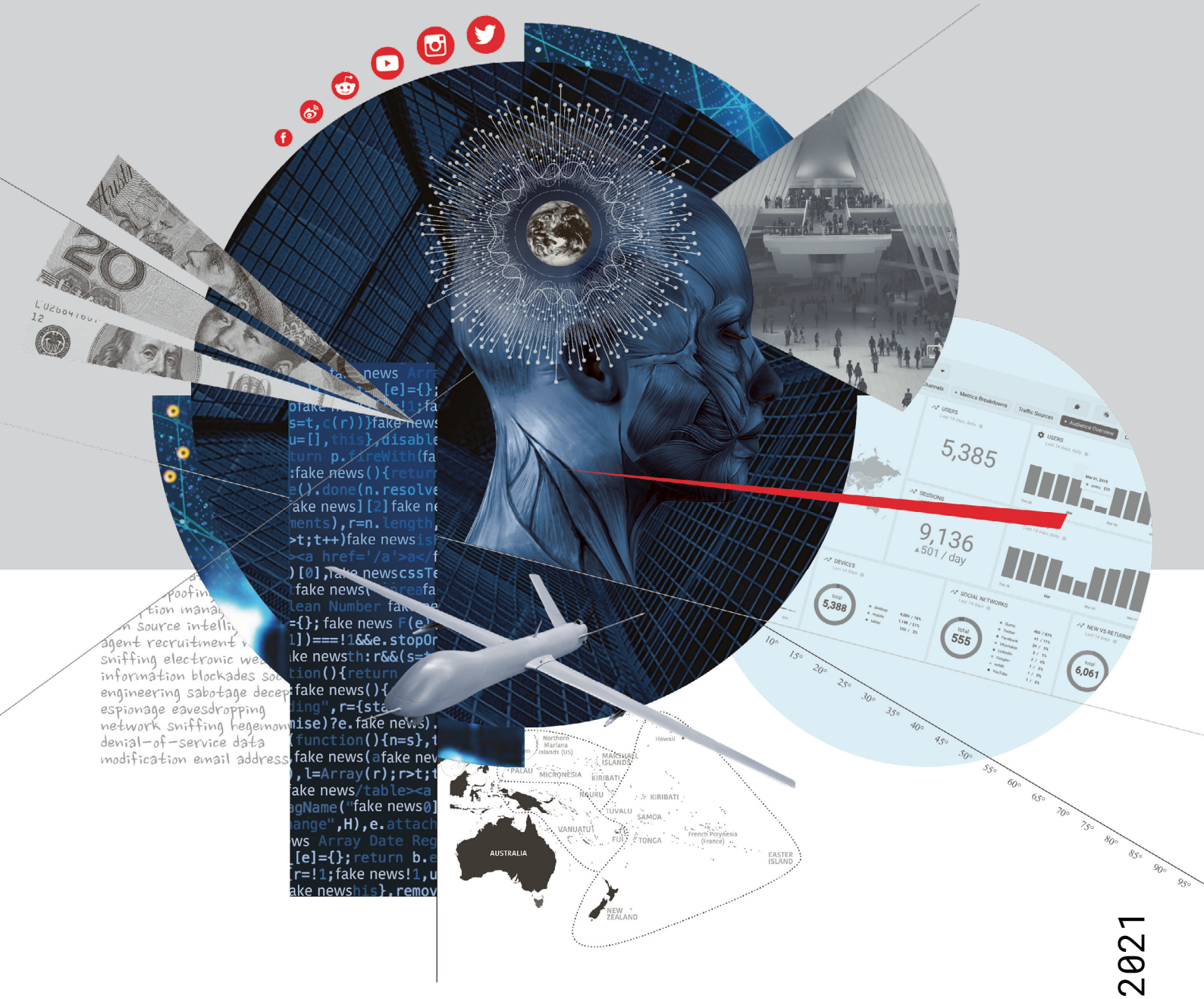
Understanding Mass Influence

Three case studies of contemporary mass influence activities



AUGUST 2021

Produced for the Department of Defence by: The University of Adelaide, The University of Melbourne, University of New South Wales, Edith Cowan University and Macquarie University

This project report is jointly submitted by the parties set out below as part of a Standard Collaborative Project pursuant to Defence Science Partnering Multi-Party Collaborative Project Agreement (Agreement No. MyIP10379) dated 11 February 2021. The ownership and use of Intellectual Property subsisting in the Report is subject to the terms of that Agreement.

Edith Cowan University
Macquarie University
University of Adelaide
University of Melbourne
University of New South Wales

Any comments or queries regarding the project report should be directed to:
UNSW Defence Research Institute
Northcott Drive, Campbell, ACT 2602, Australia
info@dri.unsw.edu.au
www.dri.unsw.edu.au

Text design and typesetting by Raye Antonelli, The Friday Collective Cover design by Raye Antonelli, The Friday Collective
Cover image by Naomi Cain, The University of Adelaide

Foreword

The 2020 Defence Strategic Update, and the strategic policy review preceding its release, highlighted the deteriorating strategic environment since the 2016 Defence White Paper was released. The Defence Strategic Update identified that a new strategy and capability investment plan for Defence was required to safeguard Australia.

With this in mind, this Report provides a multidisciplinary approach on the mass influence of technology in a dynamic strategic environment. Traditional warfighting domains have evolved to reflect the changing environment. However, military modernisation, technological disruption, and the risk of state-on-state conflict are complicating Australia's strategic circumstances. Expanding cyber capabilities – and the willingness of some countries and non-state actors to use them – are further complicating Australia's strategic environment. As one of Australia's instruments of national power, Defence's response to active interference, disinformation campaigns and economic coercion are a constant challenge.

The power of multidisciplinary research collaborations in solving complex, multi-factor challenges is evident when reviewing the perspectives provided by the three case studies in this Report. The Universities of Adelaide, Melbourne and New South Wales with Edith Cowan University and Macquarie University assembled academic teams and a panel with diverse expertise to undertake these case studies. Discipline perspectives included cognitive science, psychology, information systems, cyber security, AI, law, political science, linguistics, data science, business, human factors, computer science, digital marketing and strategic studies. The synthesis of academic perspectives identified the potential impacts, where cyber security awareness is not adhered to, allowing mass influence to effectively persuade, manipulate or target individuals or organisations.

The Australian Department of Defence commissioned this report in 2020. Despite Commonwealth funding, the views expressed in this Report are the views of the authors and do not reflect the views of the Australian Government or the Department of Defence.

For the foreseeable future, how we respond to the changing environment will be an overarching concern of Australian Governments requiring consideration and debate of all perspectives to ensure and provide the best safeguards to defend Australia.

Major General Susan Coyle, CSC, DSM
Head Information Warfare
Joint Capabilities Group

Contents

Executive Summary

Background

The digital age has changed our lives – and also the character of conflict and warfare. Our lives are increasingly connected by and dependent on the technologies the digital age has provided, and our day-to-day activities are increasingly reliant on digital information.[1]

With technology comes disruption and with connectivity comes susceptibility. In a dynamic strategic environment in which the willingness and capabilities of some countries, non-state actors and commercial entities to use cyber capabilities to influence populations psychologically, politically and economically is increasing, technological disruption and population susceptibility threaten the rules-based global order and citizens' safety and security.[2]

The Australian Government's cognisance of what are often now referred to as 'grey zone threats' is evident in the 2020 Defence Strategic Update and Force Structure Plan. The traditional warfighting domains of air, land and sea have evolved to include space and cyberspace,[3] and the Government has committed $15B over the next decade to strengthen Defence's Information and Cyber domain capabilities.[4]

Detecting and countering grey zone threats is complex and difficult, however. Adversaries seek to avoid military conflict, making it problematic for Australia to apply the Department of Defence's substantial intelligence, cyber, electronic warfare, information operations capabilities under mandate.

