may appear to a neutral observer, it is easier to ignore or resist information that opposes our own worldview. Certainly, evidence suggests that fact-checks do tend to nudge individuals' knowledge in the direction of the correct information, but it certainly doesn't replace the mis- or dis-information entirely.

This reality complicates our search for solutions to information disorder. If we accept that human brains do not always work rationally, simply disseminating more quality information is not the answer. Solutions must grapple with the social and performance characteristics that have helped make certain fabricated content so popular on Facebook. How, for example, can we make sharing false information publicly shameful and embarrassing? What can we learn from the theories of performativity, particularly in performance and identity management in an online setting that could help us experiment with some potential solutions?

What the 'interpreter' can do with a message highlights how the three elements of information disorder should be considered parts of a potential never-ending cycle. In an era of social media, where everyone is a potential publisher, the interpreter can become the next 'agent,' deciding how to share and frame the message for their own networks. Will they show support for the message by liking or commenting on it, or will they decide to share the message? If they do share the message, have they done so with the same intent as the original agent, or will they share it to, for example, show their disagreement?



Figure 7: Questions to ask about each element of an example of information disorder

Agent						
Type of Actor	Motivation	Level of organization	Use of automation	Intended audience	Intent to harm	Intent to mislead
<i>Unofficial actor</i>	<i>Financial</i>	<i>Connected to a network of sites</i>	<i>No</i>	<i>US voters</i>	<i>Yes</i>	<i>Yes</i>
Message						
Format	Durability	Level of accuracy	Imposter	Legality	Target of message	
<i>Text</i>	<i>Election campaign</i>	<i>Fabricated</i>	<i>No</i>	<i>Yes</i>	<i>Hillary Clinton</i>	
Interpreter						
How is the message interpreted by different people?	What action is taken? Is the message re-produced?					

Figure 8: Using the ‘Three Elements of Information Disorder’ to examine the ‘Pope Endorses Trump’ article

In the next section, we will review literature that helps provide a deeper historical and theoretical understanding of the three elements of information disorder.

1) The Agents: Who are they and what motivates them?

In this section, we explore the role of agents, or those who create, produce and distribute messages. Again, the motivations of the person who creates and posts a meme on an invite-only chat group on Discord could be different from the person who sees the meme on their Facebook feed and shares it with a WhatsApp group.

Official vs Unofficial Actors?

When official actors are involved, the sophistication, funding and potential impact of a message or campaign of systematic messages is far greater. Much has been written about the impact of Russian propaganda on information ecosystems in Europe and further afield. One of the most notable is the Rand Corporation's report from July 2016, entitled "The Russian 'Firehose of Falsehood' Propaganda Model,"⁴⁷ which identified four characteristics of modern Russian propaganda:

1. Voluminous and multi-channeled
2. Rapid, continuous and repetitive
3. Noncommittal to objective reality
4. Inconsistent in its messaging

The EU Stratcomm Taskforce provides regular analysis of Russian propaganda messaging across the European Union.⁴⁸ Likewise, their research shows that a key strategy of Russia is to spread as many conflicting messages as possible, in order to persuade audiences that there are too many versions of events to find the truth. As they explain, "Not only (are) big media outlets like Russia Today or Sputnik ... deployed, but also seemingly marginal sources, like fringe websites, blog sites and Facebook pages. Trolls are deployed not only to amplify dis-information messages but to bully those... brave enough to oppose them. And the network goes wider: NGOs and "GONGOs" (government organised NGOs); Russian government representatives; and other pro-Kremlin mouthpieces in Europe, often on the far-right and far-left. In all, literally thousands of channels are used to spread pro-Kremlin dis-information, all creating an impression of seemingly independent sources confirming each other's message."⁴⁹

⁴⁷ Paul, Christopher and Miriam Matthews, (June 20, 2016) *The Russian "Firehose of Falsehood" Propaganda Model: Why It Might Work and Options to Counter It*, Santa Monica, Calif.: RAND Corporation, <https://www.rand.org/pubs/perspectives/PE198.html>

⁴⁸ European Union's East StratCom Task Force, <https://euvsdisinfo.eu/>

⁴⁹ EU East StratCom Task Force, (Jan 19, 2017), Means, goals and consequences of the pro-Kremlin dis-information campaign, *ISPI* <http://www.ispionline.it/it/pubblicazione/means-goals-and-consequences-pro-kremlin-dis-information-campaign-16216>

In April 2017, Facebook published a paper by three members of its Security team, entitled “Information Operations and Facebook,” that outlines the use of the platform by state actors. They define information operations as “actions taken by organized actors (governments or non-state actors) to distort domestic or foreign political sentiment, most frequently to achieve a strategic and/or geopolitical outcome. These operations can use a combination of methods, such as false news, dis-information or networks of fake accounts aimed at manipulating public opinion (we refer to these as ‘false amplifiers’).”⁵⁰

While Russian propaganda techniques are the current focus of much concern, digital astroturfing campaigns – that is, campaigns that use troll factories, click farms and automated social media accounts – have been used by other state actors for years. A recent report by the Computational Propaganda Research Project tracked this activity across twenty-eight countries, showing the scale of these operations.⁵¹

Perhaps the most notable of these state actors is China, which has paid people to post millions of fabricated social media posts per year, as part of an effort to “regularly distract the public and change the subject” from any policy-related issues that threaten to incite protests.⁵² In countries like Bahrain and Azerbaijan, there is evidence of PR firms creating fake accounts on social media to influence public opinion.⁵³ Duterte’s government has used sophisticated ‘astroturfing’ techniques to target individual journalists and news organizations.⁵⁴

Additionally, in South Africa, an email leak in May exposed large-scale dis-information efforts by the powerful Gupta family to distract attention from its business dealings with the government. These efforts included paying Twitter users to abuse journalists and spread dis-information and the use of bots to amplify fabricated stories.⁵⁵

⁵⁰ Jen Weedon, William Nuland and Alex Stamos (April 27, 2017) *Information Operations and Facebook*, p. 4
<https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf>

⁵¹ Bradshaw, S. and P. Howard (August 2017) *Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation*
<http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/07/Troops-Trolls-and-Troublemakers.pdf>

⁵² King, G, J. Pan & M. Roberts, (May 2016) *How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument*, Harvard University,
<http://gking.harvard.edu/files/gking/files/50c.pdf?m=1463587807>

⁵³ Woolley, S and P. Howard (2017) *Social Media, Revolution and the Political Bot*, in *Routledge Handbook of Media, Conflict and Security*, edited by Piers Robinson, Philip Seib, Romy Frohlich, London:Routledge.

⁵⁴ Julie Posetti, *This is why Online Harassment Still Needs Attention*, MediaShift July 2017,
<http://mediashift.org/2017/07/online-harassment-still-needs-attention/>

⁵⁵ Eliseev, A (July 20, 2017) *The Gupta scandal: how a British PR firm came unstuck in South Africa*, *The New Statesman*, <http://www.newstatesman.com/culture/observations/2017/07/gupta-scandal-how-british-pr-firm-came-unstuck-south-africa>

In contrast to official actors, unofficial actors are those who work alone or with loose networks of citizens, and create false content to harm, make money, or entertain other like-minded people.

Following the outcry about the role of fabricated websites in the 2016 US election, journalists tracked down some of these ‘unofficial’ agents. One was Jestin Coler, who, in an interview with NPR, admitted that his “whole idea from the start was to build a site that could kind of infiltrate the echo chambers of the alt-right, publish blatantly [false] or fictional stories and then... publicly denounce those stories and point out the fact that they were fiction.” As NPR explains, “[Coler] was amazed at how quickly fake news could spread and how easily people believe[d] it.”⁵⁶

How organised are the agents?

Trolls have existed since the internet was invented.⁵⁷ Definitions vary, but one aspect is key: trolls provoke emotions by publicly offending their targets. Trolls are humans who post behind a username or handle. Yet, similar to bots, they can amplify dis-information in coordinated ways to evoke conformity among others. What they do better than bots is target those who question the veracity of a piece of information. Trolls work efficiently to silence naysayers in the early stages of dis-information distribution by posting personal attacks to undermine that person’s position on the board. And we know that some governments organize agents to pursue specific messaging goals on social media, whether through bots, cyborgs or ‘troll factories.’⁵⁸

In the report entitled ‘Media Manipulation and Disinformation Online’ from Data & Society, Alice Marwick and Rebecca Lewis analyzed ‘Gamergate’, an online campaign of bullying and harassment that took place in late 2014. They identified organized brigades, networked and agile groups, men’s rights activists and conspiracy theorists as exploiting “young men’s rebellion and dislike of ‘political correctness’ to spread white supremacist thought, Islamophobia, and misogyny through irony and knowledge of internet culture”.⁵⁹

⁵⁶ Sydell, L. (2017) (Nov 23, 2016) *We Tracked Down A Fake-News Creator In The Suburbs. Here’s What We Learned*, NPR, <http://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs>

⁵⁷ De Seta, G., “*Trolling, and Other Problematic Social Media Practices*,” in *The SAGE Handbook of Social Media*, ed. Jean Burgess, Alice E. Marwick, and Thomas Poell (Thousand Oaks, CA: SAGE Publications, 2017).

⁵⁸ Benedictus, L., “*Invasion of the troll armies: from Russian Trump supporters to Turkish state stooges*,” *Guardian* <https://www.theguardian.com/media/2016/nov/06/troll-armies-social-media-trump-russian>

⁵⁹ Marwick, A and R. Lewis (May 2017) *Media Manipulation and Dis-information Online*, Data & Society, <https://datasociety.net/output/media-manipulation-and-disinfo-online/>, p.29

Buzzfeed's Ryan Broderick examined similar, loosely-affiliated groups of Trump supporters in the US who were active during the French election.⁶⁰ Using technologies like Discord (a set of invite-only chat rooms), Google documents, Google forms and Dropmark (a file-sharing site like Dropbox), they organized 'Twitter raids' where they would simultaneously bombard Twitter accounts they hoped to influence with messages using the same hashtags.

Analysis of mis-information during the French election by Storyful and the Atlantic Council showed that such loose, online networks of actors push messages across different platforms. Anyone wishing to understand their influence needs to monitor several closed and open platforms. For example, in the context of the US election, Trump supporters produced and "audience-tested many anti-Clinton memes in 4Chan and fed the ones with the best responses into the Reddit forum 'The_Donald.' The Trump campaign also monitored the forum for material to circulate in more mainstream social media channels."⁶¹

Finally, it's worth mentioning 'fake tanks', or partisan bodies disguised as think tanks. As Transparify, the group that provides global ratings on the financial transparency of think tanks, has explained, "[T]hese [fake tanks] range from essentially fictitious entities purposefully set up to promote the very narrow agenda and vested interests of (typically one single) hidden funder at one extreme, to more established organisations that work on multiple policy issues but (occasionally or routinely) compromise their intellectual independence and research integrity in line with multiple funders' agendas and vested interests..."

Representatives from fake tanks "regularly appear on television, radio, or in newspaper columns to argue for or against certain policies, their credibility bolstered by the abuse of the think tank label and misleading job titles such as "senior scholar."⁶²

What is the motivation of the agent?

Looking into what motivates agents not only provides a deeper understanding of how dis- or mal-information campaigns work, it also points to possible ways to resist them.

It is a mistake to talk generally about agents' motivations, since they vary in each phase. It is quite likely that publishers (e.g., an editor of a cable news show) or distributors (e.g., a user

⁶⁰ Broderick, R. (Jan 24, 2017) *Trump Supporters Online Are Pretending To Be French To Manipulate France's Election*, BuzzFeed, https://www.buzzfeed.com/ryanhatesthis/inside-the-private-chat-rooms-trump-supporters-are-using-to?utm_term=.skLbZaN42#.xnOE5Jybl

⁶¹ Shaffer, K. et al. (2017) *Democracy Hacked*. Available at: <https://medium.com/data-for-democracy/democracy-hacked-a46c04d9e6d1>

⁶² Written evidence provided to the UK Parliamentary Inquiry on Fake News by Transparify <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/culture-media-and-sport-committee/fake-news/written/47967.html>

on a social network) of a message may not even be fully aware of the real purpose behind a piece of dis-information.

As we illustrate above, if a message is partly or entirely false, but no harm is intended by its producer, it doesn't fall under the definition of dis-information. For this reason, it's important to differentiate between mis-information (false, but not intended to harm) and mal-information (true, but intended to harm).

i) Political

Producers of dis-information campaigns, from Russia and elsewhere, sometimes have political motivations. A great deal has been written about Russian dis-information activity in Europe, but it's worth quoting at length from a statement given by Constanze Stelzenmüller to the US Senate Committee on Intelligence, in June 2017, on the subject of potential interference by Russia in the German Federal elections:

“Three things are new about Russian interference today. Firstly, it appears to be directed not just at Europe's periphery, or at specific European nations like Germany, but at destabilizing the European project from the inside out: dismantling decades of progress toward building a democratic Europe that is whole, free, and at peace. Secondly, its covert and overt “active measures” are much more diverse, larger-scale, and more technologically sophisticated; they continually adapt and morph in accordance with changing technology and circumstances. Thirdly, by striking at Europe and the United States at the same time, the interference appears to be geared towards undermining the effectiveness and cohesion of the Western alliance as such—and at the legitimacy of the West as a normative force upholding a global order based on universal rules rather than might alone. That said, Russia's active measures are presumably directed at a domestic audience as much as towards the West: They are designed to show that Europe and the U.S. are no alternative to Putin's Russia. Life under Putin, the message runs, may be less than perfect; but at least it is stable.”⁶³

In terms of Russian dis-information, one of the best sources of information is the EU East StratCom Task Force,⁶⁴ which has a site called ‘euvsdisinfo.eu’ that provides regular updates about Russian dis-information campaigns across Europe. As they explain, “the dis-information campaign is a non-military measure for achieving political goals. Russian authorities are explicit about this, for example through the infamous Gerasimov doctrine and through

⁶³ Stelzenmüller, C. (June 28, 2017) *Testimony to the US Senate Committee on Intelligence: The impact of Russian interference on Germany's 2017 elections*, Available at: <https://www.brookings.edu/testimonies/the-impact-of-russian-interference-on-germanys-2017-elections/>

⁶⁴ For those interested in the subject of Russian dis-information, there is great deal of excellent analysis including the recent report by Flemming Splidsboel Hansen of the Danish Institute for International study.

statements by top Russian generals that the use of ‘false data’ and ‘destabilising propaganda’ are legitimate tools in their tool kit.” Elsewhere, the Task Force wrote, “The Russian Minister of Defence describes information as ‘another type of armed forces.’”⁶⁵

One critical aspect to understanding Russian dis-information, as noted by information warfare expert Molly McKew, is that “information operations aim to mobilize actions and behavioral change. It isn't just information.”⁶⁶ As the recent revelations about Russian operatives purchasing dark ads on Facebook⁶⁷ and organizing protests via Facebook’s Events feature⁶⁸ show, the aim of these acts is to create division along socio-cultural lines.

ii) Financial

Some of those who produce or distribute dis-information may do so merely for financial gain, as in the case of PR firms and fabricated news outlets. Indeed, entire businesses might be based on dis-information campaigns.⁶⁹

Fabricated ‘news’ websites created solely for profit have existed for years. Craig Silverman documented some of the most prolific in the US in his 2015 report⁷⁰ for the Tow Center for Digital Journalism. However, the US election shone a light on how many of these sites are located overseas, but aimed at US audiences. *Buzzfeed* was one of the first news organisations to detail the phenomenon of English-language websites created by Macedonians to capitalise on US readers’ enthusiasm for sensationalist stories.⁷¹ The small city of Veles in Macedonia produced “an enterprise of cool, pure amorality, free not only of ideology but of any concern or feeling about the substance of the election. These

⁶⁵ EU East StratCom Task Force, (Jan 19, 2017), *Means, goals and consequences of the pro-Kremlin dis-information campaign*, ISPI <http://www.ispionline.it/it/pubblicazione/means-goals-and-consequences-pro-kremlin-dis-information-campaign-16216>

⁶⁶ Molly McKew on Twitter <https://twitter.com/MollyMcKew/status/907585015915171840?t=1&cn=ZmxleGlibGVfcmVjc18y&refsrc=email&iid=618ad44aaddf4116ac68a52cd832ee09&uid=20131383&nid=244+293670929>

⁶⁷ Stamos, A. (September 6, 2017) *An Update On Information Operations On Facebook*, Facebook Newsroom, <https://newsroom.fb.com/news/2017/09/information-operations-update/>

⁶⁸ *Facebook Events to Organize Anti-Immigrant Rallies on U.S. Soil*, The Daily Beast, <http://www.thedailybeast.com/exclusive-russia-used-facebook-events-to-organize-anti-immigrant-rallies-on-us-soil>

⁶⁹ Tambini, D. (2017) *How advertising fuels fake news*. LSE Media Policy Project Blog, <http://blogs.lse.ac.uk/mediapolicyproject/2017/02/24/how-advertising-fuels-fake-news/>

⁷⁰ Silverman, C. (Feb. 2015) *Lies, Damn Lies and Viral Content*, Tow Center for Digital Journalism, <http://towcenter.org/research/lies-damn-lies-and-viral-content/>

⁷¹ Silverman, C. and L. Alexander (Nov. 3, 2016) *How Teens In The Balkans Are Duping Trump Supporters With Fake News*, *Buzzfeed*, https://www.buzzfeed.com/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo?utm_term=.vxz8qAVJj#.vukajzr2R

Macedonians on Facebook didn't care if Trump won or lost the White House. They only wanted pocket money to pay for things."⁷²

This example from Veles also underscores the difficulty of assessing the true motivation of any particular agent. The dominant narrative has been that these young people were motivated by the financial benefits. We can assume this is true, as they undoubtedly made money, but we will unlikely ever know whether there was any coordinated attempt to encourage these teenagers to start this type of work in the first place.

'Fake news' websites make money through advertising. While Google and Facebook have taken steps to prevent these sites from getting money through their ad networks, there are still many other networks through which site owners can make money.

French startup Storyzy alerts brands when they appear on dubious websites. In a August 2017 write-up of their work, Frederic Filloux explains that over 600 brands had advertisements on questionable sites. When they were approached for comment, Filloux concluded that few cared, as long as their "overall return on investment was fine." They certainly did not seem to consider the ethical implications of helping to fuel a 'vast network of mis-information.'⁷³

iii) Social and Psychological

While much of the debate around dis-information has focused on political and financial motivations, we argue that understanding the potential social and psychological motivations for creating dis-information is also worth exploring.

For example, consider the motivation to simply cause trouble or entertain. There have always been small numbers of people trying to 'hoax' the news media—from Tommaso Debenedetti, who frequently uses fake Twitter accounts to announce the death of high profile people,⁷⁴ to the person behind the 'Marie Christmas' account, who fooled CNN into thinking he or she was a witness to the San Bernardino shooting.⁷⁵

Some share mis-information as a joke, only to find that people take it seriously. Most recently, during Hurricane Harvey, Jason Michael McCann tweeted the old, already-debunked image from Hurricane Sandy of a shark swimming in a flood highway. When Craig Silverman

⁷² Subramanian, S. (2017) *Inside the Macedonian Fake News Complex*, Wired, February 15, 2017. Wired. Available at: <https://www.wired.com/2017/02/veles-macedonia-fake-news>

⁷³ Filloux, F. (August 21, 2017) *More than 600 global brands still feed the fake news ecosystem*, The Monday Note on Medium, <https://mondaynote.com/more-than-600-global-brands-still-feed-the-fake-news-ecosystem-d1ddfdb80458>

⁷⁴ Kington, T. (March 30, 2010) *Twitter hoaxer comes clean and says: I did it to expose weak media*, The Guardian, <https://www.theguardian.com/technology/2012/mar/30/twitter-hoaxer-tommaso-de-benedetti>

⁷⁵ Buttry, S. (2015) *'Marie Christmas:' Some journalists fell for San Bernardino prank; others backed away*, The Buttry Diary, <https://stevebuttry.wordpress.com/2015/12/03/the-case-of-marie-christmas-verifying-eyewitnesses-isnt-simple-or-polite/>

reached out to him for comment, he explained, “Of course I knew it was fake, it was part of the reason I shared the bloomin' thing... What I had expected was to tweet that and have my 1,300 followers in Scotland to laugh at it.”

On more serious matters, the previously-mentioned research by Marwick and Lewis⁷⁶ takes a deep dive into alt-right communities and discusses the importance of considering their shared identity in understanding their actions online.

Examining the audiences of hyper-partisans sites, such as *Occupy Democrats* in the US and *The Canary* in the UK, we can also see the influence of political tribalism and identity. These types of sites do not peddle 100% fabricated content, but they are very successful in using emotive (and some would argue misleading) headlines, images and captions – which are often all of an article that’s read on platforms like Facebook – to get their audiences to share their messages.

In August 2017, Silverman and his colleagues at BuzzFeed published the most comprehensive study to date of the growing universe of US-focused, hyper-partisan websites and Facebook pages. They revealed that, in 2016 alone, at least 187 new websites launched, and that the candidacy and election of Donald Trump “unleashed a golden age of aggressive, divisive political content that reaches a massive amount of people on Facebook.”⁷⁷

Is the agent using automation?

Currently, machines are poor at creating dis-information, but they can efficiently publish and distribute it. Recent research by Shao and colleagues concluded that “[a]ccounts that actively spread mis-information are significantly more likely to be bots.” They also found that bots are “particularly active in the early spreading phases of viral claims, and tend to target influential users.”⁷⁸

Bots can manipulate majority-oriented platform algorithms to gain vast visibility and can create conformity among human agents who would then further distribute their messages.⁷⁹

⁷⁶ Marwick, A and R. Lewis (May 2017) *Media Manipulation and Dis-information Online*, Data & Society, <https://datasociety.net/output/media-manipulation-and-disinfo-online/>, p.29

⁷⁷ Silverman et al. (August 8, 2017) *Inside the Partisan Fight for your NewsFeed*, BuzzFeed News, <https://www.buzzfeed.com/craigsilverman/inside-the-partisan-fight-for-your-news-feed/>

⁷⁸ Shao, C., G.L. Ciampaglia, O. Varol, A. Flammini and F. Menczer, (July 24, 2017) *The spread of fake news by social bots*, <https://arxiv.org/pdf/1707.07592.pdf>

⁷⁹ Woolley, S & P. Howard, (2017b) *Computational Propaganda Worldwide: An Executive Summary*, <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Casestudies-ExecutiveSummary.pdf>

Many bots are designed to amplify the reach of dis-information⁸⁰ and exploit the vulnerabilities that stem from our cognitive and social biases. They also create the illusion that several individuals have independently come to endorse the same piece of information.⁸¹ As a recent report on computational amplification by Gu *et al.* concluded: “A properly designed propaganda campaign is designed to have the appearance of peer pressure—bots pretending to be humans, guru accounts that have acquired a positive reputation in social media circles—these can make a propaganda campaign-planted story appear to be more popular than it actually is.”⁸² Despite the platforms’ public commitment to stifle automated accounts, bots continue to amplify certain messages, hashtags or accounts, creating the appearance of certain perspectives being popular and, by implication, true.⁸³

A recent report by NATO StratCom entitled ‘Robotrolling’ found that two in three Twitter accounts posting in Russian about the NATO presence in the Baltics and Poland were bots. They also found the density of bots is 2 to 3 times greater among Russian-tweeting accounts than in English-tweeting accounts. The authors conclude that foreign-language sources on social networks are policed and moderated much less effectively than English-language sources.⁸⁴

It also seems possible that there may be a black market for social bots. Ferrara found that many bots who supported Trump in the 2016 election also engaged with the #MacronLeaks trend, but made few posts in-between.⁸⁵

Important research has been done on bots recently, particularly in terms of thinking about their definition, scale and influence. The most comprehensive body of research has been carried out by the Oxford Internet Institute’s Computational Propaganda Research Project.⁸⁶ They define high-frequency accounts as those that tweet more than 50 times per day on average. While often these accounts are bots, we also need to realise there are some humans

⁸⁰ Shao, C. et al., (2016) *Hoaxy: A Platform for Tracking Online Mis-information*, Proceedings of the 25th International Conference Companion on World Wide Web, pp. 745-750

⁸¹ Ratkiewicz et al., (2011) *Detecting and Tracking Political Abuse in Social Media*, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2850>

⁸² Gu, L., V. Kropotov & F. Yarochkin, (June 2017), *How Propagandists Abuse the Internet and Manipulate the Public*. Trend Micro, https://documents.trendmicro.com/assets/white_papers/wp-fake-news-machine-how-propagandists-abuse-the-internet.pdf

⁸³ Ferrara, E. et al. (2016). *The rise of social bots*. Communications of the ACM, 59(7), p.101

⁸⁴ NATO Strategic Communications Centre of Excellence (September 2017) Robotrolling, <http://www.stratcomcoe.org/robotrolling-20171>

⁸⁵ Ferrara, E. et al. (2016)

⁸⁶ Woolley, S & P. Howard, (2017)

who tweet that frequently. There are also cyborg accounts⁸⁷, which are jointly operated by people and software. As Nic Dias argues, looking at an account's posting frequency can be more useful than a fixation on whether an account is fake or not.⁸⁸

There are certainly highly partisan individuals whose accounts could be mistaken for bots. A Politico article in August 2017 described how tens of thousands of tweets per day continue to emanate from a very human, grassroots organisation. Using Group Direct Messages on Twitter, they organize people into "invite-only rooms with names like 'Patriots United' and 'Trump Train. Many rooms have accompanying hashtags to track members' tweets as they propagate, and each can accommodate as many as 50 people."⁸⁹

Returning to our discussion around agents' motivations, these examples show the power of social and psychological motivations for creating and disseminating mis- and dis-information. Being a part of the tribe is a powerful, motivating force.