Important lessons can be drawn from studying contemporary digital technologies and influence campaigns to inform the enhancement of Defence's information warfare capabilities. This report, commissioned by the Australian Department of Defence, presents analyses and findings of three case studies examining mass influence through the lenses of a state-sponsored actor, a non-state actor, and a mainstream platform enabler:

1. State-sponsored actor: Russian Internet Research Agency
2. Non-state actor: Cambridge Analytica
3. Influence platform: Facebook.

Approach and Method

The University of Adelaide, the University of Melbourne and the University of New South Wales with Edith Cowan University and Macquarie University to undertake three case studies in collaboration with staff from the Defence Science and Technology Group and the Joint Influence Activities directorate of the Information Warfare Division. Four interrelated domains and associated research questions were devised to guide and structure each of the case studies:

1. Governance and Ethics: What was the organisation's business model for operations, including their operating concept, financing arrangements, governance, legal and ethical framework?
2. Persuasive Technology and Techniques: How was the organisation able to use technology and techniques to persuade target audiences?
3. Systems and Technology: What were the organisation's foundational systems, technology and workforce skills required for operation?
4. Campaign Awareness and Sensemaking: How was the organisation able to achieve and maintain awareness of the impact of their influence activities?

Teams from the Universities of Adelaide, Melbourne and New South Wales led the conduct of each case study, supported by scientists from Defence Science and Technology Group, representatives from Information Warfare Division and a panel of approximately 50 academic experts representing highly diverse discipline specialities including:

- Computer science
- Psychology
- Law
- Political science
- Cognitive science
- Information systems
- Cyber security
- Artificial Intelligence
- Linguistics
- Data science
- Business
- Human factors
- Digital marketing
- Strategic studies.

A systematic literature review was conducted by each project team. Discipline perspectives on each research question were sought from academic and Defence experts with feedback incorporated into research and analysis. Iterations of the draft report were shared with Defence and academic experts for consideration, review, and comment throughout the project. This report summarises the key findings and recommendations of the project. Detailed analyses and findings from each case study are presented in separate case study reports available online at: https://dri.unsw.edu.au/groundbreaking_post/understanding-mass-influence/.

1. https://www.defence.gov.au/jcg/iwd.asp, downloaded 1 June 2021
2. 2020 Defence Strategic Update, pp. 3-5.
3. https://www.defence.gov.au/jcg/iwd.asp, downloaded 1 June 2021
4. 2020 Force Structure Plan, p. 27

Executive Summary

Key Findings

Across all three case studies, low levels of cyber security awareness, high levels of user credulity and strong incentives for organisations to seek to persuade, manipulate or coerce target audiences were found to have contributed substantively to detrimental outcomes – intended or otherwise – for individuals and organisations.

State-sponsored actor: Russian Internet Research Agency
The Internet Research Agency or IRA's operations illustrate the effects possible with a large, well resourced, trained, and coordinated workforce.

1. The IRA's operation demonstrated the outcomes possible when an organisation is motivated, uninhibited by laws or societal norms, well-resourced and well-coordinated.
2. The IRA demonstrated what can be achieved in practice through the adoption of an integrated suite of persuasive technologies.
3. The IRA demonstrated the benefits of incorporating psychological principles and marketing techniques in audience engagement.
4. The IRA was a 24/7 operation, enabling real-time time zone specific content creation and engagement.
5. The IRA operated within the wider Russian eco-system of disinformation and propaganda, leveraging an extensive base of expertise and experience as well as the full resources of the Russian intelligence community. The IRA workforce involved approximately 400-600 staff at any one time and 800-1000 staff over the life of the operation.
6. The IRA employed social media platforms such as Facebook that enabled precise micro-targeting of audiences. IRA staff were exceptionally well-versed in internet culture enabling deep infiltration of many diverse online communities. Furthermore, these platforms facilitated the use of many tools to enhance IRA influence operations. For example, IRA operators deployed Bots and Botnets to augment human action, narrative laundering, develop fake personas and cultivate large numbers of followers.
7. A key measure of success for IRA operators was translation of online behaviour and attitudes to offline activity, that is, actions in the physical world. Provoking offline violence between opposed groups online was viewed as a distinct success.

Non-state actor: Cambridge Analytica
Cambridge Analytica, a subsidiary of the Strategic Communication Laboratories (SCL) Group, was a private British behavioural research and strategic communication company that engaged in global information and influence operations.

1. Cambridge Analytica's business model relied on the company's ability to map and exploit the inadequacies in the regulatory environment relevant to its operations.
2. Cambridge Analytica used large cohorts of online as well as offline data from multiple sources to profile millions of individuals and groups and target them with tailored messages.
3. Cambridge Analytica took large amounts of qualitative and quantitative data and used it to develop psychological profiles that informed the design of targeted content for the purpose of shifting public opinion at scale.
   — The underpinning influence theories and models used by Cambridge Analytica for profiling and manipulating individuals and groups were simplistic, weakening its efficacy.
   — Cambridge Analytica's influence operations relied on illegal data harvesting and use.
4. Cambridge Analytica lacked mechanisms to foster and maintain its legitimacy. This made the business unsustainable in a liberal democratic operating environment.
5. Cambridge Analytica did not systematically monitor the impact of its influence operations and likely did not produce the large-scale public opinion effects attributed to them by the company.

Influence Platform: Facebook
Facebook as a leading contemporary social media platform provides insights into the key role digital technology platforms can play in mass influence campaigns.

1. Facebook and digital technology platforms like it enable influence operations. As a globally pervasive platform for social media, Facebook enables influence operations of many types with various motivations including commercial, political, public interest, and malign.
2. Facebook relies on the authenticity of users' identities to build value for their advertising clients. Similarly, Cambridge Analytica and the Russian Internet Research Agency rely on authentic identity to enable micro-targeting for malign purposes.
3. Facebook is insensitive to user intent and activity. As such Facebook cannot be relied on to respond to emerging threats or crises, except when exceptionally dire. The US insurrection provoked a response but represents a particularly high threshold for action.

4. Facebook is likely to remain an efficient platform for propagating disinformation for the foreseeable future.

Recommendations

Defence's Force Structure Plan 2020 has identified a need to "modernise ADF influence activities with an advanced internet operations capability to support Defence's capacity to shape Australia's operating environment".[5] The key recommendations of this study directly address this need and fall into four interrelated categories: governance, capabilities, workforce, and impact and effectiveness. For case-specific recommendations and the analysis from which they were derived, please refer to the case studies themselves.

5. Force Structure Plan 2020, paragraph 3.10, p.29

Governance
The application of Australian military capabilities to the protection of Commonwealth interests, States and self-governing Territories occurs under a complex set of international and Australian legislation, regulations, treaties, and other subsidiary legislation as well as social and ethical values and norms. Even greater complexity arises when considering the application of non-kinetic military capabilities in cyber space in operations short of war – or the grey zone. This complexity may be reduced by establishing principles, codes of conduct and rules of engagement to guide the ADF's information operations that:

1. Align with and protect democratic principles and Australian values
2. Accord with Australian Government policies and international treaties to which Australia is a signatory
3. Consider fundamental differences between defensive and offensive capabilities and operations
4. Allow the necessary flexibility to conduct effective operations and development of new capabilities that respond to rapidly evolving threats in the information environment
5. Establish and provide legitimacy for ADF information operations in war and operations other than war.

Capability
While driven by differing motivations and exhibiting different levels of efficacy, the Russian Internet Research Agency, Cambridge Analytica and Facebook all operate (or operated) with rapidly evolving technological constructs and each organisation embedded technological agility in their ways of working. Key capabilities that enabled this agility and, in turn, pursuit of organisational objectives, include:

1. Effective individual and collective training regimes, including ongoing performance and development review.
2. Strong and evolving understanding of end users, including changing preferences and behaviours over time.
3. Diverse data sources providing contextual knowledge, deep user knowledge and insights, and situational awareness.
4. Broad adoption across organisational functions of advanced technologies including artificial intelligence, machine learning and data science more generally, as well as integration of socio-psychological models of audience targetting.
5. Nuanced cultural competence relevant to target audiences including language, linguistic microcosms, jargon, social structures, values, beliefs including religions, rituals, and symbols.
6. Dedicated horizon scanning capabilities to identify and assess emerging threats, technologies, and techniques with abilities to rapidly address threats and adopt new technologies and techniques.
7. Models, tools, and analytical capabilities supporting measurement over time of the impact and effectiveness of influence campaigns, our own and those of competing actors, at population and sub-population levels.

Workforce
Each case studies demonstrates the importance of a workforce with diverse knowledge and expertise. One way to consider to the composition of an influence operations workforce is to consider how different skills and disciplines contribute to capability. For example:

1. Targeting: Political Science and International Security provide clarity around purpose, targets and goals for an influence operation, that is, what to do and why.
2. Planning and Situational Awareness: Psychology, Computer Science, Engineering, Science and Technology Studies and Ethnographic Research can provide insight into human factors in complex systems involving technology and humans to inform campaign planning and effectiveness assessment.
3. Capability: Social and Cognitive Psychology, Data Science, Creative Content Production and Linguistics provide insight and means into how to achieve influence, including mass influence.
4. Security: Cyber Security, Law and Ethics can facilitate secure, safe, and publicly palatable operations.
5. Organisation: Computing and Information Systems and Organisational Behaviour can provide insight into how to design organisational structures and processes that maximize agility (speed, flexibility, innovation) for operations in the information environment.

Impact and Effectiveness
The three case studies demonstrate the importance of maintaining situational awareness of the reach and impact of influence campaigns conducted by malign operators. The same applies to any capability developed implement counter-influence activities and campaigns.

1. In social influence, an indeterminate number of variables can contribute to understanding and predicting behaviour. Contemporary approaches to the science of causation, including data analysis using multi-level hierarchical models, together with emerging approaches to social network analytics, provide new tools for interrogating campaign impact and effectiveness.
2. A combination of qualitative and quantitative metrics monitored over time will be required to build:
   a. situational awareness
   b. understanding of the impact and efficacy of influence campaigns and counter-campaigns
   c. understanding connections between influence operations and behavioural changes.
3. A broad range of platform, user, usage, consumer, third party and campaign data will be required to develop, validate, and compare understanding of the reach, impact and effectiveness of influence campaigns and counter-campaigns over time.

Executive Summary

Directions for Future Research

This work has demonstrated the value and indeed criticality of combining insights and knowledge from diverse disciplines to gain an understanding of how mass influence effects in our day have been realised. Similar multi-disciplinary teams will be vital to advancing our understanding of how to respond to the here-and-now threat of foreign influence campaigns active in our areas of interest.

Existing Defence Research Programs
Defence is developing a rich array of research programs that address many of the needs associated with this area of capability need. These include:

a. Defence's Next Generation Technology Fund (NGTF) funded Cyber[6] program managed by DSTG with its focus on advances at the intersection of cyber and artificial intelligence, sometimes referred to as autonomous cyber operations
b. DSTG's Information Warfare STaRShot[7] with its focus on control of an adversary's human, information, and physical environments through an integrated information warfare capability
c. DSTG's Modelling in the Grey Zone[8] program with its focus on modelling grey zone activities; and
d. Defence funded research in International Security[9] and Law[10] as well as this present set of case studies.

International Collaboration
In keeping with DSTG's More Together Strategy[11], international collaboration with like-minded international partners facing similar challenges is vital to achieving the scale of effort necessary to respond effectively to this threat.[12] The United States of America has two complementary initiatives of particular relevance:

a. US Defense Advanced Research Projects Agency (DARPA) Information Innovation Office's program; and
b. US Department of Defense funded University Affiliated Research Centre (UARC), at the University of Maryland, College Park, the Applied Research Laboratory for Intelligence and Security[13] (ARLIS) with its focus on Information and Influence and the Human Domain.

Key Themes
Complementing existing programs, the following additional key themes are identified:

a. Sensemaking, or Situational Awareness and Modelling in the Information and Cyber Domain
b. Human Autonomy Teaming, or "AI as partner"
c. Cognitive Security and Disinformation.

Sensemaking, Situational Awareness for the Information and Cyber Domain
Making sense of observations in the information and cyber domain is a fundamental requirement for any operational influence capability. One must be able to observe and make sense of both an adversary and one's own actions in an environment if one is to successfully prosecute a response to a malicious influence campaign. Sensemaking offers a unifying framework for influence research because it is so essential to both detection and responsive capabilities. A Defence perspective would interpret sensemaking as situational awareness, that is, the perception and comprehension of events, and projection of likely futures. As such it has strong links to command and control in information warfare.

Human Autonomy Teaming (AI as partner)
"Artificial intelligence (AI) technologies have made little progress in understanding the most important component of the environments in which they operate: humans. This lack of understanding stymies efforts to create safe, efficient, and productive human-machine teams."[14] Critical to the development of an Australian influence operations capability will be the ability of human operators and AI to partner in making sense of situations and in planning appropriate responses.

Cognitive Security and Disinformation
"Disinformation is one of the most critical issues of our time, concerned with online and offline influence at scales ranging from individuals to large populations. Operations in the Information Environment are conducted within the context of Cognitive Security. The movement toward symbiotic human-machine interfaces creates an urgent demand for research to inform operations in the broadest sense."[15] Research in cognitive security is principally aimed at reducing vulnerability to misinformation and manipulation in online systems.

Conclusion

This research project identified important considerations for Defence as it works, within a whole-of-Government context, to strengthen Australian digital sovereignty in response to growing state and non-state threats to Australian governments, businesses and communities in cyber space.

Detailed Case Study Reports

The findings and recommendations summarised above are discussed in greater detail in the three following case study synopses, and in detailed Case Studies available online at: https://dri.unsw.edu.au/groundbreaking_post/understanding-mass-influence/.

6. https://www.dst.defence.gov.au/NextGenTechFund/cyber
7. https://www.dst.defence.gov.au/strategy/star-shots/information-warfare
8. https://www.dst.defence.gov.au/partner-with-us/university/modelling-complex-warfighting-strategic-research-investment/modelling
9. Defence funded research: "Countering foreign interference and influence", (ID9586) Tim Legrand, University of Adelaide.
10. Defence funded research: "In relation to Influence Operations and Commonwealth Law", (MyIP:10656) Dale Stephens, University of Adelaide.
11. https://www.dst.defence.gov.au/strategy
12. https://www.darpa.mil/work-with-us/i2o-thrust-areas
13. https://www.arlis.umd.edu
14. https://www.darpa.mil/program/artificial-social-intelligence-for-successful-teams
15. https://www.arlis.umd.edu/cogsec

Case Study 1

Internet Research Agency

Internet Research Agency

Emily Ebbott with contributions from Dr Morgan Saletta and Richard Stearne

Introduction

The term Internet Research Agency (IRA) refers to an organisation that operated from 2013, when it was officially created as a business, until 2018, when operations at its headquarters at 55 Savushkina St, St Petersburg, Russia, ended and the company was dissolved.

This report provides an overview of the four themes outlined below. We begin by presenting the key findings and recommendations then detail the strengths and weaknesses of each of the themes. This report provides additional detail in an effort to contextualize the themes in real world examples. As IRA was a company operating under the auspices of the Russian Government, the findings and recommendations of the business model, operating tactics and persuasive technology raise considerings for a Department of Defence approach to Australia's information warfare capability.

Four themes and questions were identified:

- Governance and Ethics: What was the IRA's business model for operations, including its operating concept, financing arrangements, governance, and legal and ethical framework?
- Persuasive Technology and Techniques: How did the IRA use technology and techniques to persuade its target audiences?
- Systems and Technology: What were the foundational systems, technology and workforce skills required for IRA's operation?
- Campaign Awareness and Sensemaking: How was the IRA able to achieve and maintain awareness of the impact of their influence activities?