2) The Messages: What format do they take?

In the previous section we examined the different characteristics of the 'agents' those who are involved in creating, producing or disseminating information disorder. We now turn our attention to the messages themselves.

There are four characteristics that make a message more appealing and thus more likely to be consumed, processed and shared widely:

- 1) It provokes an emotional response.
- 2) It has a powerful visual component.
- 3) It has a strong narrative.
- 4) It is repeated.

Those who create information campaigns, true or false, understand the power of this formula. Identifying these characteristics helps us to recognize dis-information campaigns which are more likely to be successful, and to inform our attempts to counter dis-information (see more in Part Three).

⁸⁷ Chu, Z., S. Gianvecchio, H. Wang, S. Jajodia, (Nov/Dec 2012) *Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?* IEEE Transactions on Dependable and Secure Computing, Vol. 9, No. 6

⁸⁸ Dias, N. (June 22, 2017), *Reporting on a new age of digital astroturfing*, First Draft, <https://firstdraftnews.com/digital-astroturfing/>

⁸⁹ Musgrave, S. (August 9, 2017) *I Get Called a Russian Bot 50 Times a Day*, Politico, <http://www.politico.com/magazine/story/2017/08/09/twitter-trump-train-maga-echo-chamber-215470>

Verbal, text or audio?

While much of the conversation about mis- and dis-information has focused on the role of the internet in propagating messages, we must not forget that information travels by word of mouth. The offline and online worlds are not separate, although the challenges researchers face in effectively studying the effects of different forms of communication simultaneously means it's easier to think about these elements separately.

The 'fake news' conversation has also focussed on text-based, fabricated news websites. As Nausicaa Renner argues, "the fake news conversation has taken place in the realm of words, but that's missing a big part of the story. Much of the content that circulates on Facebook are images, often memes. They're not attached to an article, and there's often no way to trace their source. And while Facebook's algorithm is notoriously elusive, it seems to favor images and video over text. As such, images have the potential to reach more readers than articles — whether fake, real, non-partisan or hyper-partisan."⁹⁰

Certainly, in the election-based projects First Draft led in France and the U.K., visuals were overwhelmingly the most shared and the most difficult to debunk of misleading content. In both cases, while there were almost no examples of fabricated news sites as we saw in the US context, there were large numbers of highly shareable images, infographics and memes (i.e., compelling images with large block text layered over top.)⁹¹

⁹⁰ Renner, N. (Jan. 30, 2017) *Memes trump articles on Breitbart's Facebook page*, Columbia Journalism Review. https://www.cjr.org/tow_center/memes-trump-articles-on-breitbart-faces-book-page.php

⁹¹ Matteo Moschella and Ryan Watts, (June 19, 2017) *What we Learned Fact-Checking the UK Election*, First Draft, <https://firstdraftnews.com/joint-venture-learnings/>



Figure 9: An example of a 'meme' shared widely during the UK election

As scholarship of visuals⁹² has shown, the way we understand imagery is fundamentally different to how we understand text.⁹³ Our brains process images at an incredible speed when compared with text.⁹⁴ As a result, our critical reasoning skills are less likely to engage with what we're seeing.

Technology that could identify manipulated or fabricated images lags behind technology for parsing and analysing text. While Google's Reverse Image Search engine (see also TinEye and Yandex) is a good starting point for identifying when images have circulated before, we still don't have publicly available reverse-video search engines or OCR (Optical Character Recognition) tools capable of reading the text on memes in a timely manner. We need more sophisticated, widely accessible tools to help identify problematic visual content.

Over the next few years, we will certainly see the development of artificial intelligence technologies to create as well as identify dis-information. (Simply understood, artificial intelligence is the ability of computers to undertake tasks that we previously needed human brains to work, like speech recognition or visual identification.) It is critical that the engineers

⁹² Sontag S. (1977) *On Photography*, New York: Farrar Straus and Giroux.

⁹³ Postman, N. (1985). *Amusing ourselves to death: Public discourse in the age of television*. New York: Viking.

⁹⁴ A team of neuroscientists from MIT has found that the human brain can process entire images that the eye sees for as little as 13 milliseconds. See: Potter, M. C. (2014). Detecting and remembering briefly presented pictures. In K. Kveraga & M. Bar (Eds.), *Scene Vision* (pp. 177-197). Cambridge, MA: MIT Press

who develop the new products, tools and platforms have been provided ethical training on the unintended consequences of the algorithms they write.

Upon whom are the messages focused?

While agents have particular audiences in mind when they create dis-information, the targeted subject of the message will be different. Dis-information often deliberately highlights differences and divisions, whether they be between supporters of different political parties, nationalities, races, ethnicities, religious groups, socio-economic classes or castes. As Greenhill argues, these types of messages enable discriminatory and inflammatory ideas to enter public discourse and to be treated as fact. Once embedded, such ideas can in turn be used to create scapegoats, normalize prejudices, harden us-versus-them mentalities and, in extreme cases, even catalyze and justify violence.⁹⁵

Most discussion around dis-information in the US and European contexts has focused on political messages, which, while worrying from a democratic perspective, tend not to incite violence. However, in other parts of the world, dis-information directed toward people due to their religious, ethnic or racial identities has led to violence. As Samantha Stanley explained, “perhaps the most obvious example of how mis-information can lead to violent offline action is the two-day riots in Myanmar’s second largest city, Mandalay, in July 2014. Following an unsubstantiated rumor posted on Facebook that a Muslim tea shop owner raped a Buddhist employee, a mob of almost 500 people wreaked havoc on the city and incited lingering fear amongst its Muslim citizens. Two people were killed during the riot, one Buddhist and one Muslim.”⁹⁶

3) Interpreters: How do they make sense of the messages?

As Stuart Hall explained in his seminal work on reception theory⁹⁷, messages are encoded by the producer, but then decoded by individual audience members in one of three ways:

1. **Hegemonic.** Accepting the message as it was encoded.
2. **Negotiated.** Accepting aspects of the message, but not all of it.
3. **Oppositional.** Declining the way the message was encoded.

⁹⁵ Greenhill, K. M. (forthcoming). *Whispers of War, Mongers of Fear: Extra-factual Sources of Threat Conception and Proliferation* and Greenhill, K. M., & Oppenheim B. (forthcoming). *Rumor Has It: The Adoption of Unverified Information in Conflict Zones*. *International Studies Quarterly*.

⁹⁶ Stanley, S. (May 16, 2017) *Mis-information and hate speech in Myanmar*, First Draft, <https://firstdraftnews.com/mis-information-myanmar/>

⁹⁷ Hall, S. (1973). *Encoding and Decoding in the Television Discourse*. Birmingham: Centre for Contemporary Cultural Studies

In this section, we outline the work of key cultural and social theorists who have attempted to explain how audiences make sense of messages.

George Lakoff sees rationality and emotions as being tied together to the extent that, as human beings, we cannot think without emotions. The emotions in our brains are structured around certain metaphors, narratives and frames. They help us make sense of things, and, without them, we would become disoriented. We would not know what or how to think.

Lakoff distinguishes two different kinds of reason: 'False reason' and 'real reason.'⁹⁸ False reason, he says, 'sees reason as fully conscious, as literal, disembodied, yet somehow fitting the world directly, and working not via frame-based, metaphorical, narrative and emotional logic, but via the logic of logicians alone.' Real reason, alternatively, is an unconscious thought that 'arises from embodied metaphors.'⁹⁹ He argues that false reason does not work in contemporary politics, as we've become increasingly emotional about our political affiliations.

Understanding how our brains make sense of language is also relevant here. Every word is neurally connected to a particular frame, which is in turn linked together with other frames in a moral system. These 'moral systems' are subconscious, automatic and acquired through repetition. As the language of conservative morality, for example, is repeated, frames and in turn the conservative system of thought are activated and strengthened unconsciously and automatically. Thus, conservative media and Republican messaging work unconsciously to activate and reinforce the conservative moral system, making it harder for fact-checks to penetrate.¹⁰⁰

Considering Trump's success, D'Ancona recently argued, "He communicated a brutal empathy to [his supporters], rooted not in statistics, empiricism or meticulously acquired information, but an uninhibited talent for rage, impatience and the attribution of blame."¹⁰¹ Ultimately, news consumers "face a tradeoff: they have a private incentive to consume precise and unbiased news, but they also receive psychological utility from confirmatory news."¹⁰²

As we will discuss in Part Two, the emotional allure of situating ourselves within our filter bubbles and having our worldviews supported and reinforced by 'confirmatory news' is

⁹⁸ Lakoff, George (1997). *Metaphors We Live By*. University of Chicago Press and *Moral Politics: What Conservatives Know That Liberals Don't*. University of Chicago Press

⁹⁹ Lakoff, G. (2010) "Why "Rational Reason" Doesn't Work in Contemporary Politics", <http://www.truth-out.org/buzzflash/commentary/george-lakoff-why-rational-reason-doesnt-work-in-contemporary-politics/8893-george-lakoff-why-rational-reason-doesnt-work-in-contemporary-politics>

¹⁰⁰ Lakoff, G. (2010)

¹⁰¹ D'Ancona, M. (2017) *Post-Truth*, Ebury Press.

¹⁰² Allcott, H. and M. Gentzkow, (2017) *Social Media and Fake News in the 2016 Election*, *Journal of Economic Perspectives*, 31:2, p.218

incredibly powerful. Finding solutions to this is going to require a mixture of technological and educational solutions and, ultimately, a psychological shift whereby one-sided media diets are deemed socially unacceptable.

Communication as Ritual

When town criers announced news to crowds, runners read newspapers aloud in coffeehouses and families listened to or watched the evening news together, news consumption was largely a collective experience. However, news consumption has slowly evolved into an individual behavior with the emergence of portable radios and television and, more recently, the ubiquitous adoption of laptops, tablets and smartphones.

But while we might physically consume the news alone, what we choose to consume is increasingly visible because of social media. The posts that we like or comment on and the articles, videos or podcast episodes we share are all public. Borrowing from Erving Goffman's metaphor of life as theatre, invariably, when we use social media to share news, we become performers.¹⁰³ Whatever we like or share is often visible to our network of friends, family and acquaintances, and it affects their perceptions of us.¹⁰⁴

If social media is a stage, our behaviour is a performance and our circle of friends or followers are our audience. Goffman thinks our goal for this performance is to manage our audience's perception of us.¹⁰⁵ Therefore, we tend to like or share things on social media that our friends or followers would expect us to like or share—or, in other words, what we would normally like or share.¹⁰⁶

Similarly, as Maffesoli argued in his 1996 book *The Time of the Tribes*¹⁰⁷, to understand someone's behavior, one must consider the sociological implications of the many different, small and temporary groups that he or she is a member of at any given time of day. Maffesoli's writings aptly describe the realities of users who have to navigate different online groups throughout the day, deciding what information to post or share to different 'tribes' online and off.

¹⁰³ Goffman, E. (1956) *The Presentation of Self in Everyday Life*. Random House.

¹⁰⁴ Karlova, N. A., & Fisher, K. E. (2013). *Plz RT: a social diffusion model of mis-information and dis-information for understanding human information behaviour*. *Information Research*, 18(1), 1-17.

¹⁰⁵ Goffman defines impression management as a conscious or unconscious process in which people try to influence the perceptions of other people about a person, object or event by regulating and controlling information in our daily social interaction.

¹⁰⁶ Picone, I. (2015) *Impression Management in Social Media*, Published Online: 11 FEB 2015
<http://onlinelibrary.wiley.com/doi/10.1002/9781118767771.wbiedcs071/abstract>

¹⁰⁷ Maffesoli, M. (1996) *The time of the tribes*, London:Sage.

This tribal mentality partly explains why many social media users distribute dis-information when they don't necessarily trust the veracity of the information they are sharing: they would like to conform and belong to a group, and they 'perform' accordingly.¹⁰⁸ The pressure to conform can become particularly strong when algorithms on social platforms suppress views opposing those of the user. Even if a user has a politically diverse circle of friends or followers, what she sees in her newsfeed or timeline does not necessarily reflect that diversity.

This connects with the theory of motivated cognition, which refers to the unconscious tendency of individuals to process information to fit conclusions that suit some internal goal. The classic example comes from the 1950s, when psychologists asked students from two Ivy League colleges to watch a film of a football game between their schools that featured a set of controversial officiating calls. The students from each school were more likely to see the referees' calls as correct when it favored their school than when it favored 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.¹⁰⁹

Yale University's Dan Kahan and colleagues demonstrated motivated cognition in a political context. They found that, on issues such as a gun control or climate change, participants would do mathematical somersaults with available data to 'prove' the point of view supported by their own politics.¹¹⁰ Kahan argues that while it's tempting to fixate on the 'lazy brain' theory – that humans rely heavily on mental shortcuts to compensate for the vast amount of information they encounter every day – humans are instead making decisions about what position is most appropriate to publicly support. He concludes: "Work on motivated cognition and political conflict tends to focus more on the need for maintaining a valued identity, particularly as a member of a group... But the seeming inability of economic interests to explain who believes what on issues such as climate change, the HPV vaccine, economic policies that include tax cuts or social welfare spending and the like is in fact the motivation for examining the contribution that identity-protective forms of motivated cognition make."¹¹¹

¹⁰⁸ Social platforms that do not allow anonymity are more prone to this twisted aspect of impression management, whereas on platforms which permit anonymity, other problems such as trolling and harassment can arise.

¹⁰⁹ Kahan, D. (2011) *What is Motivated Reasoning and How Does It Work?* Science and Religion Today, <http://www.scienceandreligiontoday.com/2011/05/04/what-is-motivated-reasoning-and-how-does-it-work/>

¹¹⁰ Kahan, D. et al (2013) *Motivated Numeracy and Enlightened Self-Government*, Behavioural Public Policy, 1, 54-86

¹¹¹ Kahan, D. (2011) *What is Motivated Reasoning and How Does It Work?* Science and Religion Today, <http://www.scienceandreligiontoday.com/2011/05/04/what-is-motivated-reasoning-and-how-does-it-work/>

Communication as Transmission

In a study by Van Damme and Smets in 2014, they remind us that the “human memory is not a recording device, but rather a process of (re)construction that is vulnerable to both internal and external influences.”¹¹²

The challenge for the human brain today is how these influences work in the context of social networks that are bombarding us with information, pinging us repeatedly via the smartphones in our pockets. As WikiMedia testified to the UK Parliament, “Our minds have always been a battleground for various social forces, but the sheer number of agents and institutions vying for control of our thoughts and feelings today is so large that it is confusing and destabilising for many.”¹¹³

Filippo Menczer’s¹¹⁴ most recent research highlights the challenges of our brains to make decisions about credibility when the streams of information are overwhelming. In other words, normal people are too distracted by a deluge of information to find the most accurate stories: “There are a hundred more stories you’re not seeing that are much better than those five that you thought were good.” So, according to this research, irrespective of echo chambers and confirmation bias, people are not sharing verified stories in part because they never see them.

According to research^{115 116} conducted before the heavy use of social media that we take for granted today, people used a set of key heuristics, or mental shortcuts, when evaluating the credibility of a source or message:

1. **Reputation.** Based on recognition and familiarity
2. **Endorsement.** Whether others find it credible
3. **Consistency.** Whether the message is echoed by multiple sites
4. **Expectancy violation.** Whether a website looks and behaves in the expected manner
5. **Self-confirmation.** Whether a message confirms one’s beliefs
6. **Persuasive intent.** The intent of the source in creating the message

¹¹² Van Damme, I. & K. Smets, 2014, p. 310. *The power of emotion versus the power of suggestion: Memory for emotional events in the mis-information paradigm*. *Emotion*. 14 (2): 310

¹¹³ *Evidence Provided to the UK Parliamentary Inquiry on Fake News by WikiMedia UK*, <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/culture-media-and-sport-committee/fake-news/written/48122.html>

¹¹⁴ Qiu, X. et al. (2017) *Limited individual attention and online virality of low-quality information*, *Nature Human Behaviour*, Vol 1.

¹¹⁵ Metzger, M. and A. J. Flanagin (2013) *Credibility and trust of information in online environments: The use of cognitive heuristics*, *Journal of Pragmatics*, 59 pp. 210-220

¹¹⁶ Lewandowsky, S. et al. (2012) *Mis-information and Its Correction: Continued Influence and Successful Debiasing*, *Psychological Science in the Public Interest*, 13(3), pp. 106–131

When we consider these heuristics in the context of our heavy reliance on social media as a source of information, the issues we see in this current age of mis- and dis-information become less surprising.

A very recent meta-analysis¹¹⁷ of the psychological efficacy of messages countering mis-information provides an excellent overview of the research literature pertaining to debunks and how they impact people's perceptions of mis-information. The review of the literature underlined that the effects of a debunking effect were weaker when audiences generated reasons in support of the initial mis-information, supporting what we know about the power of confirmation bias and motivated reasoning.

People are not incentivised to click out of social media to view an article in its original form. As such, the cue of 'expectancy violation' (whether the site behaves as expected) and 'consistency' (whether the information is supported by multiple sites) are unlikely to be utilized.

A most troubling finding from social media studies is how powerful 'familiarity' is as a persuasive factor.¹¹⁸ As Paul and Matthews discuss in their 2016 paper on the methods by which Russia effectively creates a 'firehose of falsehood', repetition is one of the most effective techniques for getting people to accept mal- and dis-information.

The repetition component is particularly problematic on social media due to people trying to manipulate the platforms through bots that automatically "like" or "share" stories or 'click farms'. These techniques can create false sense of popularity about content, and, by tagging influential people like celebrities, politicians or even journalists, impact the news cycle. A disturbing recent report by Trend Micro¹¹⁹ outlines the varied ways that influence is being bought, and the ways in which click farms are being used to boost hashtags, game online petitions, skew online comment and create fake accounts.

Cues like 'endorsement' also become more salient on social media. Our ability to immediately see whether friends and family have liked, shared, commented or retweeted a piece of content becomes a powerful influence on our credibility judgments. As researchers have shown¹²⁰, if you find out your friends like a song, you'll be more likely to like it too. Human

¹¹⁷ Chan, M.S., C. R.Jones, K.H. Jamieson, D. Albarracín (2017) *Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Mis-information*, *Psychological Science*, 1-16.

¹¹⁸ Pennycook, G. et al (July 5, 2017) *Prior Exposure Increases Perceived Accuracy of Fake News*, Available at SSRN: <https://ssrn.com/abstract=2958246>

¹¹⁹ Gu, L., V. Kropotov & F. Yarochkin, (June 2017), *How Propagandists Abuse the Internet and Manipulate the Public*. Trend Micro, https://documents.trendmicro.com/assets/white_papers/wp-fake-news-machine-how-propagandists-abuse-the-internet.pdf

¹²⁰ Salganik, M. et al. (2006) *Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market*, *Science*, Vol. 311, pp.854-856

beings are drawn to follow the masses, particularly when the mass is shown to include your closest friends and family. As Jonathan Stray explains, “messages received in greater volume and from more sources will be more persuasive. Quantity does indeed have a quality all its own... [R]eceiving a message via multiple modes and from multiple sources increases the message’s perceived credibility, especially if a disseminating source is one with which an audience member identifies.”¹²¹

The heuristic of self-confirmation now is also especially powerful, now that social networks are the dominant form of information dissemination. Back in 2006, research by Taber and Lodge¹²² showed the powerful effect of prior attitudes upon reasoning. Attitudinally congruent arguments are evaluated as stronger than attitudinally incongruent arguments. The algorithmic filtering that makes us much less likely to come across information that challenges us (see the section below on filter bubbles and echo chambers) means that the selective exposure that humans tend toward (as it requires less cognitive ‘work’) is done for us automatically.

In addition to self-confirmation bias, humans are also affected by motivated reasoning and a desire to be vindicated. As Sunstein et al.¹²³ found, people who believed in man-made climate change updated their beliefs more in response to bad news (e.g. temperatures are going up more than expected), whereas those who disbelieved man-made climate change were more responsive to good news. Therefore, beliefs were only changed in ways that cemented what they already thought to be true.

This is linked to recent research trying to replicate the so-called backfire effect, which was first proposed in 2010¹²⁴ to account for fact-checks that appeared to harden people’s beliefs about false information. The researchers were unable to replicate the backfire effect and found that corrections and fact-checks do nudge people toward the truth.

Specifically, their research found that, while Trump supporters were more resistant to nudging, they were nudged all the same. And there was another clear pattern of Trump supporters: corrections didn’t change participants’ feelings about Trump. As one of the researchers, Brendan Nyhan, explained, “People were willing to say Trump was wrong, but it

¹²¹ Stray, J. (Feb 27, 2017), *Defense Against the Dark Arts: Networked Propaganda and Counter-Propaganda*, Tow Center for Digital Journalism, Medium. <https://medium.com/tow-center/defense-against-the-dark-arts-networked-propaganda-and-counter-propaganda-deb7145aa76a>

¹²² Taber, C. and M. Lodge, (2006) *Motivated Skepticism in the Evaluation of Political Beliefs*, American Journal of Political Science, Vol. 50, No. 3 (Jul., 2006), pp. 755-769

¹²³ Sunstein, C. R., et al. (2016). *How People Update Beliefs about Climate Change: Good News and Bad News* (SSRN Scholarly Paper No. ID 2821919). Rochester, NY: Social Science Research Network.

¹²⁴ Nyhan, Brendan, and Jason Reifler. 2010. “When Corrections Fail: The persistence of political misperceptions.” *Political Behavior* 32 (2): 303–330.

didn't have much of an effect on what they felt about him."¹²⁵ The takeaway is that, while facts make an impression, they just don't matter for our decision-making—a conclusion that has a great deal of support in the psychological sciences.¹²⁶

The research on how best to word and visualise fact-checks and debunks is varied and at times contradictory.¹²⁷ Much of this research is US-focused, concerned with political fact-checks and mostly carried out on American undergraduate students. It's vital that more studies are replicated in different geographical settings, using mis-information in other areas—particularly health and science.

There is currently a great deal of discussion about increasing funding for individual news literacy programs, as well as integrating core elements into national curricula. We would argue those programs and curricula should include discussions of how to override the human tendency to seek out information that supports our worldview and 'tribal identifications', how to beat confirmation bias and how to be skeptical of information which produces an emotional response.

In this first section, we introduced new conceptual frameworks for discussing and researching information disorder, outlining the three types, elements and phases of information disorder:

- i) The three *types*: mis-, dis- and mal-information
- ii) The three *elements*: agents, messages and interpreters
- iii) The three *phases*: creation, production and dissemination

We need to be much more precise about the definitions we use to describe the phenomenon of information disorder, if we are to begin understanding how and why it is created, the forms that it takes, and its impact. We also need to understand how characteristics change as information flows through the different phases, and how the person who interprets a particular message can become an agent in their own right as they go on to re-share that message with their own networks. In the following section, we discuss the challenges of filter bubbles and echo chambers, underlining the importance of considering how people discover information and share it with their own networks, and the need to study the wider implications for public discourse.