Key findings

- The Russian Internet Research Agency (IRA) operated with direct approval and endorsement from Russian President Vladimir Putin.
- The IRA was funded by Russian Businessman, Yevgeny Prigozhin and operated ostensibly as a digital marketing firm, complete with corporate hierarchy and all normal expected business units.
- The IRA workforce involved approximately 400-600 staff at any one time to over 800-1000 staff. The IRA was a 24/7 operation, enabling real-time time zone specific content creation and engagement among other benefits. Two key benefits were realised by operating as a private entity: (a) plausible deniability for the Russian Government in relation to their support and involvement, and (b) creative license for the business itself.
- The IRA's overarching objectives were to sow discord and division in nations not aligned with Russian geopolitics and undermine confidence in institutions that underpin democratic principles, such as the US electoral system. Its primary mode of operation was to amplify pre-existing polarisations within society.
- The IRA operated within the wider Russian eco-system of disinformation and propaganda, leveraging an extensive base of expertise and experience as well as the full resources of Russian intelligence community.
- The IRA differed from other Russian influence operations in its use of social media platforms to reach and engage target audiences. For example, Facebook was used extensively by the IRA and provided tools that are ideally suited to the conduct of influence operations. These tools enabled precision micro-targeting of audiences, the deployment of Bots and Botnets nets to augment human action, narrative laundering, and many other techniques. Furthermore, IRA staff were exceptionally well-versed in Internet culture and online subcultures, facilitating deep infiltration of many, diverse online communities, creating websites, developing fake personas, and cultivating large numbers of followers.
- A key measure of success for IRA operators was translation of online behaviour and attitudes to offline activity, that is, actions in the physical world. Provoking offline violence between opposed groups online was viewed as a distinct success.

Recommendations

- Establish principles, codes of conduct and rules of engagement that align with democratic principles, Australian values and policies, and international treaties Australia is a signatory.
- Recruit and develop collective training regimes to establish an agile and innovative workforce that can develop experts at scanning the horizon for advancing technologies, to ensure ongoing awareness of evolving platforms and countering techniques.
- Augment native tool sets with off-the-shelf, third party tools for monitoring the social media landscape and identifying key vulnerable groups and individuals relevant to Australia's national interest.
- Apply the multidisciplinary team's expertise to survey and develop methods for identifying, monitoring and measuring the complex relationship between online behaviour, changes in attitudes and behaviour, and correlations to offline behaviour.
- Develop a coordinated ecosystem comprising Defence, Intelligence and non-government expertise that would contribute to the full spectrum of operations.
- Engage with regional partners in the Indo-Pacific to increase resilience to malign or hostile information operations and to boost, where possible and appropriate, local capabilities in the information environment.

Background

The first recorded mention of the IRA has been traced to an undercover Russian journalist who in 2013 described a "troll factory" with a collection of "internet operators" posting political propaganda and comments.[1] The "troll factory" was the IRA. The organisation has been the subject of numerous detailed investigations by journalists as well as the US Intelligence Community. It has been a "front" organisation for the Russian Government, operating in a deniable fashion and as a proxy for various arms of the Government. There is evidence its ecosystem included the GRU, or GU, the Main Intelligence Directorate, which acted alongside the IRA developing, for example, pro-Kremlin narratives, the IRA used to "narrative launder" messaging campaigns through its various social media channels.[2] The FSB, the more traditional intelligence agency, reportedly provided the ecosystem support for their cyber-hacking functions.[3] The Foreign Intelligence Service (SVR), which reportedly applies a targeted approach to cyber-espionage campaigns, retains the results of these operations to support the Kremlin's strategy.

This contrasts with the GRU's function of hacking emails and releasing information for political impact.[4]

The ecosystem's contributing workforce also consisted of official, state-run media outlets, through to proxy media outlets which enabled the Kremlin's plausible deniability and contributed, as part of the broader government apparatus, to its ability to "narrative launder". There was an intense sense of competition between these agencies, and they would often direct their respective campaigns on to the same target, despite being a part of the same ecosystem.[5]

The IRA evolved into a covert private military company carrying out influence operations in the information environment, using the methods, business model and cover of a digital marketing firm. It was, in its own words, perpetrating "information warfare" in the service of the Russian Government's domestic and geopolitical goals.[6] It was established under the direct approval of Putin and funded by Prigozhin,[7] who was allegedly directly involved in its management, meeting on a regular basis with the senior leadership.[8] Due to the increasing toxic nature of organisations associated with the Kremlin, a sense of deniability was crucial to the legitimacy of the IRA and its associated operations.[9] Relying on covert operations to achieve its geopolitical aims and expand its audience for certain messages without exposing its direct connection to the Kremlin was the initial aim.

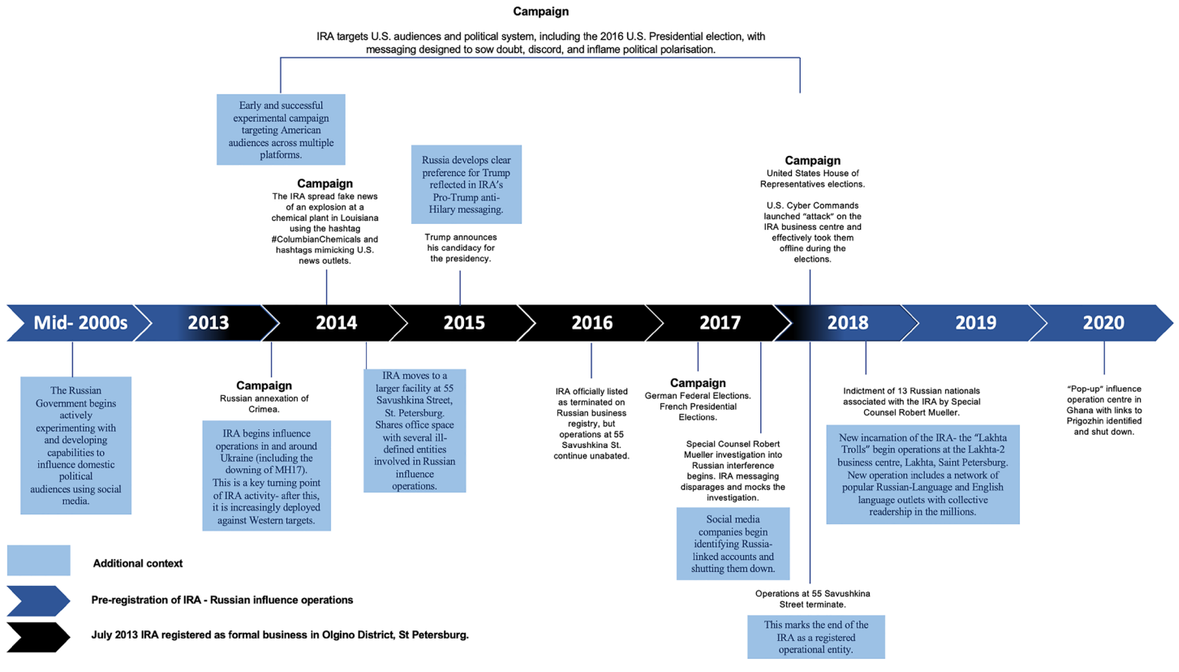


Figure 1: Internet Research Agency Timeline (Prepared by Jemma Smith, Emily Ebbott, Richard Stearne amd Dr Morgan Saletta

Internet Research Agency

Governance and Ethics

Key Strengths
- The IRA operated in alignment with Putin and the Russian Government's domestic and strategic geopolitical goals: remaining in power, returning Russia to a great power status, weakening the West by sowing doubt, division, and discord. If its influence operations were working toward these broad goals, it could, for example, operate with ideological fluidity and thus target diverse audiences across the political spectrum.
- The IRA's legal incorporation as a private company and hidden/disguised funding via Prigozhin's businesses gave President Putin and the Russian Government plausible deniability regarding the IRA's influence operations.
- The IRA's business model, based on a digital marketing firm, was tailor-made to take advantage of the new information environment created by social media platforms and lax regulation.
- The IRA's business model took maximum advantage of the data harvested by social media companies to segment populations and micro target audiences based on demographics, behaviours, and attitudes.
- The IRA's business model is easy to clone, and new pop-up influence shops have appeared in Russia and elsewhere.
- Large numbers (estimates range from 400-1000) of "trolls" with basic linguistic, cultural, and technical skills were able to quickly create and spread large amounts of content/messages across multiple platforms.

Key Weaknesses
- The IRA's business model created a paper trail, a physical and digital footprint that eventually reduced its plausible deniability.
- Criminal activity (in the United States) has resulted in indictments against Prigozhin and other IRA managers, as well as sanctions by the US Department of the Treasury against individuals including Prigozhin and associated business entities.

Governance

The IRA was a legally registered business in Russia. The IRA served as a proxy, private military company carrying out influence operations in the information environment with approval from Putin, with additional direction and substantial funding from Prigozhin funnelled through two of his existing businesses. By September 2016 records indicate the monthly funding equalled some US $1.25 million.[10]

Prigozhin is an oligarch known to the Russian Government with close ties to Putin and the Intelligence Community. He is involved in media entities that promote pro-Kremlin propaganda and engage in "narrative laundering". He also runs the Wagner Group, a private military company, and acts as a proxy for the Russian Government. To maintain the IRA's cover and decrease exposure to the Kremlin the IRA's funding source was heavily concealed. Funds were routed through 14 bank accounts of entities associated with Prigozhin's Concord Catering, and Concord Management and Consulting businesses.[14]

In contrast to the ecosystem's counterparts, the IRA did not operate like a traditional intelligence or government organisation. The IRA adopted the business model and structure of a "digital marketing firm".[11] The management included IT entrepreneurship, advertising and public relations professionals, and additional marketing firm skills. It was organised into departments and teams, including:[12]
- Content Development "bloggers"
  Content developers worked individually and in teams, depending on the desired outcome, e.g., when driving targeted comments and discussion on websites.
- Geographical regions, for example the 'American Department' also commonly known as the Translator Project
- Data analysis
- Search Engine Optimisation
- Design and Graphics
- Information Technology
- Finance.

It appears the bulk of the workforce comprised of entry-level "trolls" producing social media content. The remaining roles involved managing more sophisticated sock puppet accounts (false online accounts) and required advanced language and cultural skills.

Management conducted social media analysis and briefed lower-level employees and other content developers daily on their tasks. These daily briefings were used to identify targets and provide broad instruction on how to zero in on various audience groups. Creative licence was condoned if key benchmarks were met. Forms of creative licence might include the number of words used to create a post, or the application of authorised graphics.[13]

The business model, a combination of a corporate structure and "active measures" strategies adopted from the Soviet era, was developed to take advantage of the information environment, and to exploit social media platforms with the aim of segmenting and targeting audiences using content based on a knowledge of their behaviour and attitudes, not for commercial purposes but malign influence. The approach allowed for agility and experimentation.

Ethical and Legal Framework

At a strategic level, the US Intelligence Community assessed that the IRA operated with explicit approval of President Putin. Additionally, that Putin expressly ordered specific influence campaigns involving the wider information operations ecosystem that targeted the US political system and 2016 US elections. At an operational and tactical level, Russian influence campaigns tend to be decentralised in terms of their governance and direction. Individuals were "guided by their sense of the Kremlin's desires rather than any master plan".[15]

As the IRA was indicted by the US Government, it could be classified as a criminal enterprise, the company and 13 employees were caught conducting criminal activity. They knowingly and intentionally conspired to interfere in the sovereign political and legal process of the US elections through criminal activity and identity theft.[16] Ethically, the IRA, in contrast with Soviet era active measures, was not serving a larger ideological purpose. Operating with arguably ethical and ideological fluidity allowed the organisation to target diverse audiences across the political spectrum. The primary goal was to advance Russia's domestic and geopolitical aims with very little reliance on, or regard for facts that aligned with other narratives from their identified target groups, for example.[17]

Russia's desire to remain credible in the eyes of the international community was a primary reason the IRA operated in a covert, deniable fashion. This allowed Russia to perpetuate an image of itself as a country committed to supporting international norms and agreements. However, investigations into the IRA's activities uncovered violations against international treaties.

The following international treaties, to which Russia is a signatory, have been violated through the covert actions of the IRA because of its messaging, and the way it sought to inflame racial and ethnic tensions:
- The International Covenant on Civil and Political Rights (ICCPR)
  Article 20(2) of this treaty states that "advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law".
- The Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights).

Additionally, individual countries have adopted legislation that attempts to address the issues associated with influence operations campaigns. Alongside national legislation, social media platforms are actively developing methods, procedures, and guidelines for the removal of accounts identified as having an alignment with influence activities. These are often referred to as "coordinated inauthentic activity".[18]

As a result of efforts to counter inauthentic activity, the IRA adapted its operations. For example, it switched significant messaging efforts to Instagram when Facebook began investigating and shutting down IRA-related accounts.[19] More broadly, various actors in the Russian Information Operations ecosystem have adapted their tactics, reducing the reliance on sock puppet accounts as these require heavy 'following' cultivation, and focusing more on the use of local freelance journalists and news media outlets to create content, and drive Pro-Kremlin narratives in targeted countries.[20]

Persuasive Technology

Key Strengths
- Ability to target audiences across multiple platforms at high speed and volume.
- Ability to micro target diverse audiences across the political spectrum – from Black Lives Matter followers to 'Gun rights' activists to LGBTQ audiences.
- Ability to grow (and target) large followings using false, sock puppet accounts, websites, etc.
- Ability to cultivate and recruit assets for online and offline activities.
- Ability to analyse social media data using off-the-shelf, third party software and tools native to platforms, to segment and target audiences.
- Ability to amplify selected messages and narratives (wedge issues, conspiracy theories, etc.) using bots and botnets.
- IRA messaging and content effectively targeted emotions and social identities of audiences. This included extensive nationalistic messaging.

Key Weaknesses
- Reliance on using sock puppet and other false/counterfeit sites to grow and target large audiences. When these accounts or sites are identified and taken down (and social media platforms are getting better at this), the audiences are effectively lost.
- Reliance on botnets to amplify messaging/narratives. Tools and methods for identifying and taking down botnets are improving, evidence indicates this will become a new 'arms race'.

As previously mentioned, the IRA's business model is akin a digital marketing firm. It conducts online influence campaigns by leveraging the social media business models of advertising revenue and data harvesting. It utilises off-the-shelf (native and third party) tools to segment and micro target diverse audiences. It conducts other social media analytics to enable and identify key narratives and symbols, which are then leveraged to reach diverse audiences. To continue its covert actions, it used stolen US identities to purchase server space, and disguised activity with Virtual Private Networks (VPN), while also creating sock puppet identities.

The IRA engaged audiences across multiple platforms and channels, leveraging the content creation and curation of hundreds of employees. It developed and leveraged numerous sock puppet online accounts and audiences, to deeply embed its messaging into the target population's social media network. The IRA made extensive use of automated bots and botnets as force multipliers to amplify selected narratives and content. This created a high volume, high-speed "firehose" of content directed at targeted audiences.[21] This ecosystem created an influence multiplier and amplifier effect, which increased the outreach and significance of the messaging.[22]

Internet Research Agency

Key Tactics

- Multi-platform messaging at high-speed and volume
  The IRA was active across a social media ecosystem that enabled cross-platform links. Sending a target audience messages from multiple, seemingly 'independent' sources gave the messages credibility. Hundreds of human operators produced messaging at a high-speed and in high volumes. Amplification of the content was enabled by bots and botnets, which assisted the legitimisation process.