¹²⁵ Resnick, B. (July 10, 2017) "*Trump supporters know Trump lies. They just don't care.*" Vox, <https://www.vox.com/2017/7/10/15928438/fact-checks-political-psychology>

¹²⁶ See this analysis of the 7 most important psychological principles which explain how people make sense of politics today, also by Brian Resnick <https://www.vox.com/science-and-health/2017/3/20/14915076/7-psychological-concepts-explain-trump-politics>

¹²⁷ An excellent overview of some of the literature can be found here: Flynn, D.J. & Nyhan, Brendan & Reifler, Jason. (2017). 'The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics', *Political Psychology*. 38: 127-150

Part 2: Challenges of filter bubbles and echo chambers

‘The “public sphere”: the shared spaces – real, virtual or imagined – whereby social issues are discussed and public opinion is formed’. This theory was first shared by the German sociologist and philosopher Jurgen Habermas, who argued that a healthy public sphere is essential for democracy and must be inclusive, representative and characterised by respect for rational argument.¹²⁸ The most significant challenge to any theory of a shared public sphere is that humans, when we have a choice about who to connect with or not, tend to establish and continue relationships with people who have views similar to our own. We are programmed to enjoy spending time in ‘echo chambers,’ as it requires less cognitive work.

There is no doubt that digital technologies support us in these tendencies. In his 1998 essay, *Which Technology and Which Democracy?*, Benjamin Barber wrote, “Digitalization is, quite literally, a divisive, even polarizing, epistemological strategy... It creates knowledge niches for niche markets and customizes data in ways that can be useful to individuals but does little for common ground... [I]t obstructs the quest for common ground necessary to representative democracy and indispensable to strong democracy.”¹²⁹ At the same time, MIT Media Lab founder Nicholas Negroponte started a discussion about what this very human set of behaviours would look like online. In ‘The Daily Me’, a thought experiment, he considered the implications of completely personalised newspapers. And, in 2006, Habermas acknowledged the challenge for the public sphere in the era of the internet. He argued, “[T]he rise of millions of fragmented chat rooms across the world has tended instead to lead to the fragmentation of large but politically focused mass audiences into a huge number of isolated issue publics.”¹³⁰

These ideas moved into the mainstream in 2011 with the publication of Eli Pariser’s book, the *Filter Bubble*¹³¹. By that point, Negroponte’s thought experiment had become a reality with Facebook’s Newsfeed, and Pariser was able to explain how the social technology companies have engineered personalised experiences. Using algorithms to deliver content that we are mostly likely to enjoy, these platforms reinforce our worldviews and allow us to stay encased in our safe, comfortable echo chambers.

¹²⁸ Habermas, J. (1962) *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*

¹²⁹ Barber, B. (1998) *Which Technology and Which Democracy?*, Talk given at Democracy and Digital Media Conference, <http://web.mit.edu/m-i-t/articles/barber.html>

¹³⁰ Habermas, J. (2006) Speech to the International Communication Association.

¹³¹ Pariser, E. (2011) *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*, Penguin Books.

It is worth recounting here James Carey's description of the 'ritual view of communication', which is not about "the act of imparting information but the representation of shared beliefs." Appreciating this truth helps us explain why echo chambers are so appealing. They provide safe spaces for sharing beliefs and worldviews with others, with little fear of confrontation or division. They allow us to 'perform' our identities as shaped by our worldviews with others who share those worldviews. This behaviour is not new, but the platforms have capitalized on these human tendencies, knowing they would encourage users to spend more time on their sites.

Agents who are creating dis-information understand that, when people consume and share these messages, they will be doing so increasingly from inside these echo chambers, with no one to challenge the ideas. This means the people who will interpret their messages are much less likely to have an 'oppositional' (rejecting the way the message was encoded) or 'negotiated' (accepting only some aspects of the message) reading. As such, agents target groups that they know are more likely to be receptive to the message. If they are successful in doing that, it is very likely the message will then be shared by the initial recipient. And, as research shows, we are much more likely to trust a message coming from someone we know¹³², even if we suspect it to be false. This is why dis-information can be disseminated so quickly. It is travelling between peer-to-peer networks where trust tends to be high.¹³³

The fundamental problem is that "filter bubbles" worsen polarization by allowing us to live in our own online echo chambers and leaving us with only opinions that validate, rather than challenge, our own ideas. While confirmation bias occurs offline and the term 'selective exposure' has been used by social scientists for decades to describe how information seekers use only certain sources that share their views,¹³⁴ social media are designed to take advantage of this innate bias.

The rise in popularity of social networks as sources of news has taken place at the same time as a decline in local newspapers in some of the largest democracies in the world. In the U.S., Canada and the U.K. particularly, the local news ecosystem is struggling as the advertising model for news has collapsed. Many local newsrooms have been forced to make serious staff cuts, consolidate or close. In the U.K., there has been a net loss of approximately two hundred local newspaper titles since 2005¹³⁵ In Canada a study commissioned by Friends of Canadian

¹³² Metzger et al. (2010) *Social and Heuristic Approaches to Credibility Evaluation Online*, Journal of Communication, 60 (3):413-439

¹³³ Granovetter, M.S. (1973) *The Strength of Weak Ties*, Journal of Sociology, 78(6):1360-1380

¹³⁴ Prior, M. (2003). *Any good news in soft news? The impact of soft news preference on political knowledge*. Political Communication, 20(2), pp.149-171.

¹³⁵ Pondsford, D. (March, 31, 2017) *The decline of local journalism is a far greater threat to media plurality than Rupert Murdoch*, The Press Gazette, <http://www.pressgazette.co.uk/the-decline-of-local-journalism-is-a-far-greater-threat-to-media-plurality-than-rupert-murdoch/>

Broadcasting warned the Canadian Radio-television and Telecommunications Commission that “without intervention, half of Canada’s small and medium-market television stations could disappear by 2020.”¹³⁶ And in a Columbia Journalism Review article the scale of US News deserts was spelled out very clearly, showing how many cities have been left with just one local newspaper, and how many have none at all.¹³⁷ As ad revenue moves to Google and Facebook (in the mobile ad market these two companies earn early half of all revenue¹³⁸), it is expected that the same patterns witnessed in these countries will be felt in many other countries in the next few years.

In 2009, the U.S.-based Knight Commission on the Information Needs of Communities in a Democracy concluded that information is “as vital to the healthy functioning of communities as clean air, safe streets, good schools, and public health.”¹³⁹ While there no evidence yet to directly connect the decline of local news media to the rise in information pollution, when strong local media do not exist other sources will fill that vacuum.

As Nina Jankowicz, a fellow at the Woodrow Wilson Center’s Kennan Institute wrote in an opinion piece in the New York Times recently:

Without news that connects people to their town councils or county fair, or stories that analyze how federal policies affect local businesses, people are left with news about big banks in New York and dirty politics in Washington.... Readers compare this coverage with their dwindling bank balances and crumbling infrastructure and feel disconnected and disenfranchised, and latch onto something — anything — that speaks to them. That might be President Trump’s tweets. Or dubious “news” from an extreme right- or left-wing site might ring true. Or they might turn to Russian disinformation, which exploits this trust gap.¹⁴⁰

Local news outlets provide a shared experience for a community. When communities rely on individual feeds of news from their social networks, these shared experiences disappear. We need more research to understand the implications of this in terms of people sharing mis- and dis-information.

¹³⁶ Lindgren, A., J. Corbett & J. Hodson, (Jan. 23, 2017) *Canada’s Local News Poverty, Policy Options*, <http://policyoptions.irpp.org/magazines/january-2017/canadas-local-news-poverty/>

¹³⁷ Bucay, Y., V. Elliott, J. Kamin, A. Park, *America’s Growing News Deserts*, Columbia Journalism Review, https://www.cjr.org/local_news/american-news-deserts-donuts-local.php

¹³⁸ Thompson, D. (Nov. 3, 2016) *The Print Apocalypse and How to Survive It*, *The Atlantic*, <https://www.theatlantic.com/business/archive/2016/11/the-print-apocalypse-and-how-to-survive-it/506429/>

¹³⁹ Knight Commission on the Information Needs of Communities in a Democracy, (2009) *Informing Communities: Sustaining Democracy in the Digital Age*, The Aspen Institute, <https://production.aspeninstitute.org/publications/informing-communities-sustaining-democracy-digital-age/>

¹⁴⁰ Jankowicz, N (Sept. 25, 2017) *The Only Way to Defend Against Russia’s Information War*, The New York Times, <https://mobile.nytimes.com/2017/09/25/opinion/the-only-way-to-defend-against-russias-information-war.html>

As discussed, the technology companies are commercial entities, and therefore to keep their shareholders happy need to encourage users to stay on their site for as long as possible to maximize the number of exposures to advertisements. They do so by tweaking the algorithms to deliver more of what users have liked, shared or commented on in the past. So, while we have seen the technology companies take some steps to fight dis-information on their platforms (see Part 3), ultimately, it's difficult to imagine them making substantive changes to their algorithms to pop these filter bubbles. If the platforms changed the algorithm to provide us with more challenging material that forced us to reconsider some of our established world views, we are unlikely to spend as much time on them.

As *Wired* dramatically concluded in an article just after the US election, "The global village that was once the internet has been replaced by digital islands of isolation that are drifting further apart each day. From your Facebook feed to your Google Search, as your experience online grows increasingly personalized, the internet's islands grow farther apart."¹⁴¹ The Laboratory for Social Machines at MIT has been investigating filter bubbles, how they form, and how people can try to break out of them. In research published in December 2016, tweets sent during the US election were visualized in a network that showed almost no overlap between Trump and Clinton supporters.¹⁴² The analysis concluded that, on Twitter, Trump supporters formed a particularly insular group when talking about politics during the general election, having few connections to Clinton supporters or the mainstream media.

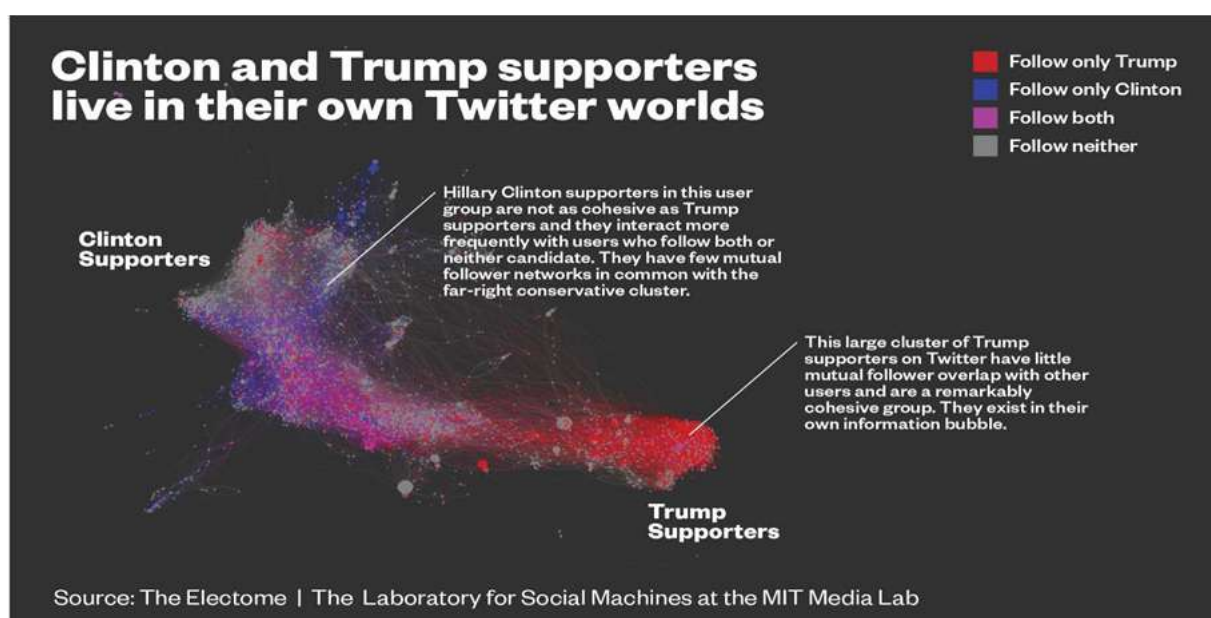


Figure 10: Visualization by the Laboratory for Social Machines at MIT of Donald Trump and Hillary

¹⁴¹ El-Bermawy, M. (Nov. 18, 2016) *Your Filter Bubble is Destroying Democracy*, *Wired*. <https://www.wired.com/2016/11/filter-bubble-destroying-democracy/>

¹⁴² Thompson, A. (Dec. 8, 2016) *Parallel Narratives, Clinton and Trump Supporters Really Don't Listen to One Another*, *Vice*. <https://news.vice.com/story/journalists-and-trump-voters-live-in-separate-online-bubbles-mit-analysis-shows>

Clinton supporters on Twitter

Research by Demos in the U.K. that analysed British Twitter accounts found similar patterns between supporters of different political parties. However, they were able to show that people with more extreme political views tended to engage with a smaller number of people than those who had more moderate political views.¹⁴³

Walter Quattrociocchi and his team have studied the dynamics of echo chambers on Facebook.¹⁴⁴ Examining the posts of 1.2 million users, his team analysed mainstream and conspiracy science news and how they are consumed and shaped by communities on Facebook. They found that polarized communities emerge around distinct types of content, and consumers of conspiracy news tended to be extremely focused on specific topics like climate change.

There has also been research that challenges these ideas about the dangers of echo chambers. A survey of 14,000 people in seven countries that was published in May 2017¹⁴⁵ concluded that “people who are interested and involved in politics online are more likely to double-check questionable information they find on the internet and social media, including by searching online for additional sources in ways that will pop filter bubbles and break out of echo chambers.”¹⁴⁶ In addition, the 2017 Digital News Report, published by the Reuters Institute for the Study of Journalism, concluded, “Echo chambers and filter bubbles are undoubtedly real for some, but we also find that – on average – users of social media, aggregators and search engines experience more diversity than non-users.”¹⁴⁷

Concerns raised since Brexit and the US election have led to new innovations by social platforms, third-party organizations and academic institutions to help people ‘prick’ their filter bubbles. With the renewed emphasis on scaling news literacy programs globally, teaching how social algorithms produce these filter bubbles should be a crucial part of any standardised curriculum.

¹⁴³ Krasodomski-Jones, A. (2016) *Talking To Ourselves? Political Debate Online and the Echo Chamber Effect*. Demos. Available at: <https://www.demos.co.uk/project/talking-to-ourselves/>

¹⁴⁴ Quattrociocchi, W. (14 Jan, 2016) *How does mis-information spread online?* World Economic Forum Blog, <https://www.weforum.org/agenda/2016/01/q-a-walter-quattrociocchi-digital-wildfires/>

¹⁴⁵ Dutton, William H., et al. (2017) "Search and Politics: A Cross-National Survey." Quello Center Working Paper No. 2944191 Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2944191

¹⁴⁶ Dutton, B. (May 2017) *Fake news, echo chambers and filter bubbles are an exaggerated threat. Here's why*. World Economic Forum Blog. <https://www.weforum.org/agenda/2017/05/fake-news-echo-chambers-and-filter-bubbles-are-an-exaggerated-threat-heres-why>

¹⁴⁷ Newman, N. (2017) *Overview and Key Findings: Digital News Report*, Reuters Institute for the Study of Journalism. <http://www.digitalnewsreport.org/survey/2017/overview-key-findings-2017/>

Facebook, for example, rolled out a new related-articles feature¹⁴⁸ that is designed to show multiple perspectives on a story. And, during the U.K. and French election, Facebook rolled out the Perspectives¹⁴⁹ feature, which allowed users to compare candidates or parties positions after clicking on an election-related article.

A third-party effort to prick the filter bubble is PolitEcho¹⁵⁰, a Chrome extension that allows you to plot your Facebook friends on a graph based on their political affiliation, which is estimated from the news organizations they have liked. Another is Flipfeed¹⁵¹, which allows you to randomly see the Twitter feed of someone with a diametrically oppositional view to your own. As the app claims, watching the 'Trump news conference', in the 'flipped' mode can be eye-opening. And yet another Chrome extension, Rbutr¹⁵², is a community-driven app which connects webpages on the basis that one page argues against the other. If you have downloaded the extension and you visit a 'rebutted' page, you will be told there are rebuttals to the particular page and it will link out to the rebutting articles.

¹⁴⁸ Constine, J. (2017) *Facebook shows Related Articles and fact checkers before you open links*, TechCrunch, <https://techcrunch.com/2017/04/25/facebook-shows-related-articles-and-fact-checkers-before-you-open-links/>

¹⁴⁹ McGregor, J. (May 31, 2017) *Facebook Wades Into Another Election*, Forbes. <https://www.forbes.com/sites/jaymcgregor/2017/05/31/facebook-wades-into-another-election/#37265336635c>

¹⁵⁰ <http://politecho.org/>

¹⁵¹ <https://flipfeed.media.mit.edu/>

¹⁵² <http://rbutr.com/#>

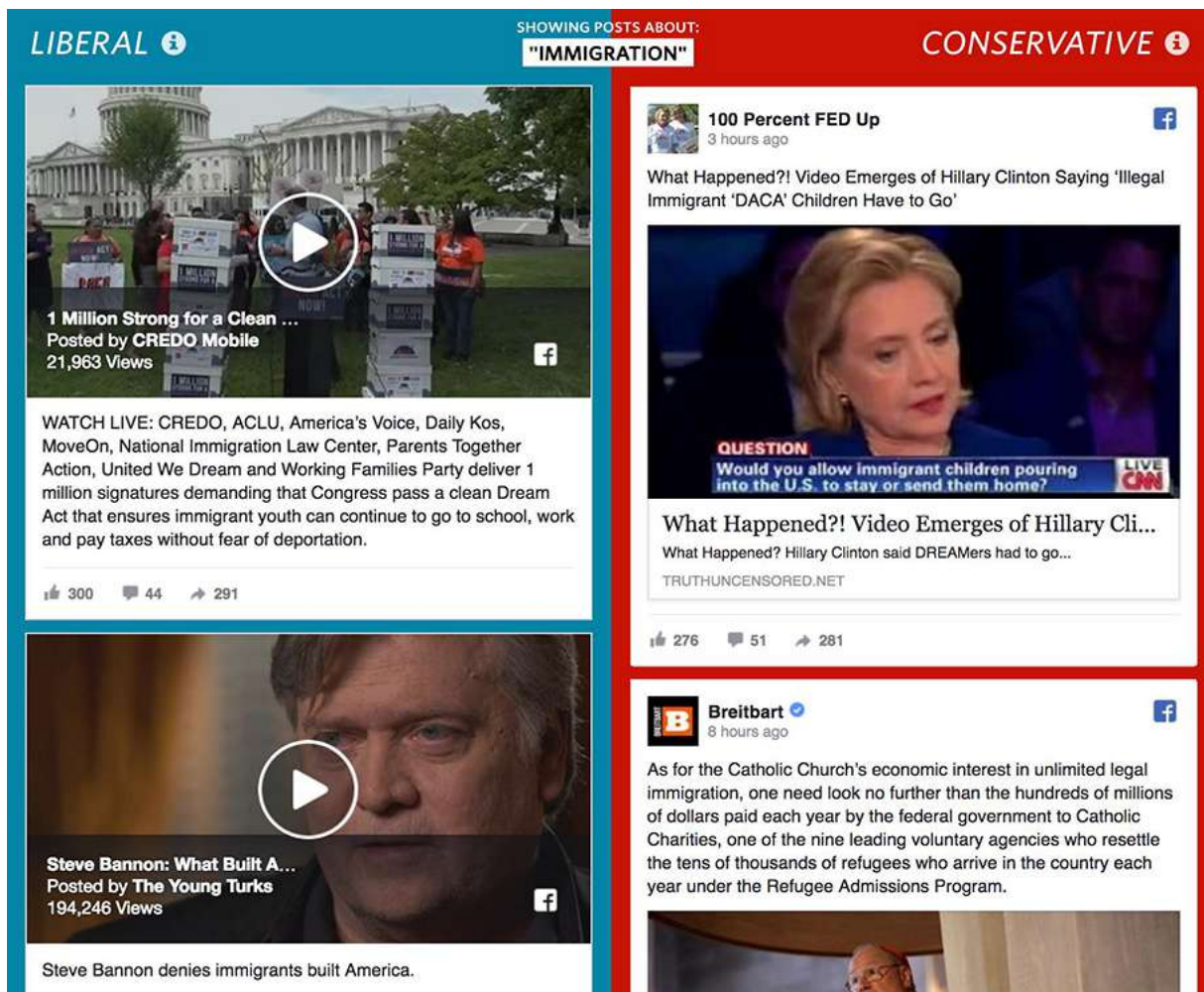


Figure 11: Screenshot of the Blue Feed/Red Feed streams on the topic of immigration (<http://graphics.wsj.com/blue-feed-red-feed/#/immigration>)

The Wall Street Journal was one of the first news outlets to try to find a way for their audiences and other people interested in this concept to compare the way different topics—like Trump, guns, health care, and immigration—were being covered and shared by people from different political perspectives. They created 'Blue Feed, Red Feed'¹⁵³ "to demonstrate how reality may differ for different Facebook users." If a source appears in the red feed, a majority of the articles shared from the source were classified as 'very conservatively aligned' in a large 2015 Facebook study. For the blue feed, a majority of each source's articles aligned 'very liberal.' These feeds aren't intended to resemble actual individual news feeds. Instead, they are rare, side-by-side looks at real conversations from different perspectives.

Buzzfeed and the *Guardian* newspaper have been testing new features that help readers navigate alternative viewpoints. *BuzzFeed's* 'Outside Your Bubble'¹⁵⁴ pulls in opinions from

¹⁵³ <http://graphics.wsj.com/blue-feed-red-feed/>

¹⁵⁴ https://www.buzzfeed.com/bensmith/helping-you-see-outside-your-bubble?utm_term=.lhjz5lx2#.yqm1R3wE5

across the web and gives them a neutral platform. Public comments, which can be the most emotional and combative part of any online story, are then removed from their normal context and rephrased as dispassionate bullet points.

The Guardian’s weekly column ‘Burst Your Bubble’¹⁵⁵ curates “five conservative articles worth reading” for the site’s liberal audience. Similarly, every week, the *Washington Post* journalist Will Sommer publishes a newsletter, ‘Right Richter’,¹⁵⁶ which aggregates right-wing perspectives for left-leaning audiences.

Finally, there is the AllSides site¹⁵⁷, whose stated mission is to expose bias and provide “multiple angles on the same story so you can quickly get the full picture, not just one slant.” Using a mixture of crowd-driven ratings and its patented algorithms, it’s the newest attempt to provide audiences with a visual guide to politically slanted journalism.