Example
In 2014 the IRA's MH17 "plane crash" campaign spread content at huge speed and volume to favour Russia and implicate Ukraine. This comprised approximately 45,000 tweets in 24 hours, which impacted the legitimate international investigation and its findings.

- Amplification
  The IRA made extensive use of bots to amplify its messaging and material from, for example, websites and influencers, whose material aligned with its goals. This amplification aided the multi-platform messaging.
- Microtargeting
  The IRA took advantage of the large volume of user data harvested from social media platforms, analysed by native and third party, analytic software, and data brokers. This provided the IRA with clear segmented populations and microtargeted audiences to target with its tailored messaging.
- Paid advertising
  The IRA used paid advertising to grow and target audiences based on behaviour and preferences gleaned from individual profiles. This information was collected alongside data that had been harvested by social media platforms and made available.
- "Doppelgänger" websites
  The IRA created an extensive environment of (evil-twin) websites that mimicked the websites of genuine social movements such as Black Lives Matter. It did this to grow, target and influence legitimate audiences. It was also a useful environment in which to cultivate witting and unwitting assets.
- Sock Puppet accounts
  Fake personas of varying levels of sophistication were created to infiltrate existing social media groups, actively engaging, and targeting members the IRA wished to cultivate as assets. These assets would go on to post and organise rallies and demonstrations.

Example
The IRA created the Twitter handle @TENN_GOP (claiming to represent Tennessee Republicans) which accrued some 100,000 followers.[23]

- Memes and audio visual/symbolic messaging
  Extensive original content was generated in the form of memes, YouTube videos, tweets, and recycled/repurposed existing memes, to contribute to, and amplify, existing messaging.
- Manipulation and "narrative laundering"
  Distributed content/messaging using Twitter handles mimicking multiple credible US (or other target) news sources to spread messages and grow and target audiences.
  Used false online personas to recruit assets to hire unwitting Western journalists to write articles for its "fake news" outlets.
  Directed traffic by using links and comments to multiple pro-Russian known 'news' websites, such as PeaceData.net, using human trolls as well as humans.
- Psychology of influence
  IRA's activity indicated it applied some fundamental psychological principles to developing and disseminating its persuasive messages. These included, but were not limited to:
  - First impressions
  - Repetition and quantity
  - Building credibility and information reinforcement
  - In-group and out-group identification
  - Emotional arousal
  - Information overload

Through analysis of various sources, the IRA's activity also indicates it applied and leveraged limited psychological principles related to influence to develop and disseminate its persuasive messages. However, it's possible this leveraging was an indirect result of applying digital marketing techniques. Specifically, IRA messaging was designed to tap into social identity and provoke emotional arousal, as research suggests messages and content are more likely to be believed and shared[24] if these elements are present.

Systems and Technology

Key Strengths
- Ability of human trolls to use their creativity to generate engaging content and messaging, using skills in cultural knowledge, linguistic knowledge, cultural capital, and social media marketing.
- Understanding of and ability to leverage the new opportunities provided by social media platforms in terms of data harvesting and microtargeting.
- Ability to deploy very large numbers of bots (estimates range from 25-60,000) to amplify messages and narratives.
- Ability to use off-the-shelf software and tools to analyse data, segment populations and manage social media messaging.
- Understanding of best practice digital marketing tools and techniques to create content, websites, etc.
- Understanding of the basic psychological principles of persuasion and messaging – focus on emotive content and messages/symbols, etc. that emphasise in-group/out-group differences, etc.
- Creativity, innovation.

Key Weaknesses
- May not have made maximum use of advertising tools available on social media platforms such as Facebook.
- Reliance on false identities and bots made influence campaigns subject to takedowns and counter messaging, bringing malign activities to light.
- Linguistic cues (grammar mistakes made by Russian-speaking, English-second-language trolls).

The IRA relied on several different technologies to plan and carry out its information operations. Strategies of legitimisation and amplification were key to the human and technological interactions. The key technologies and techniques of bots, botnets, social media analytics, social listening and search engine optimisation were contributing to the IRA's workforce skill set. This workforce is similar to what is found in digital marketing firms. The technologies consisted of:

"Trolls"
- Virtual Private Networks (VPNs) to hide its identity and location
- Cryptocurrency allowed the IRA to establish and operate hundreds of US email accounts, establish PayPal accounts, purpose political advertising, and operate false social media accounts for long periods without detection.[25]
- Developed graphic and audio-visual tools and software to produce content for the platforms and various accounts, ranging from visual memes to videos for YouTube.[26]

Management
- Social Media analytic and monitoring platforms used for situational awareness, sensemaking, audience segmentation, and to assist in the production of daily briefings given to trolls to guide their online commentaries, blog posts and other activities.
- Scheduling tools are found in the above and were likely used to assist with mass delivery across platforms and enable a coordinated strategy for bots and botnets to proliferate designated messaging.

Bots and Botnets
Automated bots were key to the IRA's strategy of amplifying its messaging and applying the tactic of repetition and volume to push its content across multiple platforms/channels. One of the main threats to the influence operations ecosystem is the advancement of Artificial Intelligence (AI) in bot usage. There are two common bots and the IRA used them both:
- Social bots mimic the social behaviour of a human social media user. Capabilities include performing social interactions, responding to questions, and generating debate by posting on trending topics.[27] Social bots are typically controlled by a 'bot-master' and form part of a network which requires medium level technical expertise.[28]
- Bots perform low tech amplification tasks such as liking or sharing content, and can be created simply and with freely available, off-the-shelf software.[29]

The IRA used social bots as a key part of its overall strategy, simulating human behaviour on platforms which gave it legitimacy in its interactions with users and helped promote its views.

In contrast, bots with advanced AI capable of generating answers to questions and producing original content are, generally, more complicated, and costly, and the technology not readily available. While some bots have limited AI capability, it appears these were not used in IRA operations.[30]

Example
In the US elections in 2016 estimates suggest bot usage in the influence campaign sat between 36,000 – 50,000. It is not known whether IRA developed its own bots or purchased and repurposed most of them.[31] However, Mikhail Burchik, one of 13 IRA employees indicted by the US Department of Justice, was a technology entrepreneur who had previously developed his own amplification software and likely had the skillset to create bots and botnets.[32]

Campaign Awareness and Sensemaking

Key Strengths
- The IRA had excellent sensemaking and situational awareness techniques that allowed the organisation to identify and target audiences, target and amplify cultural and political divisions, etc.
- The IRA tracked public and audience interests and opinions with off-the-shelf software and tools, and well-established digital marketing techniques. These methods and tools combined qualitative and quantitative analysis.
- The IRA monitored its own activities using these tools and techniques, and digital marketing metrics, to monitor and adapt their own influence campaigns.

Key Weaknesses
- Digital marketing metrics and qualitative/quantitative analytics may not reliably predict or reflect the relationship between online changes in attitudes or beliefs etc., and offline behaviour.

The IRA leveraged the availability of online data from social media platforms, applying tools for targeting, sensemaking and situational awareness, such as Facebook Advertising and Google Adwords. Additionally, it used third party social media analytic software and tools, include Twidium and Novapress. These tools were crucial to implementing its strategy of segmenting the population into discrete audiences and microtargeting these audiences using data on demographics, behaviours, and attitudes. The IRA also used these tools to

Internet Research Agency

develop a reporting mechanism, including an ability to track its activities for management, providing more information to develop ever more sophisticated online models via this feedback.

Standard digital marketing metrics were used including:
- Awareness (also called Reach) – identifies the number of individuals who have seen content.
- Engagement – indicates the number of likes, shares, comments, reach and interactions with content.
- Social Listening – identifies key narratives, events, audiences, influencers, and content.

IRA management used these metrics and tools in its briefs to 'trolls', providing guidance on the type of content likely to generate comments, creating blog posts, and spreading targeted content through sock puppet personas.

These activities amounted to online intelligence gathering missions, conducted to develop situational awareness and sensemaking, and to contribute to the IRA's cultural understanding of target groups and demographics.

Traditional intelligence gathering missions were also conducted to develop and contribute to campaign awareness. Evidence of these missions formed part of the US indictment, confirming that, at the time, they were conducted by IRA employees. There is strong evidence these individuals were either formerly part of the broader ecosystem, (although it's unknown whether they were GRU, SVR or FSB), or had received training from an arm of the ecosystem.[33]

Example
According to the US Department of Justice, two senior IRA employees conducted a three-week intelligence gathering mission in June 2014, focusing on key electoral states. They were discovered with an evacuation plan. At the time of capture, they had compiled a report of their findings on American Politics and submitted it to their superiors in St Petersburg.[34] One was head of data analysis in the IRA's American Department, the other was reportedly the third-highest ranking IRA employee, with expertise in advertising and public relations.[35]

Conclusion

This report presented an overview of the IRA case study in the broader context of the Russian influence operations ecosystem. It focused on strengths and weaknesses of the IRA as a state-sponsored entity that perpetrated information operations. It was framed around four key themes, Governance and Ethics, Persuasive Technology and Techniques, Systems and Technology, and Campaign Awareness and Sensemaking, and addressed the associated research questions.

It highlighted that the IRA derived its strength from its establishment as a digital marketing firm, contracted out as a private military company to the Russian Government. The IRA wasn't hamstrung by ethical considerations. In fact, it acted with ethical fluidity, which helped it drive its messaging to a large audience. The report shows the IRA had a workforce of approximately 1000 people with varying degrees of competency in social media fluency, systems and technology, and persuasive techniques informed by psychology. It relied heavily on embedded social media tools to assist with its large volume, high-speed messaging campaigns. The report describes the evolution of the IRA as an influence operating business, developing tactics locally in the first instance, then applying what it had learnt to operations of a more strategic geopolitical nature, for example the 2016 US elections

Therefore, based on the key findings, the report recommends that as part of its emerging influence operations capabilities the Department of Defence: (i) Recruit and develop collective training regimes to establish an agile and innovative workforce that can develop experts at scanning the horizon for advancing technologies, to ensure ongoing awareness of evolving platforms and countering techniques. (ii) Augment native tool sets with off-the-shelf, third party tools for monitoring the social media landscape and identifying key vulnerable groups and individuals relevant to Australia's national interest; (iii) survey and develop methods for identifying, monitoring and measuring the predictability, and complex relationship, between online behaviour and its correlations to offline behaviour; (iv) coordinate an ecosystem combining Defence, Intelligence and non-government personnel to provide expertise across the full spectrum of operations; (v) engage with regional partners in the Indo-Pacific to increase resilience to malign or hostile information operations, and to increase, where possible and appropriate, local capabilities to combat influence.

Case Study 2

Cambridge Analytica

Cambridge Analytica

Introduction

This report provides an overview of the findings and recommendations of the Cambridge Analytica (CA) case study conducted as part of the Joint Influence Activities' (JIA) collaborative research project on influence operations actors, to assess the enablers of CA's influence campaigns. We begin by presenting our key findings and recommendations, then detail the strengths and weaknesses of CA's operations in relation to the four themes stipulated by JIA: Governance & Ethics, Persuasive Technology, Systems & Technology, and Campaign Awareness & Sensemaking. While this report highlights the key findings, additional insights can be found in the full report. Although CA is a private corporate entity, we nonetheless find that an assessment of the core attributes of its business model and technology raises considerations for a Department of Defence approach to Australia's information warfare capability.

Key Findings

Strengths of CA's Influence Operations

- The primary strengths of CA's business model derived from the company's ability to map and exploit the regulatory environment relevant to its operations.
- CA used large cohorts of online as well as offline data from multiple sources to profile millions of individuals and groups and target them with tailored messaging.
- CA utilised traditional and "quasi-experimental", data-intensive digital techniques in its political campaigns.
- CA gathered large amounts of qualitative and quantitative data and used it to develop psychological profiles that informed the design of targeted content to shift public opinion at scale.

Weaknesses of CA's Influence Operations

- The primary weakness of CA's business model was that it lacked mechanisms to foster legitimacy. This made the business unsustainable in a liberal democratic environment.
- The underpinning influence theories and models used by CA to profile and manipulate individuals and groups were simplistic, which weakened its efficacy.
- CA's influence operations relied on illegal data harvesting and use.
- CA did not attempt to measure the impact of its influence operations and likely did not produce the large-scale public opinion effects the company claimed.

Recommendations

1. Devise a code of practice for the ethical use of persuasive technologies that guarantees protection of liberal democratic principles and gives influence operations entities legitimacy.
2. Implement stringent data harvesting procedures that ensure data is collected legally.
3. Employ multidisciplinary teams of experts to analyse target audiences and develop contextually nuanced content.
4. Develop indicators and metrics for influence at the macro, meso and micro levels, leveraging both the human and analytical sciences.
5. Develop a strategy for gaining access to social media and other online data underpinning next generation persuasive technologies.
6. Deploy a qualitative-quantitative situational awareness strategy for mapping and visualising the information and influence environment.

Background

The Cambridge Analytica Story

CA was a political campaigning firm that operated between 2013 and 2018. It was a subsidiary of SCL Group, a company that had engaged in information operations globally since the early 1990s. Although legally separate, SCL Group, CA and another subsidiary, SCL Elections, overlapped to the extent that government investigations questioned whether the companies were one and the same. These investigations deemed SCL Elections and CA to be, in practice, the same company.
CA was established primarily to influence the United States (US) electorate to favour the Republican Party but was also likely formed to engineer broader societal change. Indeed, it seems the company was founded on the ideas of far-right media mogul Steve Bannon and investment from right-wing donor Robert Mercer. CA was never intended to make a profit. Consequently, the company's political campaigns were directed towards advancing far-right politics, not only in the US but states worldwide. As the timeline on the following page shows, over the period of 2013-2018, SCL Group, SCL Elections and CA conducted operations in Nigeria, Trinidad and Tobago, Kenya, Malaysia, the Philippines, the US and, potentially, the UK.

In theory, the company's role was to develop communications strategies to help clients reach voters more effectively. In practice, this involved conducting information and influence campaigns online by microtargeting voters and spreading disinformation. CA's microtargeting strategy relied on data analytics and personality profiling. Legal and ethical problems with this, particularly CA's collection and application of personally identifiable data, instigated a scandal after which CA and SCL Group entered administration. The companies are now defunct. However, they reportedly reincarnated as a 'new' political consultancy company, Emerdata.