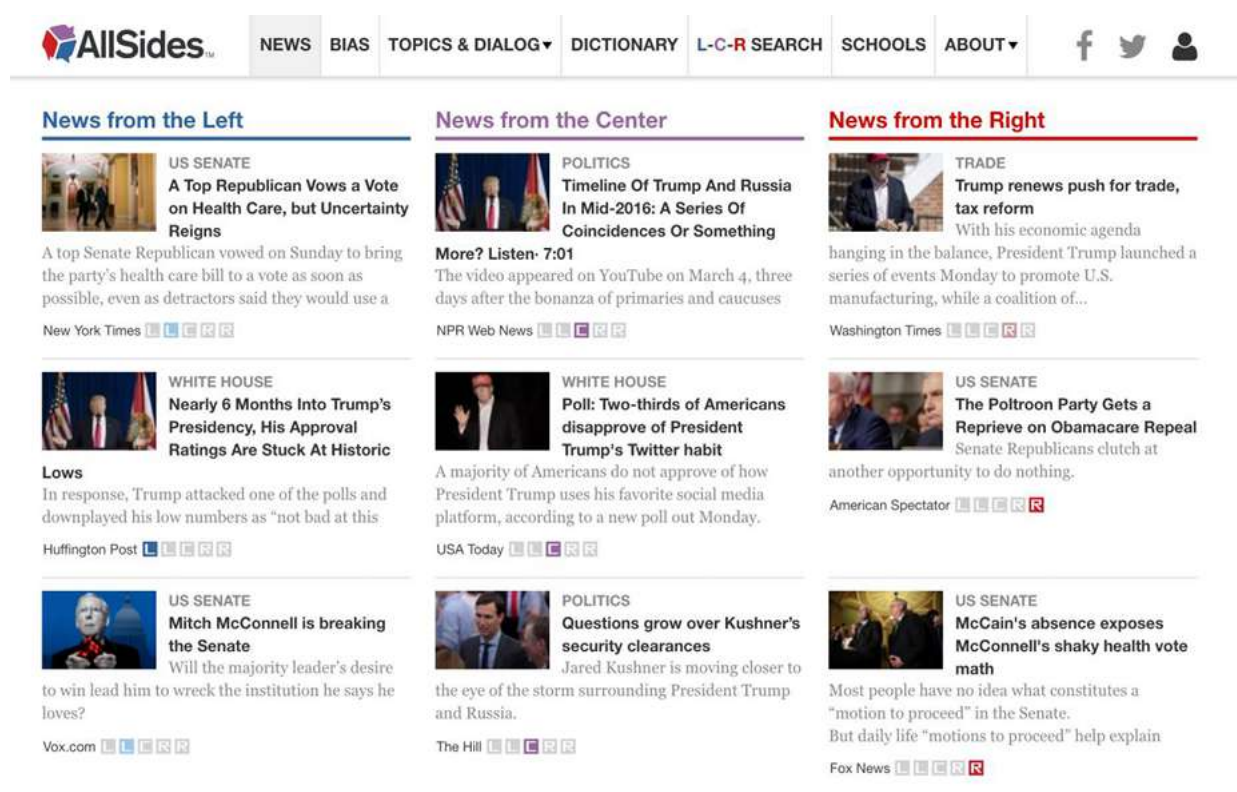


Figure 12: Screenshot of the Allsides.com website

The ultimate challenge of filter bubbles is re-training our brains to seek out alternative viewpoints. Some people have likened our informational diet to our nutritional one, claiming that in the same way as we had to be educated to see the value of a diet rich in fruit and

¹⁵⁵ <https://www.theguardian.com/us-news/series/burst-your-bubble>

¹⁵⁶ <http://tinyletter.com/rightrichter>

¹⁵⁷ <http://allsides.com/>

vegetables, we need to give ‘nutritional ‘labels to information so that people understand the value of a media diet with a variety of political viewpoints. While we can pressure the social networks to diversify our diet, we can’t force people to actually click, let alone read the content. At public events, Facebook has admitted that when they have attempted to deliver more content from an opposite view, people tend not to click on it.

As this report underlines, we have to think about information consumption from a ritual as well as transmission perspective. If we recognise that people seek out and consume content for many reasons beyond simply becoming informed – like feeling connected to similar people or affiliating with a specific identify – it means that pricking the filter bubbler requires more than simply providing diverse information.

Part 3: Attempts at solutions

One week after the election, Eli Pariser, author of *The Filter Bubble*, created a public Google document and asked people to contribute solutions to solve the mis-information problem. Within a few weeks, the document was more than 150 pages long and included comments from over 50 people. The document¹⁵⁸ includes many ideas and can be seen as a comprehensive blueprint of what solutions are possible.

One point we'd like to stress, however, is that much of the debate about solutions presupposes communication as information transmission. But this cannot explain or solve the problem of information disorder. As Carey suggests, "under a ritual view [of communication] news is not information but drama"¹⁵⁹ and "a portrayal of the contending forces in the world."¹⁶⁰ Our conversations about solutions will need to evolve in order to recognise this role information plays beyond simply transmitting messages.

Over the past twelve months, potential solutions have been discussed endlessly at conferences and in workshops, but we've seen little concrete changes from the platforms. While there is certainly more foundation money than there was,¹⁶¹ and a myriad of small projects are underway, the grand ideas are yet to be implemented. These include Apple's CEO Tim Cook call for a Public Service Announcement about dis-information, new labels to identify different types of content on social platforms, systematic programs for taking down bot accounts, the integration of critical media literacy programs in schools and best practices for making fact-checks and debunks shareable.

Facebook and Google have announced methods for preventing fabricated sites from making money through their advertising platforms. However, anecdotally, 'fake news' creators have explained that, while they experienced short term losses in revenue earlier in the year, they have returned their profits to previous levels using other ad networks that are willing to partner with them.

The only real major development we've seen has been the passing of legislation in Germany that fines platforms for hosting unlawful content, including defamation and incitement to

¹⁵⁸ Pariser, E. (2016) *Media ReDesign: The New Realities*,
https://docs.google.com/document/d/1OPghC4ra6QLhaHhW8QvPJRMKGEXT7KaZtG_7s5-UQrw/edit#heading=h.l4uvrs8m75xh

¹⁵⁹ Carey, J. (1989), p.17

¹⁶⁰ Carey, J. (1989), p.16

¹⁶¹ Yurieff, K. (April 5, 2017) *eBay founder commits \$100 million to combat 'fake news'*, CNN,
<http://money.cnn.com/2017/04/05/technology/pierre-omidyar-donation-fake-news/index.html>

hatred, and not removing such posts within 24 hours.¹⁶² And we're seeing serious discussions in Singapore about passing a similar law.¹⁶³

In this next section, we discuss potential solutions from a number of different perspectives—technological, social, media-centric, educational and regulatory.

Technological Approaches

Mis-, dis- and mal-information are incredibly complex phenomena, but the impression that these problems emerged suddenly during the US election encouraged many to believe that a solution could be found just as quickly. While changing underlying socio-economic and cultural factors takes time, the allure of an easy algorithmic tweak meant it became the popular solution. Certainly, in the crowd-sourced 'Design Solutions' Google document started by Eli Pariser, discussions about technological solutions dominate.¹⁶⁴ Even Krishna Bharat, the engineer responsible for building Google News, stepped in and wrote a detailed piece about the technical ways in which the platforms could detect mis- and dis-information in real time.¹⁶⁵

What have the social networks done?

As we have seen, one of the primary motivations for creating dis-information is financial gain. Google has therefore worked to prevent revenue flowing to the owners of “bad sites, scams and ads” and has permanently banned nearly 200 publishers from its AdSense advertising network as of late 2016.¹⁶⁶ Facebook made similar moves, updating its policies with language stating they would not display ads that show misleading or illegal content. Facebook has also taken steps to tackle ‘imposter content’, stating, “On the buying side, we’ve taken action against the ability to spoof domains, which will reduce the prevalence of sites that pretend to be real publications.”¹⁶⁷ And, in late August 2017, Facebook announced they would block ads from pages that had repeatedly shared false news, stating “Currently, we do not allow

¹⁶² Eddy, M. and M. Scott (June 30 2017), *Delete Hate Speech or Pay Up, Germany Tells Social Media Companies*, New York Times, <https://www.nytimes.com/2017/06/30/business/germany-facebook-google-twitter.html>

¹⁶³ Yi, S.B. (June 19, 2017) *New legislation to combat fake news likely to be introduced next year: Shanmugam*, Straits Times, <http://www.straitstimes.com/singapore/new-legislation-to-combat-fake-news-next-year-shanmugam>

¹⁶⁴ Eli Pariser (document creator) (2016) *Media ReDesign: The New Realities*, https://docs.google.com/document/d/1OPghC4ra6QLhaHhW8QvPJRmKGEXT7KaZtG_7s5-UQrw/edit

¹⁶⁵ Bharat, K. (April 27, 2017) *How to Detect Fake News in Real-Time*, NewsCo on Medium, <https://shift.newco.co/how-to-detect-fake-news-in-real-time-9fdae0197bfd>

¹⁶⁶ Spencer, S. (Jan. 25, 2017) *How we fought bad ads, sites and scammers in 2016*, Google Blog, <https://www.blog.google/topics/ads/how-we-fought-bad-ads-sites-and-scammers-2016>

¹⁶⁷ Written evidence by Facebook to the UK Parliamentary Inquiry on Fake News, <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/culture-media-and-sport-committee/fake-news/written/49394.html>

advertisers to run ads that link to stories that have been marked false by third-party fact-checking organizations. Now we are taking an additional step. If Pages repeatedly share stories marked as false, these repeat offenders will no longer be allowed to advertise on Facebook.”¹⁶⁸

Google News recently took steps to allow publishers to highlight fact-checked content for programmatic detection using schema.org, a structured data markup schema supported by the major search engines. This feature first appeared in the U.K. and the US last October, and has since been added to Google News in Germany, France, Brazil, Mexico and Argentina.

Google News Lab, which is distinct from Google News and whose mission is to collaborate with journalists and entrepreneurs to help build the future of media, has been very active in this space over the past couple of years.¹⁶⁹ For example, the News Lab was a founding partner of First Draft when it began in June 2015.

Facebook moved, from Mark Zuckerberg’s post-election denial¹⁷⁰ that ‘fake news’ was a problem on his platform, to rolling out a third-party fact-checking initiative on December 15, 2016 that includes the International Fact Checking Network, *The Associated Press*, *The Washington Post* and *Snopes*.¹⁷¹ They expanded the project to France and Germany in February and the Netherlands in March. In this initiative, users flag posts they think might be ‘false news’, populating a queue that the affiliated fact-checking organizations can see. After an article has been fact-checked, any user who sees that content will see that it has been disputed by one of the fact-checking organizations. If someone tries to share a disputed article, they are reminded with a pop-up notice that the content is in dispute.

The initiative was widely welcomed when launched, although there were dissenting voices, such as Data & Society’s Robyn Caplan who raised concerns about the challenge of writing algorithms to identify this type of content when the definitions are so broad. She also spoke out about the need to financially support this outsourced journalism.¹⁷²

¹⁶⁸ Shukla, S. (August 28, 2017) *Blocking Ads from Pages that Repeatedly Share False News*. Facebook Newsroom. <https://newsroom.fb.com/news/2017/08/blocking-ads-from-pages-that-repeatedly-share-false-news/>

¹⁶⁹ Google News Lab funds First Draft, and has financially supported First Draft initiatives around the French and UK elections, as well as training in Germany, Hong Kong and Korea.

¹⁷⁰ Shahari, A. (Nov 11. 2017) *Zuckerberg Denies Facebook’s Impact on the Election*, <http://www.npr.org/sections/alltechconsidered/2016/11/11/501743684/zuckerberg-denies-fake-news-on-facebook-had-impact-on-the-election>

¹⁷¹ Mosseri, A. (2016) *News Feed FYI: Addressing Hoaxes and Fake News*, Facebook Newsroom, December 15, 2016.

¹⁷² Caplan, R. “How Do You Deal with a Problem Like ‘fake News?’” *Data & Society: Points*, January 5, 2017. <https://points.datasociety.net/how-do-you-deal-with-a-problem-like-fake-news-80f9987988a9>

While there is an increasing body of evidence suggesting that many types of ‘nudge’ technologies can slow sharing,¹⁷³ and researchers, in experimental settings, have found that warning labels and pop-up boxes slow down the sharing of this content¹⁷⁴ without access to the results of this Facebook initiative we have no independent way of knowing whether the program is slowing down the spread of polluted information on the platform. Despite repeated calls for access to this data, most powerfully from the fact-checkers themselves, to date Facebook have refused to share the numbers. It’s therefore impossible to assess the success of the project.

Experimental research by Leticia Bode in 2015 suggested that when a Facebook post that includes mis-information is immediately contextualised in their ‘related stories’ feature underneath, misperceptions are significantly reduced.¹⁷⁵ In August 2017, Facebook announced¹⁷⁶ that they would be rolling out their related stories feature more widely to help contextualize mis-information with fact-checked articles.

In January 2017, Facebook launched its ‘Facebook Journalism Project’¹⁷⁷ and announced that news literacy would be a priority for the company. Beyond financially supporting non-profits working in this space, they also rolled out a Public Service Announcement-type message at the top of the New Feed in fourteen countries. This message linked to a post with 10 top tips for spotting ‘false news’.¹⁷⁸ These same tips were published as full page ads in newspapers in Germany, France and the U.K. They also committed a \$14 million fund to help establish the News Integrity Initiative based at the CUNY Journalism School in New York, which supports the development of tools, research and media literacy programs globally.¹⁷⁹

However, in a post titled ‘Facebook Must Either Innovate or Admit Defeat At The Hands Of Fake News Hoaxsters’, Craig Silverman had some strong words for the platform:

¹⁷³ Bilton, R. (Feb.2, 2017) Reddit’s /r/worldnews community used a series of nudges to push users to fact-check suspicious news, Nieman Lab, <http://www.niemanlab.org/2017/02/reddits-rworldnews-community-used-a-series-of-nudges-to-push-users-to-fact-check-suspicious-news/>

¹⁷⁴ Pennycook, G. et al. (2017) Unpublished research presented at the Harvard University event, Combating Fake News: An agenda for Research and Action.

¹⁷⁵ Bode, L. & Vraga, E (2015) *In Related News, That Was Wrong: The Correction of Mis-information Through Related Stories Functionality in Social Media*, Journal of Communication, 65 (4): 619-638.

¹⁷⁶ Su, S. (April 25, 2017) *News Feed FYI: New Test With Related Articles*, <https://newsroom.fb.com/news/2017/04/news-feed-fyi-new-test-with-related-articles/>

¹⁷⁷ Simon, F. (Jan. 11, 2017) *Introducing: Facebook Journalism Project*. Facebook Media Blog. <https://media.fb.com/2017/01/11/facebook-journalism-project/>

¹⁷⁸ Mullin, B. (April 6, 2016) *Facebook rolls out global warning against fake news*, Poynter <http://www.poynter.org/2017/facebook-rolls-out-global-warning-against-fake-news/454951/>

¹⁷⁹ Brown, C. (April 2, 2017) *Introducing the News Integrity Initiative*, Facebook Media Blog, <https://media.fb.com/2017/04/02/introducing-the-news-integrity-initiative/>

[If Facebook] “is truly committed to offering a quality Trending Topic (and News Feed) experience, its only option is to make massive strides in the detection and analysis of the factual qualities of news articles. Developing what would likely be the world’s first algorithm to do this job with accuracy and consistency will require significant engineering resources. But it’s what is necessary to actually stop Facebook from being the world’s biggest platform for false and fake news without the use of editors. Right now Facebook has no editors, a flawed algorithm, and a weak product.”¹⁸⁰

Facebook closed 30,000 automated accounts in France ahead of the election and ‘tens of thousands’ in the U.K. ahead of their election. To date, these kinds of moves by Facebook have timed with elections, but there are calls for Facebook to take an ongoing approach to policing automation on the platform.¹⁸¹

As discussed, the challenge of bots disseminating dis-information is a significant one. Ongoing research by the Computational Propaganda Project¹⁸² at the Oxford Internet Institute continues to identify the scale of the problem globally. In April 2017, Facebook’s Security team published a paper on ‘Information Operations’, defining it as “actions taken by organized actors (governments or non-state actors) to distort domestic or foreign political sentiment, most frequently to achieve a strategic and/or geopolitical outcome.”¹⁸³ It was the first-time Facebook had admitted the scale of the problem they face in terms of official, tightly organised, networked agents using their platform to disseminate automated dis-information.

In September 2017, Facebook admitted that they had found evidence that ‘dark ads’ (ads that are only visible to the intended audience, rather than publicly viewable on a page) had been purchased by a Russian organization and directed at US citizens. Facebook explained, “[T]he ads and accounts appeared to focus on amplifying divisive social and political messages across the ideological spectrum — touching on topics from LGBT matters to race issues to immigration to gun rights.”¹⁸⁴ A few days later, an investigation by the Daily Beast found

¹⁸⁰ Silverman, C. (April, 30, 2017) *Facebook Must Either Innovate Or Admit Defeat At The Hands Of Fake News Hoaxsters*, BuzzFeed, https://www.buzzfeed.com/craigsilverman/how-facebook-fell-into-a-fake-news-trap-of-its-own-making?utm_term=.uqGPIGAKp#.km8Vb7AmX

¹⁸¹ Hofileña, C. F. (Oct. 9, 2016) *Fake accounts, manufactured reality on social media*, Rappler, <https://www.rappler.com/newsbreak/investigative/148347-fake-accounts-manufactured-reality-social-media>

¹⁸² <http://comprop.oii.ox.ac.uk/>

¹⁸³ Weedon, J., W. Nuland & A. Stamos (April 27, 2017) *Information Operations and Facebook*, p. 4 <https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf>

¹⁸⁴ Stamos, A. (September 6, 2017) *An Update On Information Operations On Facebook*, Facebook Newsroom, <https://newsroom.fb.com/news/2017/09/information-operations-update/>

inauthentic accounts, seemingly located in Russia, had used the Facebook events function to organize anti-immigration protests in the US.¹⁸⁵

While Facebook is struggling on this topic, Twitter's open APIs mean it is much easier for academics and think tanks to visualise the bot networks that exist on their platform. Thus, calls for action have been focused on Twitter. In a June 2017 blog post, Twitter explained its efforts to fight against bots: "We're working hard to detect spammy behaviors at source, such as the mass distribution of Tweets or attempts to manipulate trending topics. We also reduce the visibility of potentially spammy Tweets or accounts while we investigate whether a policy violation has occurred. When we do detect duplicative, or suspicious activity, we suspend accounts. We also frequently take action against applications that abuse the public API to automate activity on Twitter¹⁸⁶, stopping potentially manipulative bots at the source."¹⁸⁷

Ultimately, the question remains as to whether the social networks and technology companies, as commercial entities, can ever spearhead serious solutions to the problem of information disorder. As Martin Moore, Director of the Center for the Study of Media, Communication and Power argued in his submission of evidence to the U.K. Parliamentary Select Committee on Fake News:

"Remedies solely or heavily based on technological fixes or market-driven corrections will not, on their own, address these problems. Technology should be able to reduce the spread of certain types of news (such as that which is shared without being read first), and to show where news is disputed. However, the long history of fake news, the political, social and economic motivations for producing it, and the ease of self-publishing online, mean that technology will only ever partly address the problem. It also elicits dangers of its own with regard to the value-driven choices that engineers will have to make when determining which news to promote and which to suppress. Nor are market-driven corrections likely to solve, or even alleviate, the problem. The technology platforms on which this news travels are reliant on advertising that prioritises popular and engaging content that is shared widely. The content is not distinguished by its trustworthiness, authority or public interest, since these are not criteria that drive likes and shares."

Emily Bell, director of the Tow Center for Digital Journalism argues, "The business of publishing and monetizing information is never neutral; it is always deeply political. It shapes

¹⁸⁵ Collins, B. et al (September 11, 2017) *Exclusive: Russia Used Facebook Events to Organize Anti-Immigrant Rallies on U.S. Soil*, The Daily Beast, <http://www.thedailybeast.com/exclusive-russia-used-facebook-events-to-organize-anti-immigrant-rallies-on-us-soil>

¹⁸⁶ Twitter, (April 6, 2016) *Automation Rules*, <https://support.twitter.com/articles/76915>

¹⁸⁷ Crowell, C., (June 14, 2017) *Our Approach to Bots and Automation*, https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Mis-information.html

opinion, informs markets, reinforces biases, creates understanding, and spreads confusion. Facebook has said more than once that it does not want to be an arbiter of the truth, but it also does not want to be the purveyor of lies. Journalists have known for a long time what technology companies are just finding out: what you don't publish is as brand-defining as what you do."¹⁸⁸

In addition to these steps to shut down automated accounts by the social networks, the ability to identify and discredit sources of dis-information in real time is an increasingly necessary skill for newsrooms to master. We hope technology companies will support newsrooms in being able to identify the agents of a post. If newsrooms were provided with the tools to identify whether the agents of a piece of dis-information are official, organised and automated, they would be able to quickly raise red flags for audiences.

When assessing moves by the technology companies over the past year, one of the most frustrating elements has been the failure to connect with the research, education, library, civil society, and policy communities at any substantive level. There are decades of research on mis-information, how people 'read' and make sense of information, the factors that slow down or exacerbate rumours. But the responses have often felt knee-jerk and atheoretical, and at times public relations moves rather than serious attempts to tackle the complexity of the problem. On this topic, when the scale and seriousness require sophisticated responses, the technology companies must work more closely with those who have research expertise on this subject, as well as those working on the ground around the world, and experience first-hand the real-world repercussions of information pollution.

Blacklisting, flagging and credibility scores

Creating lists of problematic sites was an early and popular suggestion, but as US academic Melissa Zimdars found to her detriment¹⁸⁹, attempting to be the 'arbiter of truth' can cause you to become incredibly unpopular. Notably, her list, now hosted at opensources.co, has been used by a number of technologists building tools to help 'flag' problematic content through browser extensions such as Check This¹⁹⁰. One such tool from the French daily newspaper *Le Monde* is based on a database of sites reviewed by *Le Monde's* fact-checkers and¹⁹¹ allows readers to search a website's URL to check whether it has published unreliable

¹⁸⁸ Bell, E. (Dec. 15, 2015), *Facebook drains the fake news swamp with new, experimental partnerships*, Columbia Journalism Review,

https://www.cjr.org/tow_center/facebook_drains_fake_news_swamp_new_experimental_partnerships.php

¹⁸⁹ Zimdars, M. (Nov 18, 2016) My 'fake news list' went viral. But made-up stories are only part of the problem, <https://www.washingtonpost.com/posteverything/wp/2016/11/18/my-fake-news-list-went-viral-but-made-up-stories-are-only-part-of-the-problem/>

¹⁹⁰ <https://chrome.google.com/webstore/detail/check-this-by-metacert/felmjclcjadopolhjmlbemfekjaojfbn?hl=en>

¹⁹¹ <http://www.lemonde.fr/verification/>

content. The goal of most of these projects is to build a system that could be integrated with Google and Facebook and used to down-rank certain content from ‘less credible’ sources so that users are less likely to see it.

Recently, the social news agency Storyful teamed up with the advertising analytics firm Moat and CUNY Journalism School to create the Open Brand Safety Framework, an attempt to create the master black list of ‘fake news’ sites that advertisers can avoid.¹⁹² The desire for advertising companies to stay away from problematic content has become increasingly clear over the past few months, as the success of the online activist group Sleeping Giants¹⁹³ demonstrates.¹⁹⁴

Categorising content, while seemingly a well-meaning exercise for providing people with additional context, can quickly backfire when people question the authority of those who create the labels. Indeed, one can easily imagine algorithms programmed to identify, de-rank or take down certain types of content might produce unintended consequences. A clampdown on dis-information should not become an excuse for suppressing dissenting or minority views. An organization labelling something as ‘fake’ should provide full transparency around how it makes its ‘blacklists.’

Credibility scores

In conversations about mis-information, there are regular comparisons drawn between deceptive articles and email spam. Questions are often raised about why techniques similar to those used to combat email spam can’t be used to identify and down-rank poor quality content.

The Trust Project¹⁹⁵, led by Sally Lehrman at Santa Clara University, has been working on a set of criteria that would help audiences know what content to trust. Standards include whether an outlet has a corrections policy and whether a reporter has written on the topic previously. The hope is that if newsrooms added this information as metadata to online articles, Facebook and Google could ‘read’ these signals and place them higher algorithmically. There are also two US-based projects working on credibility: the Technical Schema for Credibility¹⁹⁶, led by Meedan in collaboration with Hacks/Hackers, and the News Quality Score Project, led

¹⁹² Doctor, K., (May 2, 2017) *Can a Master Blacklist Choke Off Fake News*, Nieman Lab.

<http://www.niemanlab.org/2017/05/newsonomics-can-a-master-blacklist-choke-off-fake-news-money-supply/>

¹⁹³ <https://www.facebook.com/slpnggiants/>

¹⁹⁴ Alba, D. (Dec. 15, 2016) *Meet the Ad Companies Ditching Breitbart and Fake News*, Wired, Available at:

<https://www.wired.com/2016/12/fake-news-will-go-away-tech-behind-ads-wont-pay/>

¹⁹⁵ <http://thetrustproject.org/>

¹⁹⁶ Mina, A.X. (March 2, 2017) *Building Technical Standards for Credibility*, MisinfoCon on Medium.

<https://misinfocon.com/building-technical-standards-for-credibility-59ef9ee4ab73>

by Frederic Filloux¹⁹⁷. The projects are developing markers of credibility, to see if there can be a programmatic way for social networks use their scores to influence algorithmic rankings.

Stronger media

CUNY Journalism Professor Jeff Jarvis has said “our problem isn’t ‘fake news.’ Our problem is trust.”¹⁹⁸ As has been well documented, trust in mainstream media has been falling for decades, as has (it must be noted) trust in other public institutions.¹⁹⁹ Ethan Zuckerman’s recent essay on the topic describes the slow decline of trust:

“Addressing the current state of mistrust in journalism will require addressing the broader crisis of trust in institutions. Given the timeline of this crisis, which is unfolding over decades, it is unlikely that digital technologies are the primary actor responsible for the surprises of the past year. While digital technologies may help us address issues, like a disappearing sense of common ground, the underlying issues of mistrust likely require close examination of the changing nature of civics and public attitudes to democracy.”²⁰⁰

It’s worth recognizing how trust in journalism varies geographically. Recent comparative analysis by the Reuters Institute in its annual Digital News Report²⁰¹ shows how news media enjoy different levels of trust in different countries:

¹⁹⁷ Filloux, F. (June 25, 2017) *The News Quality Scoring Project*, Medium, <https://mondaynote.com/the-news-quality-scoring-project-surfacing-great-journalism-from-the-web-48401ded8b53>

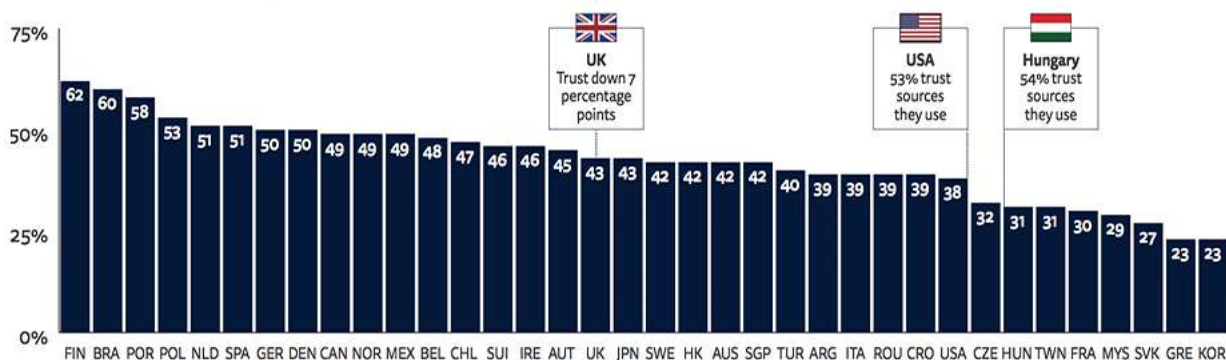
¹⁹⁸ Jarvis, J. (2017) *Our problem isn’t ‘fake news.’ Our problems are trust and manipulation.* <https://medium.com/whither-news/our-problem-isnt-fake-news-our-problems-are-trust-and-manipulation-5bfbc716440>

¹⁹⁹ Harrington, M (Jan. 16, 2017) *Survey: People’s Trust Has Declined in Business, Media, Government, and NGOs*, Harvard Business Review, <https://hbr.org/2017/01/survey-peoples-trust-has-declined-in-business-media-government-and-ngos>

²⁰⁰ Zuckerman, E. (August 2017) *Mistrust, Efficacy and the New Civics*, A whitepaper for the Knight Foundation, <https://dspace.mit.edu/handle/1721.1/110987#files-area>

²⁰¹ Newman, N. (2017) *Digital News Report*, Reuters Institute for the Study of Journalism, https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf

OVERALL TRUST IN NEWS MEDIA – ALL MARKETS



Q6_2016_1/6. Please indicate your level of agreement with the following statements. - I think you can trust most news most of the time/I think I can trust most of the news I consume most of the time Base: Total sample in each market.

Figure 13: From the Digital News Report, 2017 by the Reuters Institute for the Study of Journalism - Overall Trust in the News Media

There are many reasons for the decline of trust in media. Improving these numbers is not going to happen quickly, but initiatives to help build trust and credibility go hand in hand with any initiatives aiming to combat mis- and dis-information.

Strategic Silence

As Data & Society outlined in their May 2017 report Media Manipulation and Dis-information Online, “for manipulators, it doesn’t matter if the media is reporting on a story in order to debunk or dismiss it; the important thing is getting it covered in the first place.”²⁰² Certainly, during the weekend of the #MacronLeaks, Ryan Broderick of BuzzFeed reported that members of 4Chan discussion boards were linking to stories debunking the information, and celebrating them as a form of engagement.²⁰³ While reporting on these stories, and the people behind the stories, feels a natural response by journalists at this point in time, there is a real need for the industry to come together to discuss the impact of reporting on dis-information, and providing oxygen for rumours or fabricated content that otherwise would stay in niche communities online. We would recommend cross-industry meetings whereby senior editors could discuss whether there is a need to reach a shared agreement on when a rumour or piece of content crosses a tipping point, and moves from niche online communities to a wider audience. The French rules which prevented any discussion of election related topics for the forty-eight hours before the polls closed, meant there was no discussion of the

²⁰² Marwick, A and R. Lewis (May 2017) *Media Manipulation and Dis-information Online*, Data & Society, <https://datasociety.net/output/media-manipulation-and-disinfo-online/>, p.39

²⁰³ Ryan Broderick, (broderick) “I covered a different 4chan Macron rumor last week. They don't care if it's not true. They want it debunked”. (May 5, 2017, 11.19am) <https://twitter.com/broderick/status/860423715842121728?lang=en>

leaks by the mainstream media in France, something which raised eyebrows amongst US journalists. The idea of strategic silence in the coverage of mal- and dis-information might sit uncomfortably with some, but we would argue there is a need for these conversations to take place.

Identifying the sources of dis-information

In Paul & Matthews report on Russian propaganda for the RAND corporation, they argue that one of the most effective ways of tackling the issue is to inoculate users, or to “forewarn audiences of mis-information, or merely reach them first with the truth, rather than retracting or refuting false ‘facts.’”²⁰⁴

However, the current trend is fact-checking initiatives.²⁰⁵ Since 2016, we have seen the creation of numerous fact-checking organizations, new teams²⁰⁶ and election-based initiatives like CrossCheck,²⁰⁷ which worked to debunk rumours and claims around the French election. The difficulty here is that if “fake news isn’t about facts, but about power, then independent fact-checking alone won’t fix it — particularly for readers who already distrust the organizations that are doing the fact-checking.”²⁰⁸ (Borel, 2017)

However, as Jeff Jarvis argues, there are other techniques which are not currently a natural part of reporting. For example, “journalism should cover the manipulators’ methods but not their messages.... We should not assume that all our tried-and-true tools — articles, explainers, fact-checking — can counteract manipulators’ propaganda. We must experiment and learn what does and does not persuade people to favor facts and rationality.” (Jarvis, 2017)

A Belgium start-up, *Saper Vedere* is making a similar claim, based on its analysis of the effectiveness of fact-checks during the French election²⁰⁹. In its visualisation below, one can see the audience for the rumour that Macron was funded by Saudi Arabia, as well as the audience of its debunk. There is almost no overlap between these two groups.

²⁰⁴ Paul and Matthews, (2016) p.9

²⁰⁵ Mantzarlis, A. (2016) *There’s been an explosion of international fact-checkers, but they face big challenges*, Poynter, <http://www.poynter.org/2016/theres-been-an-explosion-of-international-fact-checkers-but-they-face-big-challenges/415468/>

²⁰⁶ See the launch of the BBC’s Reality Check: Jackson, J. (Jan. 12, 2017) BBC sets up team to debunk fake news, *The Guardian*, <https://www.theguardian.com/media/2017/jan/12/bbc-sets-up-team-to-debunk-fake-news>

²⁰⁷ <https://firstdraftnews.com/project/crosscheck/>

²⁰⁸ Borel, B. (Jan. 4, 2017) *Fact-checking Won’t Save Us from Fake News*, FiveThirtyEight, <https://fivethirtyeight.com/features/fact-checking-wont-save-us-from-fake-news/>

²⁰⁹ <http://www.saper-vedere.eu/>

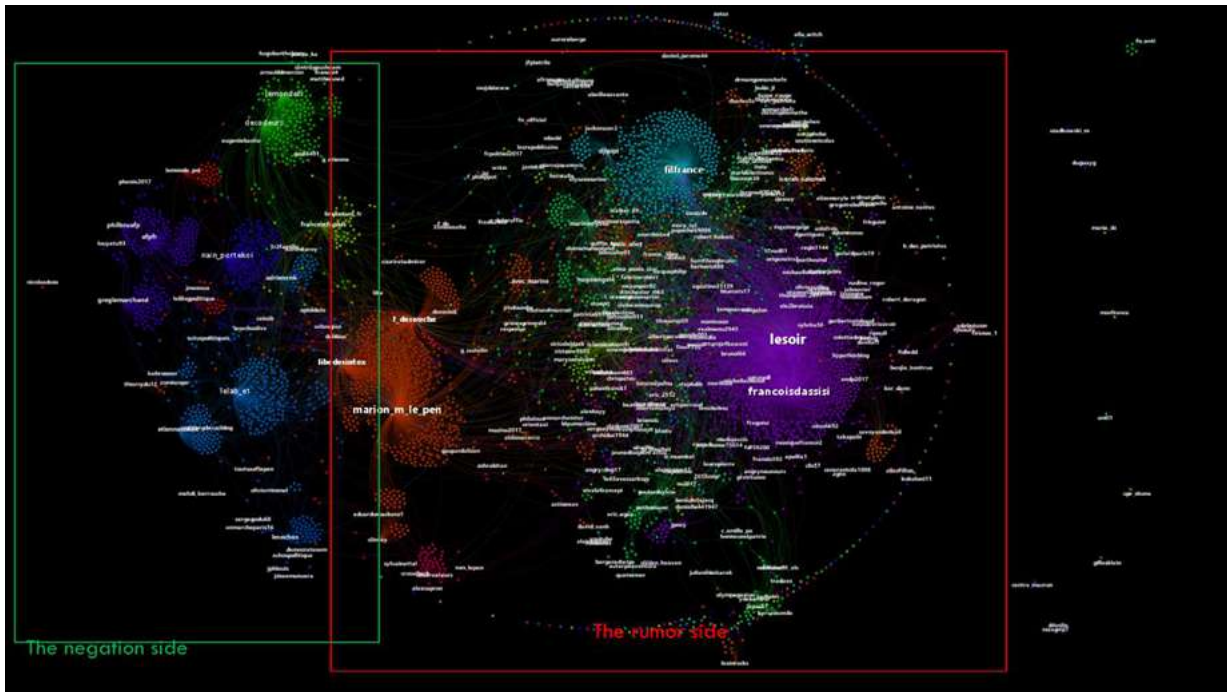


Figure 14: Visualization of Twitter accounts discussion a rumour that Emmanuel Macron was being funded by Saudi Arabia. The accounts on the left in the green box were discussing the debunk. The account in red were discussing the rumour. The two communities hardly overlap. Credit: Alexandre Alaphilippe and Nicolas Vanderbiest.

Instead, they argue that journalists need better tools to be able to identify the sources of disinformation in real-time: source-checking. When bot accounts who originated a rumour appear to be based in a country other than the one connected with said rumour, it could prove to be a faster way of encouraging skepticism in the audience than debunking the fact itself.

Education

In a large-scale exercise designed to evaluate students' ability to evaluate information sources online, researchers at Stanford University were surprised by the degree to which respondents were unable to identify an advert from editorial content or question the partisan nature of facts presented to them.²¹⁰ The call for more news literacy programs²¹¹ has been deafening recently, and they are one solution on which almost everyone can agree.

Danah boyd in a provocative piece titled 'Did Media Literacy Backfire' from January 2017 argued media literacy has actually taught students not to trust Wikipedia while failing to give

²¹⁰ Stanford History Education Group, (Nov. 22, 2016) Evaluation Information: The Cornerstone of Civic Online Reasoning. <https://sheg.stanford.edu/upload/V3LessonPlans/Executive%20Summary%2011.21.16.pdf>

²¹¹ This document from 1999 titled '7 Great Debates in Media Literacy, is still incredibly relevant' <http://files.eric.ed.gov/fulltext/ED439454.pdf>

them sufficient critical research skills to know how to ascertain the credibility of any one piece of information.²¹² boyd identified a significant problem: news literacy has been distorted into distrust of the media and selective research that reaffirms beliefs.

The specifics of how such programs should be rolled out, in terms of the format, structure and content of program curriculum, have supported very vibrant discussions. In addition to more traditional ideas around news literacy, such as how to differentiate between opinion and hard news, there have been calls to include elements like the critical assessment of statistical and quantitative statements in the media²¹³, a deep understanding of algorithms and artificial intelligence²¹⁴ and greater emotional skepticism.²¹⁵

There is also a need to educate people on the power of images to manipulate and persuade. As discussed earlier, the way we understand visuals is fundamentally different to how we think about text. While much of the ‘fake news’ debate to date has been about text-based dis-information, the election monitoring projects First Draft has worked on in the US, UK, France and Germany have shown how frequently dis-information appears in visual formats—whether doctored images, fabricated videos, misleading visualizations or ‘memes’ (striking images with text superimposed over top). In an investigation carried out in the run up to the French election, BuzzFeed discovered loose networks of US teenagers creating ‘meme shells’ (generic images related to the candidates) that anyone could use to create memes for social media.

A recent study at Stanford University observed 10 Ph.D. historians, 10 professional fact checkers and 25 Stanford University undergraduates as they evaluated live websites and searched for information on social and political issues. They found that “historians and students often fell victim to easily manipulated features of websites, such as official-looking logos and domain names. They read vertically, staying within a website to evaluate its reliability. In contrast, fact checkers read laterally, leaving a site after a quick scan and opening new browser tabs in order to judge the credibility of the original site. Compared to

²¹² boyd, d. (Jan. 5, 2017) *Did Media Literacy Backfire*, Data and Society: Points <https://points.datasociety.net/did-media-literacy-backfire-7418c084d88d>

²¹³ Written evidence submitted by the Royal Statistical Society to the UK Parliamentary Inquiry on Fake News, <https://www.parliament.uk/business/committees/committees-a-z/commons-select/culture-media-and-sport-committee/inquiries/parliament-2015/inquiry2/publications/>

²¹⁴ Written evidence submitted by the UCL Knowledge Lab to the UK Parliamentary Inquiry on Fake News <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/culture-media-and-sport-committee/fake-news/written/48571.html>

²¹⁵ On The Media, (2017) Rise and Fall of Fake News, WNYC, <http://www.wnyc.org/story/rise-and-fall-fake-news/>

the other groups, fact checkers arrived at more warranted conclusions in a fraction of the time.”²¹⁶

The question is how to make this type of ‘reading’ habitual amongst students.

Ultimately, there is an acceptance that any curriculum should not lecture students. Telling students they are wrong is not a solution, and may even be counter-productive. As InformAll and the CILIP Information Literacy Group testified to the UK Parliament, “[t]he essence of any solution lies in stimulating curiosity and a spirit of enquiry, and crucially, finding effective ways of triggering this curiosity, in the education system and beyond.”²¹⁷

One of the most impressive initiatives is The Digital Polarization Initiative²¹⁸, which was launched by the American Association of State Colleges and Universities and is led by Mike Caulfield. It is a national effort to build student civic, information and web literacy by having students participating in a broad, cross-institutional project to fact-check, annotate and provide context to the different news stories that show up in their Twitter and Facebook feeds. As Caulfield explains: “The point is to get students to understand the mechanisms and biases of Facebook and Twitter in ways that most digital literacy programs never touch. The point is not to simply decode what’s out there, but to analyze what is missing from our current online environment, and, if possible supply it.”²¹⁹

Programs that have focused on critical thinking, source evaluation and emotional manipulation have seen success. In Ukraine, the nongovernmental organization IREX, trained 15,000 people on a program called Learn to Discern, which was designed to teach citizens how to separate fact from fiction and recognize manipulation and hate speech. In their evaluation of the project, they found an observed 24% increase in participants’ ability to distinguish trustworthy news from false news, a 22% increase in those who cross-check the information in the news they consume, and a 26% increase in participants' confidence in analyzing news.²²⁰

²¹⁶ McGrew, S., T. Ortega, J. Breakstone & S. Wineburg, (Fall 2017) *The Challenge That’s Bigger Than Fake News: Teaching Students to Engage in Civic Online Reasoning*. American Educator.

²¹⁷ Written evidence submitted by InformAll and the CILIP Information Literacy Group to the UK Parliamentary Inquiry on Fake News, <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/culture-media-and-sport-committee/fake-news/written/48215.html>

²¹⁸ <http://www.aascu.org/AcademicAffairs/ADP/DigiPo/>

²¹⁹ Caulfield, M. (Dec 7. 2016) *Announcing the Digital Polarization Initiative*. Hapgood. <https://hapgood.us/2016/12/07/announcing-the-digital-polarization-initiative-an-open-pedagogy-joint/>

²²⁰ Susman-Peña, T. and Vogt, Katya (June 12, 2017) *Ukrainians’ self-defense against information war: What we learned from Learn to Discern*, IREX, <https://www.irex.org/insight/ukrainians-self-defense-against-information-war-what-we-learned-learn-discern>

Established programs like the News Literacy Project²²¹, which is focused on providing materials and curricula to High School students; the Stonybrook Center for News Literacy²²², which offers skills training to University students; and a new online course being offered by Hong Kong University²²³ are all also currently leading the thinking on best practices in this area.

It seems there is a need for a task force on the best approaches for teaching news literacy, creative thinking about a standardized curriculum and rigorous testing of new techniques. Suggested elements of any curriculum include: (i) traditional news literacy skills; (ii) forensic social media verification skills; (iii) information about the power of algorithms to shape what is presented to us; (iv) the possibilities but also the ethical implications offered by artificial intelligence; (v) techniques for developing emotional scepticism to override our brain's tendency to be less critical of content that provokes an emotional response; and (vi) statistical numeracy.

Regulation

The First Amendment of the US Constitution means that, for all the discussion in the US about the impact of fabricated and manipulated content, there is very little appetite in the US for any type of regulatory intervention.²²⁴ In Europe however, the regulatory wheels have been turning slowly, and we are starting to see legislation directed at information disorder. Germany, for instance, recently passed the Network Enforcement Law, which concentrates primarily on hate speech, and introduces a possible imposition of fines on the social networks if they don't take down hateful or defamatory content within twenty-four hours. The BBC World Service survey measuring attitudes to information and the internet of eighteen countries, showed that only in two of the eighteen countries, China and the UK, did a majority want their governments to regulate the internet.²²⁵

However, there's little denying that the regulatory discourse in Europe has been loud since late December 2016, when Giovanni Pitruzzella, the chairman of the Italian Competition

²²¹ <http://www.thenewsliteracyproject.org/>

²²² <https://www.centerfornewsliteracy.org/>

²²³ <https://www.coursera.org/learn/news-literacy>

²²⁴ Gillespie, T. (in press) *Governance of and by platforms*. In Burgess, J., Poell, T., & Marwick, A. (Eds), SAGE handbook of social media. Retrieved from <http://culturedigitally.org/wp-content/uploads/2016/06/Gillespie-Governance-ofby-Platforms-PREPRINT.pdf>

²²⁵ Cellan-Jones, R. (Sept 22, 2017) Fake news worries 'are growing' suggests BBC poll, BBC News, <http://www.bbc.com/news/technology-41319683>

Authority, told the Financial Times that EU countries should deal with "post-truth" politics by setting up antitrust-like agencies devoted to spotting and removing fake news.²²⁶

And then, in January 2017, Andrus Ansip, the European Commission (EC) Vice President for the Digital Single Market, warned that if Facebook and other tech companies didn't take tougher stances on fake news, the EC might have to step in. Ansip told the Financial Times in an interview, "I really believe in self-regulatory measures, but if some kind of clarifications are needed then we will be ready for that."²²⁷ However, on Twitter, he stressed²²⁸ that he was not referring to a 'ministry of truth.'

The European Commission has already pushed Facebook, Twitter, YouTube and Microsoft to sign up to a code of conduct²²⁹ that aims to tackle online hate speech and take down the majority of potentially illegal content within 24 hours. Many fear this code of conduct might become a blueprint for regulating fabricated content online.

In the United Kingdom, the Culture, Media and Sport Committee set up a Fake News Inquiry, and evidence was submitted²³⁰ by 79 experts and organizations. The inquiry was closed when the election was called, and it is unknown whether it will reconvene.

In the Czech Republic, officials are monitoring fake news directly. Ahead of the country's general election in October, the Czech government has set up a "specialised analytical and communications unit"²³¹ within the Ministry of the Interior that, as part of its work to monitor threats to internal security, will also target "dis-information campaigns." According to the Ministry, it will "not force the 'truth' on anyone, or censor media content." Rather, as the unit's Twitter page explains, it will assess whether the dis-information seriously affects internal security, and, if so, it will respond by publicising available facts and data that disprove the fake story.

Whatever happens in Europe will set an important global precedent. Already, Singapore's Law and Home Affairs Minister K. Shanmugam stated that laws to tackle the "scourge of fake news" are expected to be introduced next year.²³²

²²⁶ Politi, J. (Dec. 30, 2016) 'Italy antitrust chief urges EU to help beat fake news'
<https://www.ft.com/content/e7280576-cddc-11e6-864f-20dcb35cede2?mhq5j=e2>

²²⁷ Bond, D. and Robinson, D. (Jan. 29, 2017), 'European Commission fires warning at Facebook over fake news'
Financial Times, <https://www.ft.com/content/85683e08-e4a9-11e6-9645-c9357a75844a?mhq5j=e2>

²²⁸ https://twitter.com/Ansip_EU/status/826085369493995522

²²⁹ http://europa.eu/rapid/press-release_IP-16-1937_en.htm

²³⁰ <https://www.parliament.uk/business/committees/committees-a-z/commons-select/culture-media-and-sport-committee/inquiries/parliament-2015/inquiry2/publications/>

²³¹ <http://www.mvcr.cz/cthh/clanek/centre-against-terrorism-and-hybrid-threats.aspx>

²³² Yi, S.B. (June 19, 2017) 'New legislation to combat fake news likely to be introduced next year: Shanmugam',
Straits Times, <http://www.straitstimes.com/singapore/new-legislation-to-combat-fake-news-next-year-shanmugam>

Any attempt to create a regulatory framework will be problematic without appropriate definitions for information disorder. When politicians or policymakers talk about ‘fake news’, what are they targeting? Fabricated news sites created for profit? Twitter raids created by loose networks of bored teens?

As Jan Kleijssen, the Director of the Information Society and Action against Crime Department of the Council of Europe reminds us, “When we speak about freedom of expression today, we often hear a ‘but’ - and then mention is made of ‘hate speech’ and ‘fake news’. At the Council of Europe, we believe that we have to be very careful with that ‘but’ after freedom of expression. We are talking about one of the most important foundations of democracy, one of the most important foundations of democratic security.” The topics of mis-, mal- and dis-information are too important to start legislating and regulating around until we have a shared understanding of what we mean by these terms.



Figure 15: Cartoon by Cathy Wilcox, drawn for UNESCO for World Press Freedom Day 2017.

An easier regulatory move that we are likely to see soon is connected to online advertising on Facebook. With the news that Russia was buying ‘dark’ posts on Facebook and targeting them at US citizens in the run up to the 2016 US election, there is a growing pressure for increased

transparency around these types of advertisements.²³³ Without any oversight on what is being published and to whom, there can be no accountability. In most democracies paid-for election related communication is held to certain standards before it can be broadcast or published. In 2011, the Federal Communications Commission ruled that Facebook did not have to require disclaimers on its paid-for posts, but we expect this to be revisited as the opportunities presented by this technology, to those trying to sow dis-information, become clearer.

While Mark Zuckerberg announced on September 21, 2017 that Facebook will ensure that anyone advertising on Facebook will have to disclose which page paid for an ad, and will also ensure that you can visit an advertiser's page and see the ads that they are currently running to any audience on Facebook. While this seems like a positive step, as a group of distinguished academics wrote in response to the announcement via an open letter:

Transparency is a first step in the right direction. Digital political advertising operates in a dynamic tension between data and humans, commerce and politics, power and participation. Some of these tensions can be resolved by transparency, others not. The way forward is to engage with governments, regulators, election monitoring bodies, civil society and academics to develop public policies and guidelines for ensuring fairness, equality, and democratic oversight in digital political campaigns.²³⁴

²³³ Vaidhyanathan, S. (Sept 8. 2017) *Facebook Wins, Democracy Loses*. New York Times, <https://www.nytimes.com/2017/09/08/opinion/facebook-wins-democracy-loses.html>

²³⁴ Helberger, N. et al. (Sept 22. 2017) *Dear Mark: An Open Letter to Mark Zuckerberg in response to his statement on political advertising on Facebook*, Available at: https://www.dropbox.com/s/7v3vpk9yw5sa18b/Dear%20Mark_final..pdf?dl=0

Part 4: Future trends

Messaging apps

As discussed, much of the recent debate around mis- and dis-information has focused on their political impact, and this debate has been largely shaped by events during the US election. As a result, much of the focus has been on the Facebook News Feed. But even a cursory glance outside of the US demonstrates that the next frontier for mis- and dis-information is closed-messaging apps.²³⁵

According to the Digital News Report published by the Reuters Institute for the Study of Journalism²³⁶, the use of the Facebook-owned WhatsApp as a news source rivals Facebook in a number of markets, including Malaysia, Brazil and Spain. While WhatsApp is clearly the most dominant messaging app globally, the popularity of different apps in other countries is quite startling. For example, WeChat, the most popular messaging app in China, has 963 million users as of Q2 2017.²³⁷

The obvious challenge of tackling rumours and fabricated content on these messaging apps is that it's impossible to know what is being shared. There are innovative projects attempting to tackle rumours being shared on these apps. One example is the Thai News Agency's 'Sure and Share' project, which encourages audience members to submit questions they have about content, rumours or stories circulating on the messaging app LINE. The news agency then creates engaging infographics or YouTube videos based on their fact-checking and shares them on their LINE channel. Similar initiatives are emerging for WhatsApp in Colombia²³⁸ and India²³⁹.

Augmented reality, artificial reality and voice recognition

As we continue to undertake research and work collaboratively on solutions, our biggest challenge will be the speed at which technology is refining the creation of fabricated video and audio.

²³⁵ Dias, N. (August 17, 2017) *The Era of Whatsapp Propaganda is Upon Us*, Foreign Policy, <http://foreignpolicy.com/2017/08/17/the-era-of-whatsapp-propaganda-is-upon-us/>

²³⁶ Newman, N. (2017), p. 10.

²³⁷ Tencent. (n.d.). Number of monthly active WeChat users from 2nd quarter 2010 to 2nd quarter 2017 (in millions). In Statista - The Statistics Portal. Retrieved September 15, 2017, from <https://www.statista.com/statistics/255778/number-of-active-wechat-messenger-accounts/>

²³⁸ Serrano, C. (March 20, 2017) To slow the spread of false stories on WhatsApp, this Colombian news site is enlisting its own readers, Nieman Lab, <http://www.niemanlab.org/2017/03/to-slow-the-spread-of-false-stories-on-whatsapp-this-colombian-news-site-is-enlisting-its-own-readers/>

²³⁹ https://twitter.com/boomlive_in/status/861559074378452992

Research by Justus Thies and colleagues²⁴⁰ has demonstrated how technologists can change facial expressions in live video. And, as Nick Bilton wrote in an article for *Vanity Fair*, “The technology out of Stanford that can manipulate a real-time news clip doesn’t need an array of high-end computers like those used by Pixar; it simply needs a news clip from YouTube and a standard Webcam on your laptop.”

More recently, researchers at the University of Washington used artificial intelligence to create visually convincing videos of Barack Obama saying things he had said before, but in a completely different context.²⁴¹ The researchers fed a neural network seventeen hours of footage from the former president’s weekly addresses as ‘training data’. The resulting algorithm was able to generate mouth shapes from Obama’s voice and overlay them onto Obama’s face in a different “target” video.

Audio can be manipulated even more easily than video. Adobe has created Project VoCo, which has been nicknamed ‘Photoshop for audio’. The product allows users to feed a 10-to-20 minute clip of someone’s voice into the application and then dictate words in that person’s exact voice. Another company called Lyrebird²⁴² is working on voice generation. On its site, it claims to “need as little as one minute of audio recording of a speaker to compute a unique key defining her/his voice. This key will then generate anything from its corresponding voice.” It also plans to create an API whereby other platforms could easily use those voices.

Finally, Mark Zuckerberg at Facebook’s Developer Conference, F8, in April 2017, demonstrated new Augmented Reality technology that allows users to seamlessly ‘add’ features and filters to their images or videos. Zuckerberg used the example of adding more steam to the image of his morning coffee. While this is a harmless example, the darker versions of augmented reality are easy to imagine.

²⁴⁰ Thies, J. et al. (2016) Face2Face: Real-time Face Capture and Reenactment of RGB Videos, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2387-2395

²⁴¹ Suwajanakorn, S. et al. (July 2017) Synthesizing Obama: Learning Lip Sync from Audio, ACM Transactions on Graphics, 36 (4). Article 95. http://grail.cs.washington.edu/projects/AudioToObama/siggraph17_obama.pdf

²⁴² <https://lyrebird.ai/demo>

Part 5: Conclusions

This report has provided a conceptual framework for thinking about information disorder. We hope that the definitions explained here will provide a structure for conversations by policymakers, legislators and researchers who are investigating the phenomenon. Only by beginning with a shared understanding can we start constructively discussing solutions. We also hope that our conceptual framework, which outlines the different elements and phases of information disorder, will help bring nuance to debates about this issue.

First, we need to understand communication as something beyond just a transmission of messages. People’s consumption of news and information is, first and foremost, a way to reaffirm their affinity with a larger dramatic narrative about the world and their place in it, and transcends facts and figures.

Second, if we are serious about creating solutions, we need to consider the specific motivations of different types of ‘Agents’, the characteristics of different types of ‘Messages’ and the factors impacting how people ‘Interpret’ those messages. We also need to recognise how messages and the motivations about them can shift and transform as other agents re-produce and disseminate these messages.

We have also outlined research from different disciplines and methodological backgrounds. In particular, we want to connect the excellent experimental work that has helped us to understand how people process information with the sociological and cultural theories that highlight how and why people seek out information and use it to position themselves within certain ‘tribes’.

To us, it is clear that any solutions will need to be based on a multi-disciplinary approach. While some recent psychological research in the US context has highlighted the effects of fact-checking initiatives in making people question information, it has also shown that these initiatives can have little impact on people’s underlying beliefs—as in the case of Donald Trump’s supporters. We need more research on the influence of emotions on the way humans make sense of and use information in their lives.

As D’Ancona underlines, conspiracy theories are effective because they are based on powerful narratives. They unconsciously tap into deep-seated fears. “Veracity will be drowned out unless it is resonant.”²⁴³ There is research that shows that for false information to be challenged effectively, our brains need it to be replaced with an alternative narrative²⁴⁴.

²⁴³ D’Ancona, (2017), p.131

²⁴⁴ Nyan, B. and J. Reifler (2015) ‘Displacing Mis-information about Events: An Experimental Test of Causal Corrections’, *Journal of Experimental Political Science*, 2 (1) : 81-93

So to use rumours about Obama’s religious affiliation as an example, rather than stating ‘Barack Obama is not a Muslim’ it is more effective to provide a story (preferably with a powerful narrative structure) of Obama going to his local Christian church with his family.

We need to fight rumours and conspiracy with engaging and powerful narratives that leverage the same techniques as dis-information. As discussed in Part 1, effective strategies for dis-information include: provoking an emotional response, repetition, a strong visual aspect and a powerful narrative. If we remember the powerful, ritualistic aspects to information seeking and consumption, the importance of integrating these elements into our solutions is obvious.

While the explosion of fact-checking and debunking initiatives is admirable, there is an urgent need to understand the most effective formats for sparking curiosity and skepticism in audiences about the information they consume and the sources from which that information comes. Simply pushing out more ‘factual information’ into the ecosystem, without sufficiently understanding the emotional and ritualistic elements of communication, is potentially a waste of time and resources.

We are seeing success with the use of nudge technology to remind people to check the veracity of information before re-sharing, as well as hopeful initiatives such as the International Center For Journalist’s TruthBuzz competition, which is encouraging people to design fact-checking and debunking formats that are highly engaging and shareable. We are also seeing games being created to help teach news literacy skills.²⁴⁵ Formats for reporting on mis-information do seem matter in terms of connecting with audiences. A recent experiment found that videos were considered “more interesting and understandable” than a comparable print-based fact-checking story.²⁴⁶

And as the director of the International Fact-Checking Network, Alexios Mantzarlis concludes: “We need to find formats for people who are bored with reading long articles stuffed with

²⁴⁵ Schmidt, C. (Aug. 3, 2017) *Games might be a good tool for fighting fake news. Here’s what three developers have learned*, Nieman Lab, <http://www.niemanlab.org/2017/08/games-might-be-a-good-tool-for-fighting-fake-news-heres-what-three-developers-have-learned/>

²⁴⁶ Young, D. et al. (2017) *Fact-Checking Effectiveness as a Function of Format and Tone: Evaluating FactCheck.org and FlackCheck.org*, *Journalism and Mass Communication Quarterly*, 1-27.

hyperlinks. On this front I am glad to note some success on Snapchat²⁴⁷ and bots²⁴⁸. But we haven't seen a breakaway podcast, and seen the struggles with TV."²⁴⁹

The next area of research and testing should be devoted to understanding the performative element of why people share mis-, mal- and dis-information. How can that be slowed down? What cultural factors would make it shameful or embarrassing? In addition to teaching emotional scepticism as part of news literacy education programs, how do we teach people how to call out their friends and family when they post fabricated or misleading content on social networks or in closed messaging apps? Brooke Borel, in a piece titled 'How to Talk to Your Facebook Friends about Fake News,'²⁵⁰ provided tips based on academic literature and a case study of two old high school friends who managed to overcome their political differences via a long exchange on Facebook. We need more research, and accessible resources derived from this research, to help people navigate the challenges of peer-to-peer news literacy.

The technological developments outlined in the previous section demonstrate that, as we run to catch up with the current phenomenon of polluted information streams, we are going to have to run faster if we want to be well-placed to deal with these technological advancements. In a short amount of time, audiences will have little trust in the information they find online, dismissing any image, video or audio clip as potentially fabricated or manipulated. The implications of this reality are truly terrifying, particularly as our societies become increasingly polarized and divided. We must work together on solutions driven by research and experimentation to mitigate dis-information and significantly improve information literacy. Knee-jerk reactions based on poor definitional frameworks, or simplistic calls to limit access to information, will only create more problems in the long run. Information disorder cannot be solved overnight, but the first step is understanding the complexity of the issue. We hope this report has provided useful framing and context.

²⁴⁷ Mantzarlis, A (Feb. 2, 2016) *Fact-checkers experiment with Snapchat, GIFs and other stuff millennials* ♥. Poynter.

<http://www.poynter.org/2016/fact-checkers-experiment-with-snapchat-gifs-and-other-stuff-millennials-%E2%99%A5/393992/>

²⁴⁸ Mantzarlis, A (May 10, 2016) *Did Clinton or Trump twist the facts? This messaging bot will tell you*. Poynter. <http://www.poynter.org/2016/did-clinton-or-trump-twist-the-facts-this-messaging-bot-will-tell-you/409457/>

²⁴⁹ Mantzarlis, A (June 7, 2016) *There's been an explosion of international fact-checkers, but they face big challenges*, Poynter. <http://www.poynter.org/2016/theres-been-an-explosion-of-international-fact-checkers-but-they-face-big-challenges/415468/>

²⁵⁰ Borel, B, (2017) *How to Talk to your Facebook Friends about Fake News*, Open Notebook <http://www.theopennotebook.com/2017/02/21/how-to-talk-to-your-facebook-friends-about-fake-news/>

Part 6: Recommendations

*What could **technology companies** do?*

1. **Create an international advisory council.** We recommend the creation of an independent, international council, made up of members from a variety of disciplines that can (1) guide technology companies as they deal with information disorder and (2) act as an honest broker between technology companies.
2. **Provide researchers with the data related to initiatives aimed at improving the quality of information.** While technology companies are understandably nervous about sharing their data (whether that's metrics related to how many people see a fact-check tag, or the number of people who see a 'disputed content' flag and then do not go on to share the content), independent researchers must have better access to this data in order to properly address information disorder and evaluate their attempts to enhance the integrity of public communication spaces. As such, platforms should provide whatever data they can—and certainly more than they are currently providing.
3. **Provide transparent criteria for any algorithmic changes that down-rank content.** Algorithmic tweaks or the introduction of machine learning techniques can lead to unintended consequences, whereby certain types of content is de-ranked or removed. There needs to be transparency around these changes so the impact can be independently measured and assessed. Without this transparency, there will be claims of bias and censorship from different content producers.
4. **Work collaboratively.** Platforms have worked together to fight terrorism and child abuse. Slowly, collaboration is also beginning to happen around information disorder, and we encourage such collaboration, particularly when it involves sharing information about attempts to amplify content.
5. **Highlight contextual details and build visual indicators.** We recommend that social networks and search engines automatically surface contextual information and metadata that would help users ascertain the truth of a piece of content (for example automatically showing when a website was registered or running a reverse image search to see whether an image is old). The blue verification tick is an example of a helpful visual indicator that exists across platforms. We argue that technology companies should collaborate to build a consistent set of visual indicators for these contextual details. This visual language should be developed in collaboration with cognitive psychologists to ensure efficacy.
6. **Eliminate financial incentives.** Technology companies as well as advertising networks more generally must devise ways to prevent purveyors of dis-information from gaining financially.
7. **Crack down on computational amplification.** Take stronger and quicker action against automated accounts used to boost content.

8. **Adequately moderate non-English content.** Social networks need to invest in technology and staff to monitor mis-, dis- and mal-information in all languages.
9. **Pay attention to audio/visual forms of mis- and dis-information.** The problematic term ‘fake news’ has led to an unwarranted fixation on text-based mis- and dis-information. However, our research suggests that fabricated, manipulated or falsely-contextualized visuals are more pervasive than textual falsehoods. We also expect fabricated audio to become an increasing problem. Technology companies must address these formats as well as text.
10. **Provide metadata to trusted partners.** The practice of stripping metadata from images and video, (for example location information, capture date and timestamps), although protective of privacy and conservative of data, often complicates verification. Thus, we recommend that trusted partners be provided increased access to such metadata.
11. **Build fact-checking and verification tools.** We recommend that technology companies build tools to support the public in fact-checking and verifying rumors and visual content, especially on mobile phones.
12. **Build ‘authenticity engines’.** As audio-visual fabrications become more sophisticated, we need the search engines to build out ‘authenticity’ engines and water-marking technologies to provide mechanisms for original material to be surfaced and trusted.
13. **Work on solutions specifically aimed at minimising the impact of filter bubbles:**
 - a) **Let users customize feed and search algorithms.** Users should be given the chance to consciously change the algorithms that populate their social feeds and search results. For example, they should be able to choose to see diverse political content or a greater amount of international content in their social feeds.
 - b) **Diversify exposure to different people and views.** Using the existing algorithmic technology on the social networks that provides suggestions for pages, accounts, or topics to follow, these should be designed to provide exposure to different types of content and people. There should be a clear indication that this is being surfaced deliberately, and while the views or content might be uncomfortable or challenging, it is necessary to have an awareness of different perspectives.
 - c) **Allow users to consume information privately.** To minimize performative influences on information consumption, we recommend that technology companies provide more options for users to consume content privately, instead of publicizing everything they ‘like’ or ‘follow’.
 - d) **Change the terminology used by the social networks.** Three common concepts of the social platforms unconsciously affect how we avoid different views and remain in our echo chambers. ‘To follow’, for most people subconsciously implies a kind of agreement, so it emotionally creates a resistance against

exposure to diverse opinion. ‘Friend’ also connotes a type of bond you wouldn’t want to have with those you strongly disagree with but are curious about. So is the case with ‘like’, when you want to start reading a certain publication on Facebook. We should instead institute neutral labels such as connecting to someone, subscribing to a publication, bookmarking a story, etc.

*What could **national governments** do?*

1. **Commission research to map information disorder.** National governments should commission research studies to examine information disorder within their respective countries, using the conceptual map provided in this report. What types of information disorder are most common? Which platforms are the primary vehicles for dissemination? What research has been carried out that examines audience responses to this type of content in specific countries? The methodology should be consistent across these research studies exercises, so that different countries can be accurately compared.
2. **Regulate ad networks.** While the platforms are taking steps to prevent fabricated ‘news’ sites from making money, other networks are stepping in to fill the gap. States should draft regulations to prevent any advertising from appearing on these sites.
3. **Require transparency around Facebook ads.** There is currently no oversight in terms of who purchases ads on Facebook, what ads they purchase and which users are targeted. National governments should demand transparency about these ads so that ad purchasers and Facebook can be held accountable.
4. **Support public service media organisations and local news outlets.** The financial strains placed on news organisations in recent years has led to ‘news deserts’ in certain areas. If we are serious about reducing the impact of information disorder, supporting quality journalism initiatives at the local, regional and national level needs to be a priority.
5. **Roll out advanced cyber-security training.** Many government institutions use bespoke computer systems that are incredibly easy to hack, enabling the theft of data and the generation of mal-information. Training should be available at all levels of government to ensure everyone understands digital security best practices and to prevent attempts at hacking and phishing.
6. **Enforce minimum levels of public service news on to the platforms.** Encourage platforms to work with independent public media organisations to integrate quality news and analysis into users’ feeds.

*What could **media organisations** do?*

1. **Collaborate.** It makes little sense to have journalists at different news organisations fact-checking the same claims or debunking the same visual content. When it comes to debunking mis- or dis-information, there should be no ‘scoop’ or ‘exclusive’. Thus,

- we argue that newsrooms and fact-checking organisations should collaborate to prevent duplications of effort and free journalists to focus on other investigations.
2. **Agree policies on strategic silence.** News organisations should work on best practices for avoiding being manipulated by those who want to amplify mal- or dis-information.
 3. **Ensure strong ethical standards across all media.** News organizations have been known to sensationalize headlines on Facebook in ways that wouldn't be accepted on their own websites. News organizations should enforce the same content standards, irrespective of where their content is placed.
 4. **Debunk sources as well as content.** News organisations are getting better at fact-checking and debunking rumours and visual content, but they must also learn to track the sources behind a piece of content in real time. When content is being pushed out by bot networks, or loose organised groups of people with an agenda, news organisations should identifying this as quickly as possible. This will require journalists to have computer programming expertise.
 5. **Produce more segments and features about critical information consumption.** The news media should produce more segments and features which teach audiences how to be critical of content they consume. When they write debunks, they should explain to the audience how the process of verification was undertaken.
 6. **Tell stories about the scale and threat posed by information disorder.** News and media organisations have a responsibility to educate audiences about the scale of information pollution worldwide, and the implications society faces because of it, in terms of undermining trust in institutions, threatening democratic principles, inflaming divisions based on nationalism, religion, ethnicity, race, class, sexuality or gender.
 7. **Focus on improving the quality of headlines.** User behaviour shows the patterns by which people skim headlines via social networks without clicking through to the whole article. It therefore places greater responsibility on news outlets to write headlines with care. Research²⁵¹ using natural language processing techniques are starting to automatically assess whether headlines are overstating the evidence available in the text of the article. This might prevent some of the more irresponsible headlines from appearing.
 8. **Don't disseminate fabricated content.** News organisations need to improve standards around publishing and broadcasting information and content sourced from the social web. There is also a responsibility to ensure appropriate use of headlines, visuals, captions and statistics in news output. Clickbait headlines, the misleading use of statistics, unattributed quotes are all adding to the polluted information ecosystem.

²⁵¹ Chesney, S., M. Liakata, M. Poesio and M. Purver (2017) *Incongruent Headlines: Yet Another Way to Mislead Your Readers*, <http://www.eecs.qmul.ac.uk/~mpurver/papers/chesney-et-al17nlpj.pdf>

What could *civil society* do?

1. **Educate the public about the threat of information disorder.** There is a need to educate people about the persuasive techniques that are used by those spreading dis- and mal-information, as well as a need to educate people about the risks of information disorder to society, i.e., sowing distrust in official sources and dividing political parties, religions, races and classes.
2. **Act as honest brokers.** Non-profits and independent groups can act as honest brokers, bringing together different players in the fight against information disorder, including technology companies, newsrooms, research institutes, policy-makers, politicians and governments.

What could *education ministries* do?

1. **Work internationally to create a standardized news literacy curriculum.** Such a curriculum should be for all ages, based on best practices, and focus on adaptable research skills, critical assessment of information sources, the influence of emotion on critical thinking and the inner workings and implications of algorithms and artificial intelligence.
2. **Work with libraries.** Libraries are one of the few institutions where trust has not declined, and for people no longer in full time education, they are a critical resource for teaching the skills required for navigating the digital ecosystem. We must ensure communities can access both online and offline news and digital literacy materials via their local libraries.
3. **Update journalism school curricula.** Ensure journalism schools teach computational monitoring and forensic verification techniques for finding and authenticating content circulating on the social web, as well as best practices for reporting on information disorder.

What could *Grant-Making Foundations* do?

1. **Provide support for testing solutions.** In this rush for solutions, it is tempting to support initiatives that 'seem' appropriate. We need to ensure there is sufficient money to support the testing of any solutions. For example, with news literacy projects, we need to ensure money is being spent to assess what types of materials and teaching methodology are having the most impact. It is vital that academics are connecting with practitioners working in many different industries as solutions are designed and tested. Rather than small grants to multiple stakeholders, we need fewer, bigger grants for ambitious multi-partner, international research groups and initiatives.
2. **Support technological solutions.** While the technology companies should be required to build out a number of solutions themselves, providing funding for smaller startups

to design, test and innovate in this space is crucial. Many solutions need to be rolled out across the social platforms and search engines. These should not be developed as proprietary technology.

- 3. Support programs teaching people critical research and information skills.** We must provide financial support for journalistic initiatives which attempt to help audiences navigate their information ecosystems, such as public-service media, local news media and educators teaching fact-checking and verification skills.

Appendix: European Fact-checking and Debunking Initiatives

According to a study published in 2016 by Lucas Graves and Federica Cherubini, there are 34 permanent fact checking operations that exist across 20 European countries. There are two different types: those attached to news organisations (around 40%) and those that operate as nonprofits (around 60%).²⁵²

Much of the following information is drawn from the [fact-checking database](#) created by Duke University's Reporter's Lab.

Austria: [Fakt ist Fakt](#), an independent fact checking organisation. They examine the truths of public figures, especially politicians.

Bosnia: [Istinomjer](#), a project of Bosnia's Zašto ne? (Why Not?), a peace-building group, founded in 2001. Its goal is to make political and public discourse in Bosnia & Herzegovina "more relevant to the promotion of political accountability as the fundamental principle of democracy." Zašto ne? has been financially supported by the National Endowment for Democracy (NED) since its inception.

Croatia: [Faktograf](#) rates political claims on a scale from "Fact" to "Not even the F of Fact" and keeps a database that supports researchers' efforts in collecting information on the degree to which promises made by public officials are fulfilled. Their rating system emulates the one popularized by PolitiFact's Truth-O-Meter and was adopted with modifications by the most external fact-checkers worldwide. Zašto ne?, which is behind the Bosnian *Istinomjer*, did the programming and designing for *Faktograf*'s website with support from the National Endowment for Democracy and TechSoup.

Czech Republic: [Demagog](#) was created in February 2012 and is linked to the Slovak version of the project, Demagog.sk. It originated as a voluntary and independent initiative of Matej Hruška and Ondrej Lunter, students of Masaryk University in Brno.

Denmark:

- ❑ [Detektor](#) is a fact-checking show produced by Denmark's public broadcaster, DR.
- ❑ [TjekDet](#) is the fact-checking vertical of *Mandag Morgen*, a Danish weekly business and political magazine.

Finland: [Faktabaari](#) is a Finnish site launched in 2014 by an NGO called the Open Society Association. It is managed by a voluntary staff of professional journalists, EU experts and technical staff with the help of a broader network of topical experts.

France:

- ❑ Libération's [Désintox](#), which was launched in 2008, fact-checks politicians and debunks outright rumors and fake stories.
- ❑ [Le Vrai du Faux](#), from the radio and television network franceinfo, is a news site that

²⁵² Graves, L.& Cherubini, F. (2016). *The rise of fact-checking sites in Europe*. Reuters Institute for the Study of Journalism. <http://reutersinstitute.politics.ox.ac.uk/publication/rise-fact-checking-sites-europe>

sifts through mis-information circulating on websites and social networks. It also has a programme that airs daily on franceinfo.

- ❑ [Les Observateurs](#), an online hub and television programme from the multilingual network France24, has a collaborative site in four languages (French, English, Arabic and Persian) and a TV show on France 24. They cover international current affairs by using eyewitness accounts from people who were at the heart of events. Eyewitnesses send photos and videos, which *Les Observateurs*' team of professional journalists in Paris verify and contextualize.
- ❑ [Les Décodeurs](#), Le Monde's dedicated fact checking unit, verifies declarations, assertions and rumours of all kinds. It has a [ten-point charter](#), which shapes their work. They also built [Decodex](#), a browser extension that helps fight false information by informing readers when they see an article that comes from a site that frequently posts fabricated or misleading content.
- ❑ [LuiPresident.fr](#) is a financially and politically independent website created in 2012 by three students from the Ecole Supérieure de Journalisme de Lille (ESJ-Lille) to verify Francois Hollande's campaign commitments during his term in office. In 2017, *LuiPresident* was transformed into a project of the ESJ-Lille to follow Emmanuel Macron's promises during the five-year period of his presidency 2017-2022. It is run by students under the supervision of professional journalists.
- ❑ [CrossCheck](#)²⁵³ was a collaborative news verification project led by FirstDraft and funded by Google News. It involved 37 different partners, including newsrooms and technology companies.

Georgia: [FactCheck Georgia](#) is a political news-and-information project established by Georgia's Reforms Associates (GRASS). Modelled on successful international political news-and-information watchdog services, it aims to rate the factual accuracy of statements made by Members of Parliament (MPs), the Prime Minister, the President and the Government economic team. The service is offered in a fully bilingual Georgian-English format. FactCheck is implemented with the support of the Embassy of the Kingdom of the Netherlands in Tbilisi, the German Marshall Fund, the European Endowment for Democracy and the US Embassy in Georgia.

Germany:

- ❑ [Fakt oder Fake](#) is a vertical on Zeit Online.
- ❑ [Faktenfinder](#) is a collaboration between regional members of the German public broadcasting consortium ARD, who contribute multimedia fact checks and explainers to investigate widely spread claims and rumours from across the country and world. The project is an initiative of tagesschau24, the digital TV news channel produced for ARD by one of its regional partners, Norddeutscher Rundfunk (NDR) in Hamburg.

²⁵³ <https://crosscheck.firstdraftnews.com/france-en/>

- ❑ [Correctiv](#) is an independent fact-checking site. It was launched to focus on political claims and viral mis-information with financial support from the Open Society Foundation. It was Facebook's German partner for its third party fact-checking initiative. It also partnered with First Draft to monitor dis-information around the 2017 German Federal election.

Ireland: [theJournal.ie](#), a commercial digital news outlet, started a fact-checking project during the county's February 2016 general election and continued after the result. It often focuses on claims flagged by its readers.

Italy:

- ❑ [Pagella Politica](#) is a website that fact-checks claims by politicians. It also produces fact-checks for *Virus*, a public affairs programme on RAI.
- ❑ [Agi Fact-Checking](#), a service of the Italian news service Agi, distributes reports by Pagella Politica.

Kosovo²⁵⁴: [Kryptometer](#) is a fact-checking vertical on Kallxo.com and produced for the video program "Jeta në Kosovë". Both initiatives are projects of the Balkans Investigative Reporting Network (Rrjeti Ballkanik i Gazetarisë Hulumtuese) and Internews Kosova, a non-governmental organization that supports regional media projects and training. Kallxo is an online accountability platform for Kosovo citizens that was funded by the United Nations Development Programme.

Latvia: [Melu Detektors](#) is hosted by LSM.lv, the digital news portal for Latvijas Sabiedriskie Mediji (Latvian Public Media), which includes TV and radio channels. The project operates on its own using a combination of government funding and advertising. It began as a partnership with Re:Baltica, a non-profit investigative news outlet supported by the Open Society Institute in Latvia, and launched with support from the U.S. Baltic Foundation and the U.S. State Department.

Lithuania: [Patikrinta](#) is a branch of the established Lithuanian news website '15 min' that focuses on statements by Lithuanian politicians, often based on recommendations from readers.

Macedonia: [Vistinomer](#) is a fact-checking website run by the Macedonian NGO Metamorphosis. It is connected to Why not? from Bosnia and Herzegovina and LINnet from Serbia.

Norway: [Faktisk](#) is a new fact-checking collaboration between the four biggest news organizations in Norway.

Poland:

- ❑ [Demogog](#) is a project of the Demogog Association, the first professional fact-checking organization in Poland.
- ❑ [OKO.press](#) fact-checks statements made by Polish public figures. It was founded in 2016 and is supported entirely by individual donations.

²⁵⁴ All reference to Kosovo, whether to the territory, institutions or population, in this text shall be understood in full compliance with United Nations Security Council Resolution 1244 and without prejudice to the status of Kosovo.

Portugal: [Fact Checks do Observador](#) is a reporting project of El Observador, a free, ad-supported online newspaper based in Lisbon.

Romania: [Factual](#) is run by volunteer contributors and a team at Funky Citizens, an accountability and budget transparency organization. It is funded by voluntary contributions and a grant from the Alumni Grants Programme of the Open Society Institute.

Serbia: [Istinomer](#) was established in 2009 by the Center for Research, Transparency and Accountability.

Slovakia: [Demagoq](#) was founded in 2010 by a pair of political science students at Masaryk University in Brno, and quickly spread to sister sites in the Czech Republic and Poland.

Spain

- ❑ [El Objetivo con Ana Pastor](#) is a highly rated weekly public affairs programme on the Spanish television network La Sexta, which goes out to between 1.5 and 2 million viewers each Sunday.
- ❑ [Maldito Buló](#) is an online-only fact-checking initiative linked to El Objetivo.
- ❑ [La Chistera](#) is a blog published by the data journalism unit at El Confidencial, a commercial digital news service based in Madrid and operated by Titania Compañía Editorial SL.

Sweden: [Viralgranskaren](#) is the fact-checking project of the Swedish newspaper *Metro*.

Switzerland:

- ❑ [Swissinfo.ch](#) is the fact-checking initiative of the Swiss Broadcasting Corporation. Reports are posted in multiple languages and appear most frequently during the country's voting periods.
- ❑ [Tages Anzeiger Faktenchecks](#) is a project of the Swiss German-language daily newspaper *Tages Anzeiger*. The frequency of its fact-checks increases during periods of public campaigning and political debate.

Turkey:

- ❑ [Do ruluk Payı](#) (or 'share of truth') is a fact-checking initiative established by the Dialogue for Common Future Association. It is funded by the National Endowment for Democracy.
- ❑ [Teyit](#) is a verification and debunking service that monitors news reports and social media for mis- and dis-information. The initiative is a non-profit, social enterprise based in Ankara and is supported by the European Endowment for Democracy.

Ukraine:

- ❑ [StopFake](#) was founded by students and faculty of the Kyiv Mohyla School of Journalism. The site is now in 11 different languages.
- ❑ [Slovo i Dilo](#) was created by a non-governmental organization called the People's Control System to track the political promises of Ukrainian officials at the national and local level.
- ❑ [VoxCheck](#) is a branch of VoxUkraine, which does research-based policy analysis and is funded in part by the National Endowment for Democracy. It fact-checks Ukrainian

politicians and has utilized crowd-funding for much of its support.

- ❑ [FactCheck Ukraine](#) is an independent fact-checking initiative that examines claims by Ukrainian politicians and public figures is funded by individual donations and non-governmental organizations.

United Kingdom:

- ❑ [Full Fact](#) is the UK's largest independent, non-partisan fact-checking organization. As well as publishing fact-checks, the organisation actively pushes for corrections where necessary and works with government departments and research institutions to improve the quality and communication of information at source. It is supported by individuals, charitable trusts and foundation support. They recently received money from the Omidyar Network to build on their work on automated fact-checking. They partnered with First Draft to monitor mis-information during the 2017 UK snap election.
- ❑ [FactCheck](#) is a fact-checking feature from Channel 4 News, which appears on their website.
- ❑ [Reality Check](#) is the BBC's dedicated fact-checking project. It was introduced in 2015 to cover the Brexit referendum and then re-started in the autumn of 2016. Fact-checkers for the BBC's Reality Check appear on high-reach outlets, including the BBC News Channel, BBC World Television, Radio 5 Live, Radio 4 and the BBC World Service.
- ❑ [The Ferret](#) is an investigative news organisation based in Scotland. In the Spring of 2017, they launched Fact Service, Scotland's first non-partisan fact-checking service. They check statements from politicians, pundits and prominent public figures. Fact checks are also undertaken on some viral claims, hoaxes and memes.
- ❑ [FactCheckNI](#) is an independent fact-checking organisation. They provide tools, information and advice to citizens so they can undertake their own fact-checks on information they hear from politicians and and the media.

References

- Alba, D. (2016) Meet the Ad Companies Ditching Breitbart and Fake News, *Wired*, December, 15, 2016. Available at: <https://www.wired.com/2016/12/fake-news-will-go-away-tech-behind-ads-wont-pay/>
- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w23089> Political Behavior
- Bakamo (2017a) The Role and Impact of Non-Traditional Publishers in the 2017 French Presidential Election. Available at: <https://www.bakamosocial.com/frenchelection/>
- Bakamo (2017b) Patterns of Dis-information in the 2017 French Presidential Election. Available at: Available at: <https://www.bakamosocial.com/frenchelection/>
- Barber, B. (1998) 'Which Technology and Which Democracy?', Talk given at Democracy and Digital Media Conference, <http://web.mit.edu/m-i-t/articles/barber.html>
- Bell, E. (2016) Facebook drains the fake news swamp with new, experimental partnerships, *Columbia Journalism Review*, December 15, 2016. Available at: https://www.cjr.org/tow_center/facebook_drains_fake_news_swamp_new_experimental_partnerships.php
- Benedictus, L., "Invasion of the troll armies: from Russian Trump supporters to Turkish state stooges," *The Guardian*. <https://www.theguardian.com/media/2016/nov/06/troll-armies-social-media-trump-russian>
- Benkler Y. et al (March 3, 2017) Study: Breitbart-led right-wing media ecosystem altered broader media agenda. *Columbia Journalism Review*. <https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>
- Beran, Dale. "4chan: The Skeleton Key to the Rise of Trump." *Medium*, February 14, 2017. <https://medium.com/@DaleBeran/4chan-the-skeleton-key-to-the-rise-of-trump-624e7cb798cb>
- Bernstein, J. (April 5, 2017) Never Mind The Russians, Meet The Bot King Who Helps Trump Win Twitter, *Buzzfeed News*, <https://www.buzzfeed.com/josephbernstein/from-utah-with-love/>
- Bessi, A. (2015) Science vs Conspiracy: Collective Narratives in the Age of Mis-information. *PLoSOne*, 10, (2)
- Bharat, K. (April 27, 2017) How to Detect Fake News in Real-Time, *NewsCo on Medium*, <https://shift.newco.co/how-to-detect-fake-news-in-real-time-9fdae0197bfd>

Bilton, N. (Jan. 26, 2017) Fake News is About to Get Scarier than You'd Ever Dreamed. *Vanity Fair*. <http://www.vanityfair.com/news/2017/01/fake-news-technology>

Bilton, R. (Feb. 2, 2017) Reddit's /r/worldnews community used a series of nudges to push users to fact-check suspicious news, *Nieman Lab*, <http://www.niemanlab.org/2017/02/reddits-rworldnews-community-used-a-series-of-nudges-to-push-users-to-fact-check-suspicious-news/>

Birdsell, D. S., & Groarke, L. (1996). Toward a Theory of Visual Argument. *Argumentation and Advocacy*, 33(1), 1-10.

Born, K. (2017) The future of truth: Can philanthropy help mitigate mis-information? *Hewlett Foundation*. Available at: <http://www.hewlett.org/future-truth-can-philanthropy-help-mitigate-mis-information/>

Bordia, P. (2013). Rumor Clustering, Consensus, and Polarization: Dynamic Social Impact and Self-Organization of Hearsay. *Journal of Experimental Social Psychology*, 49(3), 378-399.

boyd, d.. (2017) "Did Media Literacy Backfire?" *Data & Society: Points*, January 5, 2017. <https://points.datasociety.net/did-media-literacy-backfire-7418c084d88d>

boyd, d. (March 27, 2017) "Google and Facebook can't just make fake news disappear", *Backchannel*, <https://medium.com/backchannel/google-and-facebook-cant-just-make-fake-news-disappear-48f4b4e5fbe8>

Bond, D. and Robinson, D. (Jan. 29, 2017), 'European Commission fires warning at Facebook over fake news' *Financial Times*, <https://www.ft.com/content/85683e08-e4a9-11e6-9645-c9357a75844a?mhq5j=e2>

Borel, B. (2017) Fact Checking Won't Save Us From Fake News, *FiveThirtyEight*, January 4, 2017. Available at: <https://fivethirtyeight.com/features/fact-checking-wont-save-us-from-fake-news/>

Borel, B. (2017) How to Talk to your Facebook Friends about Fake News, *Open Notebook* <http://www.theopennotebook.com/2017/02/21/how-to-talk-to-your-facebook-friends-about-fake-news/>

Boyd, L. & Vraga, E (2015) In Related News, That Was Wrong: The Correction of Mis-information Through Related Stories Functionality in Social Media, *Journal of Communication*, 65 (4): 619-638.

Bradshaw, S. and P. Howard (August 2017) Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/07/Troops-Trolls-and-Troublemakers.pdf>

Broderick, R. (2017) Trump Supporters Online Are Pretending To Be French To Manipulate

France's Election, *Buzzfeed News*, January 24, 2017. Available at:
https://www.buzzfeed.com/ryanhatesthis/inside-the-private-chat-rooms-trump-supporters-are-using-to?utm_term=.aaMXl5lvgG#.crwErDrpYj

Brown, C. (April 2, 2017) Introducing the News Integrity Initiative, *Facebook Media Blog*,
<https://media.fb.com/2017/04/02/introducing-the-news-integrity-initiative/>

Bucay, Y., V. Elliott, J. Kamin, A. Park, America's Growing News Deserts, *Columbia Journalism Review*, https://www.cjr.org/local_news/american-news-deserts-donuts-local.php

Caplan, Robyn. "How Do You Deal with a Problem Like 'Fake News?'" *Data & Society: Points*, January 5, 2017. <https://points.datasociety.net/how-do-you-deal-with-a-problem-like-fake-news-80f9987988a9>

Carey, J. (1989) *Communication as Culture: Essays on Media and Society*, London: Routledge.

Caulfield, M. (Dec 7, 2016) Announcing the Digital Polarization Initiative. *Hapgood*.
<https://hapgood.us/2016/12/07/announcing-the-digital-polarization-initiative-an-open-pedagogy-joint/>

Cellan-Jones, R. (Sept 22, 2017) Fake news worries 'are growing' suggests BBC poll, *BBC News*,
<http://www.bbc.com/news/technology-41319683>

Chan, M.S., C. R. Jones, K.H. Jamieson, D. Albarracín (2017) Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Mis-information, *Psychological Science*, 1-16.

Chesney, S., M. Liakata, M. Poesio and M. Purver (2017) Incongruent Headlines: Yet Another Way to Mislead Your Readers, <http://www.eecs.qmul.ac.uk/~mpurver/papers/chesney-et-al17nlpj.pdf>

Chu, Z., S. Gianvecchio, H. Wang, S. Jajodia, (Nov/Dec 2012) Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg? *IEEE Transactions on Dependable and Secure Computing*, Vol. 9, No. 6

Collins, B., K. Poulsen & S. Ackerman (September 11, 2017) Exclusive: Russia Used Facebook Events to Organize Anti-Immigrant Rallies on U.S. Soil, *The Daily Beast*,
<http://www.thedailybeast.com/exclusive-russia-used-facebook-events-to-organize-anti-immigrant-rallies-on-us-soil>

Constine, J. (2017) Facebook shows Related Articles and Fact Checkers Before you Open Links, *TechCrunch*, <https://techcrunch.com/2017/04/25/facebook-shows-related-articles-and-fact-checkers-before-you-open-links/>

Crowell, C., (June 14, 2017) Our Approach to Bots and Automation, Twitter Blog.
https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Mis-information.html

d'Ancona, M. (2017). *Post-Truth: The New War on Truth and How to Fight Back*. London: Ebury Press.

de Benedictis-Kessner, J., Baum, M.A., Berinsky, A.J., & Yamamoto, T. "Persuasion in Hard Places: Accounting for Selective Exposure When Estimating the Persuasive Effects of Partisan Media." Unpublished Manuscript, Harvard University and MIT.

De Feyter, S. (2015). 'They are like Crocodiles under Water': Rumour in a Slum Upgrading Project in Nairobi, Kenya. *Journal of Eastern African Studies*, 9(2), 289-306

De Seta, G., (2017) "Trolling, and Other Problematic Social Media Practices," in *The SAGE Handbook of Social Media*, ed. Jean Burgess, Alice E. Marwick, and Thomas Poell. Thousand Oaks, CA: SAGE Publications.

Derakhshan H. (July 14, 2015) The Web We Have to Save. *Matter*. Available at: <https://medium.com/matter/the-web-we-have-to-save-2eb1fe15a426>

Derakhshan H. (Nov 29, 2016) Social Media Is Killing Discourse Because It's Too Much Like TV, *MIT Technology Review*. <https://www.technologyreview.com/s/602981/social-media-is-killing-discourse-because-its-too-much-like-tv/>

Dias, N. (June 22, 2017), Reporting on a new age of digital astroturfing, *First Draft News*, <https://firstdraftnews.com/digital-astroturfing/>

Dias, N. (August 17, 2017) The Era of Whatsapp Propaganda is Upon Us, *Foreign Policy*, <http://foreignpolicy.com/2017/08/17/the-era-of-whatsapp-propaganda-is-upon-us/>

DiFonzo, N., Beckstead, J. W., Stupak, N., & Walders, K. (2016). Validity judgments of rumors heard multiple times: the shape of the truth effect. *Social Influence*, 11(1), 22-39.

DiFonzo, N., Bourgeois, M. J., Suls, J., Homan, C., Stupak, N., Brooks, B. P., ... & Doctor, K., (May 2, 2017) Can a Master Blacklist Choke Off Fake News, *Nieman Lab*. <http://www.niemanlab.org/2017/05/newsonomics-can-a-master-blacklist-choke-off-fake-news-money-supply/>

Rachel Donadio, (May 8, 2017) Why the Macron Hacking Attack Landed with a Thud in France, *The New York Times*, <https://www.nytimes.com/2017/05/08/world/europe/macron-hacking-attack-france.html>

Dutton, B. (May 2017) Fake News, Echo chambers and Filter Bubbles are an Exaggerated Threat. Here's Why. *World Economic Forum Blog*. <https://www.weforum.org/agenda/2017/05/fake-news-echo-chambers-and-filter-bubbles-are-an-exaggerated-threat-heres-why>

Dutton, William H., et al. (2017) "Search and Politics: A Cross-National Survey." Quello Center Working Paper No. 2944191 Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2944191

Eddy, M. and M. Scott (June 30 2017), Delete Hate Speech or Pay Up, Germany Tells Social Media Companies, *New York Times*,
<https://www.nytimes.com/2017/06/30/business/germany-facebook-google-twitter.html>

El-Bermawy, M. (Nov. 18, 2016) Your Filter Bubble is Destroying Democracy, *Wired*.
<https://www.wired.com/2016/11/filter-bubble-destroying-democracy/>

Eliseev, A (July 20, 2017) The Gupta scandal: how a British PR firm came unstuck in South Africa, *The New Statesman*,
<http://www.newstatesman.com/culture/observations/2017/07/gupta-scandal-how-british-pr-firm-came-unstuck-south-africa>

EU East StratCom Task Force, Means, Goals and Consequences of the Pro-Kremlin Dis-information Campaign, *Italian Institute for International Political Studies (ISPI)*, January 19, 2017. Available at: <https://euvsdisinfo.eu/commentary-means-goals-and-consequences-of-the-pro-kremlin-dis-information-campaign/>

Evaluating Information: The Cornerstone of Civic Online Reasoning. *Stanford History Education Group*. 2016. Available at:
<https://sheg.stanford.edu/upload/V3LessonPlans/Executive%20Summary%2011.21.16.pdf>

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). *The Rise of Social Bots*. *Communications of the ACM*, 59(7), 96-104.

Filloux, F. (August 21, 2017) More than 600 global brands still feed the fake news ecosystem, *The Monday Note on Medium*, <https://mondaynote.com/more-than-600-global-brands-still-feed-the-fake-news-ecosystem-d1ddfbd80458>

Filloux, F. (August 6, 2017) You can't sell news for what it costs to make, *The Walkley Magazine on Medium*, <https://medium.com/the-walkley-magazine/you-cant-sell-news-for-what-it-costs-to-make-7a4def964ffa>

Filloux, F. (June 25, 2017) The News Quality Scoring Project, *Medium*,
<https://mondaynote.com/the-news-quality-scoring-project-surfacing-great-journalism-from-the-web-48401ded8b53>

Flynn, D.J. & Nyhan, Brendan & Reifler, Jason. (2017). 'The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics', *Political Psychology*. 38: 127-150

Gillespie, T. (in press) Governance of and by platforms. In Burgess, J., Poell, T., & Marwick, A. (Eds), *SAGE Handbook of Social Media*. Retrieved from <http://culturedigitally.org/wp-content/uploads/2016/06/Gillespie-Governance-ofby-Platforms-PREPRINT.pdf>

Goffman, E, (1956) *The Presentation of Self in Everyday Life*. Random House.

Goodman, E. (2017) Write up from the Westminster Media Forum Keynote Seminar: Fake news - scope, public trust and options for policy, LSE, <http://blogs.lse.ac.uk/mediapolicyproject/2017/08/10/the-evolving-conversation-around-fake-news-and-potential-solutions/>

Granovetter, M.S. (1973) The Strength of Weak Ties, *Journal of Sociology*, 78(6):1360-1380

Graves, L.& Cherubini, F. (2016). The Rise of Fact-Checking Sites in Europe. Available at: <http://reutersinstitute.politics.ox.ac.uk/publication/rise-fact-checking-sites-europe>

Greenhill, K. M. (forthcoming). Whispers of War, Mongers of Fear: Extra-factual Sources of *Threat Conception and Proliferation*

Greenhill, K. M., & Oppenheim B. (forthcoming). Rumor Has It: The Adoption of Unverified Information in Conflict Zones. *International Studies Quarterly*.

Gu, L., Kropotov, V. & Yarochkin, F. (2017) The Fake News Machine: How Propagandists Abuse the Internet and Manipulate the Public. Oxford University. Available at: https://documents.trendmicro.com/assets/white_papers/wp-fake-news-machine-how-propagandists-abuse-the-internet.pdf

Habermas, J. (1962) (1989 translation, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. MIT Press.

Haigh, M., Haigh, T., & Kozak, N. I. (2017). Stopping Fake News: The work practices of peer-to-peer counter propaganda. *Journalism Studies*, 1-26.

Hall, S. (1973). *Encoding and Decoding in the Television Discourse*. Birmingham: Centre for Contemporary Cultural Studies

Hansen, F. S. (August 2017) Russian Hybrid Warfare: A Study of Dis-information, *Danish Institute for International study*. http://pure.diis.dk/ws/files/950041/DIIS_RP_2017_6_web.pdf

Harrington, M (Jan. 16, 2017) Survey: People's Trust Has Declined in Business, Media, Government, and NGOs, *Harvard Business Review*, <https://hbr.org/2017/01/survey-peoples-trust-has-declined-in-business-media-government-and-ngos>

Helberger, N. et al. (Sept 22. 2017) Dear Mark: An Open Letter to Mark Zuckerberg in Response to his Statement on Political Advertising on Facebook, Available at: https://www.dropbox.com/s/7v3vpk9yw5sa18b/Dear%20Mark_final..pdf?dl=0

Hendrix, J. and Carroll, D. (2017) Confronting a Nightmare for Democracy. Available at: <https://medium.com/@profcarroll/confronting-a-nightmare-for-democracy-5333181ca675>

Higgins, A., McIntire, M., & Dance, G. J. (2016). Inside a Fake News Sausage Factory: 'This is all about income'. *New York Times*, November 25, 2016. Available at:

<https://www.nytimes.com/2016/11/25/world/europe/fake-news-donald-trump-hillary-clinton-georgia.html>

Hofileña, C. F. (Oct. 9, 2016) Fake accounts, manufactured reality on social media, *Rappler*, <https://www.rappler.com/newsbreak/investigative/148347-fake-accounts-manufactured-reality-social-media>

Iyengar S. and S. J. Westwood (2015) Fear and Loathing across Party Lines: New Evidence on Group Polarization, *American Journal of Political Science* Vol. 59, No. 3 (July 2015), pp. 690-707

Jack, C. (2017) Lexicon of Lies, *Data & Society*, https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf

Jackson, J. (Jan. 12, 2017) BBC sets up team to debunk fake news, *The Guardian*, <https://www.theguardian.com/media/2017/jan/12/bbc-sets-up-team-to-debunk-fake-news>

Jankowicz, N. (Sept. 25, 2017) The Only Way to Defend Against Russia's Information War, *The New York Times*, <https://mobile.nytimes.com/2017/09/25/opinion/the-only-way-to-defend-against-russias-information-war.html>

Jarvis, J. (2017) Our problem isn't 'fake news.' Our problems are trust and manipulation. *Whither News*. Available at: <https://medium.com/whither-news/our-problem-isnt-fake-news-our-problems-are-trust-and-manipulation-5bfbcd716440>

Jolley, D., & Douglas, K. M. (2014) "The Social Consequences of Conspiracism: Exposure to Conspiracy Theories Decreases the Intention to Engage in Politics and to Reduce One's Carbon Footprint." *British Journal of Psychology*, 105 (1): 35–56.

Kahan, D. et al (2013) Motivated Numeracy and Enlightened Self-Government, *Behavioural Public Policy*, 1, 54-86

Kahan, D. (2011) What is Motivated Reasoning and How Does It Work? *Science and Religion Today*, <http://www.scienceandreligiontoday.com/2011/05/04/what-is-motivated-reasoning-and-how-does-it-work/>

Karlova, N. A., & Fisher, K. E. (2013). Plz RT: a Social Diffusion model of Mis-information and Dis-information for Understanding Human Information Behaviour. *Information Research*, 18(1), 1-17.

King, G, J. Pan & M. Roberts, (May 2016) How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument, Harvard University, <http://gking.harvard.edu/files/gking/files/50c.pdf?m=1463587807>

Kington, T. (March 30, 2010) Twitter hoaxer comes clean and says: I did it to expose weak media, *The Guardian*, <https://www.theguardian.com/technology/2012/mar/30/twitter-hoaxer-tommaso-de-benedetti>

Knight Commission on the Information Needs of Communities in a Democracy, (2009) *Informing Communities: Sustaining Democracy in the Digital Age*, *The Aspen Institute*, <https://production.aspeninstitute.org/publications/informing-communities-sustaining-democracy-digital-age/>

Krasodonski-Jones, A. (2016) Talking to Ourselves? Political Debate Online and the Echo Chamber Effect. *Demos*. Available at: <https://www.demos.co.uk/project/talking-to-ourselves/>

Lakoff, G. (2010) "Why "Rational Reason" Doesn't Work in Contemporary Politics", *Truth Out*. <http://www.truth-out.org/buzzflash/commentary/george-lakoff-why-rational-reason-doesnt-work-in-contemporary-politics/8893-george-lakoff-why-rational-reason-doesnt-work-in-contemporary-politics>

Lakoff, George (1996) *Moral Politics: What Conservatives Know That Liberals Don't*. University of Chicago Press.

Lakoff, George (1997). *Metaphors We Live By*. University of Chicago Press

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. Available at: <https://shorensteincenter.org/combating-fake-news-agenda-for-research/>

Lev-on, A. and Manin, Bernard, (2009) Happy Accidents: Deliberation and Online Exposure to Opposing Views, *Online Deliberation: Design, Research, and Practice* (edited by Todd Davies and Seeta Peña Gangadharan)

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Mis-information and its correction continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106-131.

Lewandowsky, S., Stritzke, W. G., Freund, A. M., Oberauer, K., & Krueger, J. I. (2013). Mis-information, dis-information, and violent conflict: From Iraq and the "War on Terror" to future threats to peace. *American Psychologist*, 68(7), 487.

Lin, X., Spence, P. R., & Lachlan, K. A. (2016). Social media and credibility indicators: The Effect of Influence Cues. *Computers in Human Behavior*, 63, 264-271.

Lindgren, A., J. Corbett & J. Hodson, (Jan. 23, 2017) Canada's Local News Poverty, *Policy Options*, <http://policyoptions.irpp.org/magazines/january-2017/canadas-local-news-poverty/>

Maffesoli, M. (1996) *The Time of the Tribes*, London: Sage.

Marwick, A. and Lewis, R (2017) Media Manipulation and Dis-information Online. *Data and Society*. Available at: <https://datasociety.net/output/media-manipulation-and-disinfo-online/>

Mantzaris, A (Feb. 2, 2016) Fact-checkers experiment with Snapchat, GIFs and other stuff millennials ♥. *Poynter*. <http://www.poynter.org/2016/fact-checkers-experiment-with-snapchat-gifs-and-other-stuff-millennials-%E2%99%A5/393992/>

Mantzaris, A (May 10, 2016) Did Clinton or Trump twist the facts? This messaging bot will tell you. *Poynter*. <http://www.poynter.org/2016/did-clinton-or-trump-twist-the-facts-this-messaging-bot-will-tell-you/409457/>

Mantzaris, A (June 7, 2016) There's been an explosion of international fact-checkers, but they face big challenges, *Poynter*. <http://www.poynter.org/2016/theres-been-an-explosion-of-international-fact-checkers-but-they-face-big-challenges/415468/>

Mantzaris, A. (July 12, 2017), European policy-makers are not done with Facebook, Google and fake news just yet, *Poynter*, <https://www.poynter.org/2017/european-policy-makers-are-not-done-with-facebook-google-and-fake-news-just-yet/465809/>

McGregor, J. (May 31, 2017) Facebook Wades Into Another Election, *Forbes*. <https://www.forbes.com/sites/jaymcgregor/2017/05/31/facebook-wades-into-another-election/#37265336635c>

McGrew, S., T. Ortega, J. Breakstone & S. Wineburg, (Fall 2017) The Challenge That's Bigger Than Fake News: Teaching Students to Engage in Civic Online Reasoning. *American Educator*.

Meeker, M (2017) Internet Trends, 2017. Available at: <http://dq756f9pzlyr3.cloudfront.net/file/Internet+Trends+2017+Report.pdf>

Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, 41(8), 1042-1063.

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, 59, 210-220.

Metzger et al. (2010) Social and Heuristic Approaches to Credibility Evaluation Online, *Journal of Communication*, 60 (3):413-439

Mina, A.X. (March 2, 2017) Building Technical Standards for Credibility, MisinfoCon on Medium. <https://misinfocon.com/building-technical-standards-for-credibility-59ef9ee4ab73>

Moschella, M. and R. Watts, (June 19, 2017) What we Learned Fact-Checking the UK Election, *First Draft Newst*, <https://firstdraftnews.com/joint-venture-learnings/>

Mosseri, A. (2016) News Feed FYI: Addressing Hoaxes and Fake News, *Facebook Newsroom*, December 15, 2016.

Mullin, B. (April 6, 2016) Facebook rolls out global warning against fake news, *Poynter*

<http://www.poynter.org/2017/facebook-rolls-out-global-warning-against-fake-news/454951/>

Musgrave, S. (August 9, 2017) I Get Called a Russian Bot 50 Times a Day, *Politico*, <http://www.politico.com/magazine/story/2017/08/09/twitter-trump-train-maga-echo-chamber-215470>

NATO Strategic Communications Centre of Excellence (September 2017) Robotrolling, <http://www.stratcomcoe.org/robotrolling-20171>

Newman, N. (2017) Digital News Report, *Reuters Institute for the Study of Journalism*, https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf

Nielsen, J. (2003). "IM, Not IP (Information Pollution)". *ACM Queue*. 1 (8): 75–76.

Nyhan, B. and J. Reifler (2015) 'Displacing Mis-information about Events: An Experimental Test of Causal Corrections', *Journal of Experimental Political Science*, 2 (1) : 81-93

Nyhan, B., and J. Reifler (2010) "When Corrections Fail: The persistence of political misperceptions." *Political Behavior*, 32 (2): 303–330.

Oyeyemi, S. O., Gabarron, E., & Wynn, R. (2014). Ebola, Twitter, and Mis-information: a Dangerous Combination?. *BMJ*, 349.

Pariser, E. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. London: Penguin.

Pariser, E (2016) & many others, Design Solutions for Fake News, [November 2016 ongoing]. Available at: https://docs.google.com/document/d/1OPghC4ra6QLhaHhW8QvPJRMKGEXT7KaZtG_7s5-UQrw/edit

Paul, C. & Matthews, M. (2016) The Russian 'Firehose of Falsehood' Propaganda Model: Why It Might Work and Options to Counter It, *Rand Corporation*, July 2016. Available at: http://www.rand.org/content/dam/rand/pubs/perspectives/PE100/PE198/RAND_PE198.pdf

Pennycook, G. et al (July 5, 2017) Prior Exposure Increases Perceived Accuracy of Fake News, Available at SSRN: <https://ssrn.com/abstract=2958246>

Picone, I. (2015) Impression Management in Social Media, Published Online: 11 FEB 2015. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/9781118767771.wbiedcs071/abstract>

Politi, J. (Dec. 30, 2016) 'Italy antitrust chief urges EU to help beat fake news'. *Financial Times*, <https://www.ft.com/content/e7280576-cddc-11e6-864f-20dcb35cede2?mhq5j=e2>

Pondsford, D. (March, 31, 2017) The decline of local journalism is a far greater threat to media plurality than Rupert Murdoch, *The Press Gazette*, <http://www.pressgazette.co.uk/the-decline-of-local-journalism-is-a-far-greater-threat-to-media-plurality-than-rupert-murdoch/>

Posetti, P. (July 2017) This is why Online Harassment Still Needs Attention, *MediaShift* <http://mediashift.org/2017/07/online-harassment-still-needs-attention/>

Postman, N. (1985). *Amusing Ourselves to Death: Public Discourse in the Age of Television*. New York: Viking.

Potter, M. C. (2014). Detecting and remembering briefly presented pictures. In K. Kveraga & M. Bar (Eds.), *Scene Vision* Cambridge, MA: MIT Press, pp. 177-197.

Prior, M. (2002). Any good news in soft news? The impact of soft news preference on political knowledge. *Political Communication*, 20(2), pp.149-171.

Qiu, X., D. F. M. Oliveira, A.S. Shirazi, A. Flammini and F. Menczer (2017) Limited Individual Attention and Online Virality of Low-Quality Information, *Nature Human Behaviour*, 1

Quattrociochi, W. (14 Jan, 2016) How does mis-information spread online? *World Economic Forum Blog*, <https://www.weforum.org/agenda/2016/01/q-a-walter-quattrociochi-digital-wildfires/>

Ratkiewicz et al., (2011) Detecting and Tracking Political Abuse in Social Media, *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2850>

Renner, N. (January 30, 2017) Memes trump articles on Breitbart's Facebook page, *Columbia Journalism Review*. https://www.cjr.org/tow_center/memes-trump-articles-on-breitbart-facebook-page.php

Resnick, B. (May 6, 2017) 7 Psychological Concepts that Explain the Trump Era of Politics, *Vox*, <https://www.vox.com/science-and-health/2017/3/20/14915076/7-psychological-concepts-explain-trump-politics>

Resnick, B. (July 10, 2017) "Trump supporters know Trump lies. They just don't care." *Vox*, <https://www.vox.com/2017/7/10/15928438/fact-checks-political-psychology>

Rutenberg, J. (Sept. 13, 2017) RT, Sputnik and Russia's New Theory of War, *New York Times*, <https://www.nytimes.com/2017/09/13/magazine/rt-sputnik-and-russias-new-theory-of-war.html?>

Salganik, M. et al. (2006) Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market, *Science*, Vol. 311, pp.854-856

Schmidt, C. (Aug. 3, 2017) Games might be a good tool for fighting fake news. Here's what three developers have learned, *Nieman Lab*, <http://www.niemanlab.org/2017/08/games-might-be-a-good-tool-for-fighting-fake-news-heres-what-three-developers-have-learned/>

Schwartz, J. (May 24, 2016) Most Popular Messaging App in Every Country, *Market Intelligence Blog*, <https://www.similarweb.com/blog/worldwide-messaging-apps>

Shaffer, K, Carey, CE & Starling, B. (2017) *Democracy Hacked*. Available at: <https://medium.com/data-for-democracy/democracy-hacked-a46c04d9e6d1>

Shahari, A. (Nov 11. 2017) Zuckerberg Denies Facebook's Impact on the Election, *NPR*, <http://www.npr.org/sections/alltechconsidered/2016/11/11/501743684/zuckerberg-denies-fake-news-on-facebook-had-impact-on-the-election>

Shao, C., G.L. Ciampaglia, O. Varol, A. Flammini and F. Menczer, (July 24, 2017) The Spread of Fake News by Social Bots, <https://arxiv.org/pdf/1707.07592.pdf>

Shao, C. et al., (2016) Hoaxy: A Platform for Tracking Online Mis-information, *Proceedings of the 25th International Conference Companion on World Wide Web*, pp. 745-750

Shukla, S. (August 28, 2017) Blocking Ads from Pages that Repeatedly Share False News. *Facebook Newsroom*. <https://newsroom.fb.com/news/2017/08/blocking-ads-from-pages-that-repeatedly-share-false-news/>

Silverman, C. (2015), Lies, Damn Lies and Viral Content, *Tow Center for Digital Journalism*. Available at: http://towcenter.org/wp-content/uploads/2015/02/LiesDamnLies_Silverman_TowCenter.pdf

Silverman, C. (2016a) Facebook Must Either Innovate or Admit Defeat At The Hands Of Fake News Hoaxsters, *Buzzfeed News*, August 30, 2016. Available at: https://www.buzzfeed.com/craigsilverman/how-facebook-fell-into-a-fake-news-trap-of-its-own-making?utm_term=.tpdEQ2QJnN#.poVnBOBx2E

Silverman, C. (2016b) This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook, *Buzzfeed News*, November 16, 2016. Available at: https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.aaMXl5lvgG#.wtVW3y3jYn

Silverman et al. (August 8, 2017) Inside the Partisan Fight for your NewsFeed, *Buzzfeed News*, <https://www.buzzfeed.com/craigsilverman/inside-the-partisan-fight-for-your-news-feed/>

Silverman, C. et al., (2016) Hyperpartisan Facebook Pages Are Publishing False And Misleading Information At An Alarming Rate, *Buzzfeed News*, October 20, 2016. Available at: https://www.buzzfeed.com/craigsilverman/partisan-fb-pages-analysis?utm_term=.lgVgaAa8Mo#.xmk3BQBoEp

Silverman, C. & Singer-Vine, J. (2016a). "Most Americans Who See Fake News Believe It, New

Survey Says." *BuzzFeed News*, December 6. Available at: https://www.buzzfeed.com/craigsilverman/fake-news-survey?utm_term=.uov6wAwgJK#.jvLRNKN136

Silverman, C. & Singer-Vine, J. (2016b). "The True Story Behind The Biggest Fake News Hit Of The Election.", *Buzzfeed News*, December 16. Available at: https://www.buzzfeed.com/craigsilverman/the-strangest-fake-news-empire?utm_term=.yrzPyEpLXq#.nqXD9N7opO

Silverman, C. & Alexander, L. (2016) "How Teens In the Balkans Are Duping Trump Supporters With Fake News." *Buzzfeed News*, November 3, 2016. Available at: https://www.buzzfeed.com/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo?utm_term=.ftzDj7japR#.ex0dMBM37O

Smith, L. C., Lucas, K. J., & Latkin, C. (1999). Rumor and gossip: Social discourse on HIV and AIDS. *Anthropology & Medicine*, 6 (1), 121-131.

Sontag S. (1977) *On Photography*, New York: Farrar Straus and Giroux.

Spencer, S. (Jan. 25, 2017) How we fought bad ads, sites and scammers in 2016, *Google Blog*, <https://www.blog.google/topics/ads/how-we-fought-bad-ads-sites-and-scammers-2016>

Stamos, A. (September 6, 2017) An Update on Information Operations On Facebook, *Facebook Newsroom*, <https://newsroom.fb.com/news/2017/09/information-operations-update/>

Stanley, S. (May 16, 2017) Mis-information and hate speech in Myanmar, *First Draft News*, <https://firstdraftnews.com/mis-information-myanmar/>

Stelzenmüller, C. (June 28, 2017) Testimony to the US Senate Committee on Intelligence: The Impact of Russian interference on Germany's 2017 Elections, Available at: <https://www.brookings.edu/testimonies/the-impact-of-russian-interference-on-germanys-2017-elections/>

Stelter, B. (Feb. 11, 2016) Apple CEO Tim Cook calls for "massive campaign" against fake news, *CNN*, <http://money.cnn.com/2017/02/11/media/fake-news-apple-ceo-tim-cook/index.html>

Stray, J. (Feb 27, 2017), Defense Against the Dark Arts: Networked Propaganda and Counter-Propaganda, *Tow Center*, Medium. <https://medium.com/tow-center/defense-against-the-dark-arts-networked-propaganda-and-counter-propaganda-deb7145aa76a>

Subramanian, S. (2017) Inside the Macedonian Fake News Complex, *Wired*, February 15, 2017. *Wired*. Available at: <https://www.wired.com/2017/02/veles-macedonia-fake-news>

Sunstein, C. R., Bobadilla-Suarez, S., Lazzaro, S. C., & Sharot, T. (2016). How People Update Beliefs about Climate Change: Good News and Bad News (*SSRN Scholarly Paper* No. ID

2821919). Rochester, NY: Social Science Research Network.

Sunstein, Cass R., and Adrian Vermeule. (2009) "Conspiracy Theories: Causes and Cures". *Journal of Political Philosophy*. 17 (2): 202–227.

Sullivan, M (Jan 6, 2017,) It's Time To Retire the Tainted Term Fake News, *The Washington Post*, https://www.washingtonpost.com/lifestyle/style/its-time-to-retire-the-tainted-term-fake-news/2017/01/06/a5a7516c-d375-11e6-945a-76f69a399dd5_story.html

Susman-Peña, T. and Vogt, Katya (June 12, 2017) Ukrainians' self-defense against information war: What we learned from Learn to Discern, *IREX*, <https://www.irex.org/insight/ukrainians-self-defense-against-information-war-what-we-learned-learn-discern>

Suwajanakorn, S. et al. (July 2017) Synthesizing Obama: Learning Lip Sync from Audio, *ACM Transactions on Graphics*, 36 (4). Article 95.
http://grail.cs.washington.edu/projects/AudioToObama/siggraph17_obama.pdf

Sydell, L. (2017) (Nov 23, 2016) We Tracked Down A Fake-News Creator In The Suburbs. Here's What We Learned, *NPR*,
<http://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs>

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755-769.

Tambini, D. (2017) How advertising fuels fake news. *LSE Media Policy Project Blog*, available at: <http://blogs.lse.ac.uk/mediapolicyproject/2017/02/24/how-advertising-fuels-fake-news/>

Tandoc, Jr., E. C, Lim, Z. W., and Ling, R. (Aug. 2017) Defining 'Fake News': A Typology of Scholarly Definitions, *Digital Journalism*, 5 (7): 1-17

Thies, J. at al. (2016) Face2Face: Real-time Face Capture and Reenactment of RGB Videos, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2387-2395

Thompson, A. (Dec. 8, 2016) Parallel Narratives, Clinton and Trump Supporters Really Don't Listen to One Another, *Vice*. <https://news.vice.com/story/journalists-and-trump-voters-live-in-separate-online-bubbles-mit-analysis-shows>

Thompson, D. (Nov. 3, 2016) The Print Apocalypse and How to Survive It, *The Atlantic*.
<https://www.theatlantic.com/business/archive/2016/11/the-print-apocalypse-and-how-to-survive-it/506429/>

Uberti, D. (December 15, 2016) The Real History of Fake News, *Columbia Journalism Review*,
https://www.cjr.org/special_report/fake_news_history.php

Venkatraman, A., Mukhija, D., Kumar, N., & Nagpal, S. J. S. (2016). Zika virus mis-information on the internet. *Travel medicine and Infectious Disease*, 14(4), 421-422.

Vaidhyathan, S. (Sept 8, 2017) Facebook Wins, Democracy Loses. *New York Times*, <https://www.nytimes.com/2017/09/08/opinion/facebook-wins-democracy-loses.html>

Van Damme, I. & K. Smets, 2014, p. 310. The power of emotion versus the power of suggestion: Memory for emotional events in the mis-information paradigm. *Emotion*. 14 (2): 310-320

Weedon, J., W. Nuland & A. Stamos (April 27, 2017) Information Operations and Facebook, *Facebook Newsroom*. <https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf>

Wardle, C. (February 16, 2017) Fake News. It's Complicated. *First Draft News*. Available at: <https://firstdraftnews.com/fake-news-complicated/>

Wardle, C. (2016) Timeline: Key Moments in the Fake News Debate, *First Draft News*, (Sept 30 – Dec 1, 2016). Available at: <https://medium.com/1st-draft/key-moments-in-the-fake-news-debate-bce5fb6547d4>

Woolley, S and P. Howard (2017) 'Social Media, Revolution and the Political Bot', in *Routledge Handbook on Media, Conflict and Security*, (edited by Piers Robinson, Philip Seib, Romy Frohlich), London: Routledge.

Woolley, S & P. Howard, (2017) Computational Propaganda Worldwide: An Executive Summary, <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Casestudies-ExecutiveSummary.pdf>

Yale University's Information Society Project's 'Fighting Fake News' Workshop https://law.yale.edu/system/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf

Yi, S.B. (June 19, 2017) New legislation to combat fake news likely to be introduced next year: Shanmugam, *Straits Times*, <http://www.straitstimes.com/singapore/new-legislation-to-combat-fake-news-next-year-shanmugam>

Young, D. G., K.H. Jamieson, S. Poulsen and A. Goldring (2017) Fact-Checking Effectiveness as a Function of Format and Tone: Evaluating FactCheck.org and FlackCheck.org, *Journalism And Mass Communication Quarterly*, 1-27

Yurieff, K. (April 5, 2017) eBay founder commits \$100 million to combat 'fake news', *CNN*, <http://money.cnn.com/2017/04/05/technology/pierre-omidyar-donation-fake-news/index.html>

Zaller, J. (1992) *The Nature and Origins of Mass Opinion*, Cambridge University Press.

Zimdars, M. (Nov 18, 2016) My 'fake news list' went viral. But made-up stories are only part of the problem, *The Washington Post*,

<https://www.washingtonpost.com/posteverything/wp/2016/11/18/my-fake-news-list-went-viral-but-made-up-stories-are-only-part-of-the-problem/>

Zuckerman, E. (January 30, 2017) Stop Saying 'Fake News' It's Not Helping, *Ethanzuckerman.com*, Available at: <http://www.ethanzuckerman.com/blog/2017/01/30/stop-saying-fake-news-its-not-helping/>

Zuckerman, E. (August 2017) Mistrust, Efficacy and the New Civics, *A Whitepaper for the Knight Foundation*, <https://dspace.mit.edu/handle/1721.1/110987#files-area>

This report provides a new framework for policy-makers, legislators, researchers, technologists and practitioners working on the theoretical and practical challenges related to mis-, dis- and mal-information – the three elements of information disorder. While the historical impact of rumours and fabricated content have been well documented, the complexity and scale of information pollution in our digitally-connected, increasingly polarised world presents an unprecedented challenge. There is an immediate need to work collaboratively on workable solutions and this report provides a framework for the different stakeholders involved in research, policy discussions, and technical innovations connected to this phenomenon of information disorder.

This report has been commissioned by the Council of Europe and produced in co-operation with First Draft and the Shorenstein Center on Media, Politics and Public Policy.

**FIRST
DRAFT**

 **HARVARD Kennedy School**
SHORENSTEIN CENTER
on Media, Politics and Public Policy

www.coe.int/freedomofexpression

PREMS 112208

ENG

www.coe.int

The Council of Europe is the continent's leading human rights organisation. It comprises 47 member states, 28 of which are members of the European Union. All Council of Europe member states have signed up to the European Convention on Human Rights, a treaty designed to protect human rights, democracy and the rule of law. The European Court of Human Rights oversees the implementation of the Convention in the member states.