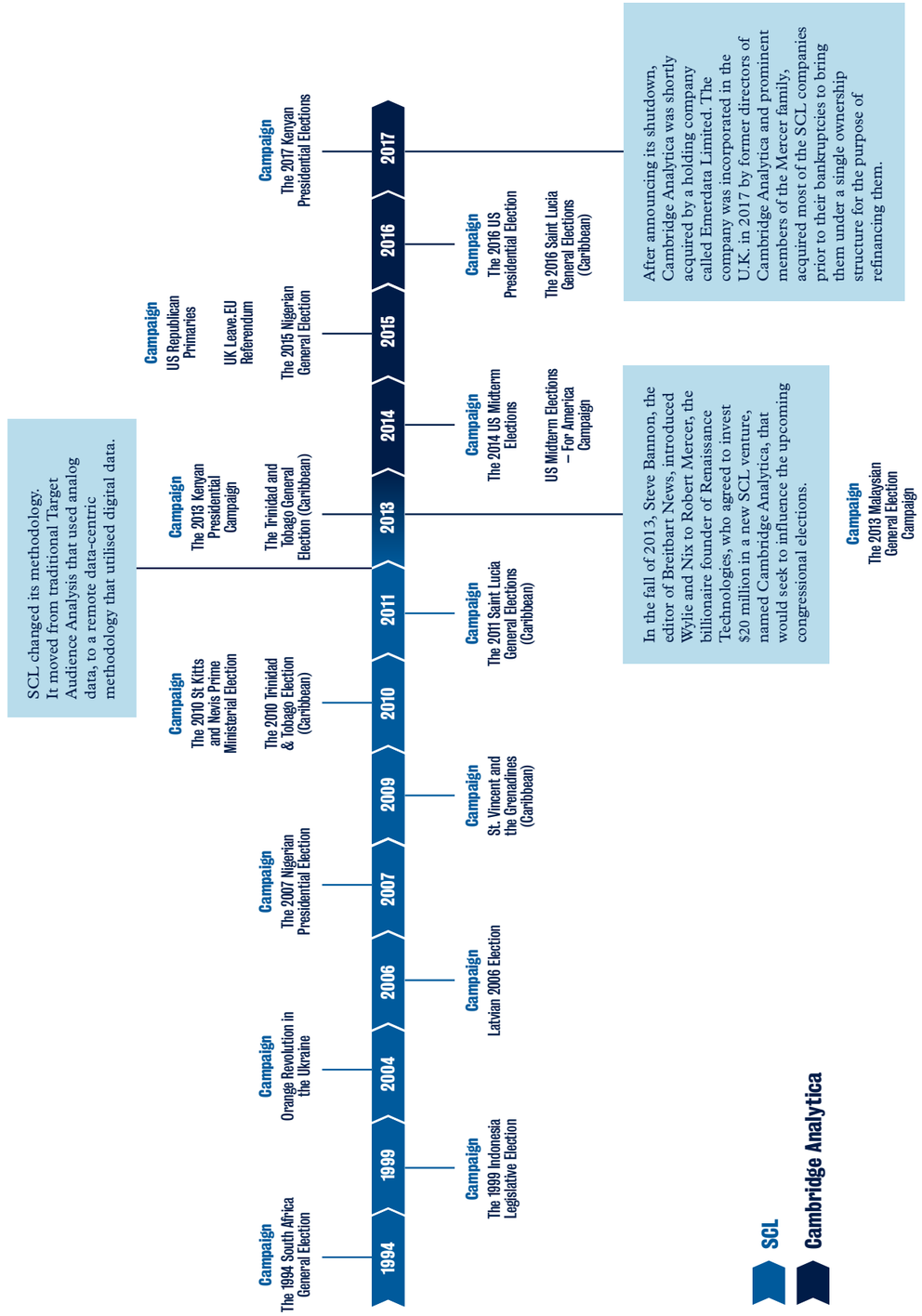


Figure 1: Cambridge Analytica Timeline (prepared by Stephanie Meek and Carmen Jacques)

Cambridge Analytica

Governance and Ethics

Melissa-Ellen Dowling

An understanding of Cambridge Analytica's (CA) governance structure is necessary to understand the organisational enablers and disablers of its influence operations. Accordingly, we evaluate the strengths and weaknesses of CA's business model from the perspective of legality and ethics. The benchmark for what constitutes 'ethical' conduct in this case study is the extent to which CA adhered to, or deviated from, norms and laws with respect to two core values of liberal democracy, privacy and consent. We find that CA's organisational structure and business model lacked sufficient legitimacy and transparency to sustain its operations in a liberal democratic environment.

Cambridge Analytica's objectives and strategies

As a private political consultancy, CA's objective was to influence voting behaviour in accordance with its main donor's preferred outcomes. Accordingly, CA's business model was built on the manipulation of elections via its ability to influence preference formation (e.g. deciding who to vote for) and articulation (e.g. casting a ballot) in the decision-making process. Its influence objectives informed every aspect of its business model. Across its campaigns, CA aspired to covertly prevent the articulation of preferences by the opposition, promote the articulation of preferences of its clients' supporters and persuade swing voters to reshape their preferences. The clandestine nature of its operations, in conjunction with the vast amounts of demographic and psychographic data it used, set CA apart from other political campaigning firms. Its unique selling point was its market position as a data analytics firm that could – given it allegedly had in its possession 'over 5000 points of data' on every voter – identify and target swing voters more effectively than other consultancies, particularly those operating outside digital data analytics. In short, the core capability of CA's business model was its ability to identify swing voters and target them.

Strengths of CA's business model

The key strengths of CA's business model derived from the company's ability to map and exploit the regulatory environment relevant to its operations.

A complex corporate structure shielded SCL (the UK parent company) and the Mercer Family from scrutiny (and potential liability) regarding CA's operations. Incorporated in the US, CA was one of several subsidiaries of SCL Group. CA was reportedly 90% owned by the Mercer Family Foundation due to the Foundation's upfront $15-$20 million investment in the company. The complex structure created a circular situation which gave CA the IP rights of its parent yet, due to an accompanying exclusivity agreement, transferred contracts to SCL staff who performed the work. It enabled CA to operate under the radar and mask the Mercer Family's involvement in an influence operations firm.

CA's business model was founded on the acquisition and use of accessible data. CA knew it would be able to access both psychographic and demographic data and accordingly made data collection and analytics central to its business. Aside from Facebook data, CA purchased data through data brokers such as Infogroup, Experian and Data Trust, and legally utilised other sources such as gun licence registries.

CA consistently operated according to the ideological preferences of shareholders which increased the prospects of the business. CA was driven primarily by ideological objectives. It was founded with a view to bolstering support for the US Republican Party. Its shareholders favoured right-wing political ideologies and its campaign history confirms its right-wing disposition. CA was therefore an ideologically driven actor, rather than a purely profit-motivated business. This shaped its business model and informed its strategies, goals and clients.

CA understood its online operating environment and with that knowledge was able to leverage a loosely regulated digital domain to its advantage. CA harnessed regulatory loopholes to collect data and post political content to social media platforms knowing that the attribution of information and disinformation is a major challenge in cyberspace. However, for CA, the attribution 'problem' was not a problem but an enabler. It meant the company could post inflammatory content knowing that it would be 'virtually' impossible to trace.

Weaknesses of CA's business model

CA's business model lacked mechanisms to foster its legitimacy which meant it was able – under its own guidelines at least – to operate in an unacceptable way. This led to a loss of legitimacy.

Although the complex corporate structure protected SCL legally, the porosity between SCL and CA precipitated both companies' insolvency due to perceptions of legal and ethical misconduct. Despite their legal separation, the practical separation between SCL Group and CA was negligible as the companies shared staff, offices, intellectual property and techniques. The porous corporate structure meant it was impossible for SCL Group to escape negative associations with CA. SCL reportedly lost clients to the point it was no longer a viable business. However, insolvency enabled CA/SCL to reincarnate as a 'new' consultancy, Emerdata.

CA's business model was unsustainable in a liberal democratic environment because its purpose necessitated the degradation of liberal democratic norms. CA's business model necessitated interference in democratic principles, processes and norms, thereby undermining democracy. CA's covert manipulation of decision-making processes degraded the integrity and legitimacy of the electoral process, and electoral outcomes. Its business model did not preclude collaboration with known malign foreign entities and, thus, potentially made CA a vector for foreign interference. By eroding the legitimacy of democratic processes, CA concurrently diminished its legitimacy as a business.

The 'unique selling point' that differentiated CA from other firms relied on illegal data harvesting. Although CA lawfully purchased demographic and consumer data from data brokers, its operations depended on collecting personally identifiable psychographic data. For CA, data needed to be personal and identifiable to enable microtargeting. It would have defeated the purpose of CA's convert influence campaigns if people knew they were being targeted with unattributed political messaging. Shortcomings in matters of privacy and consent contributed to CA's insolvency. CA's collection of Facebook data was deemed illegal in the US and UK because it (1) misrepresented the type of data it would collect, (2) collected personal identifiable data, (3) collected data without user consent and (4) used data for purposes outside reasonable user expectation – i.e., for political campaigning.

Key Findings

- The primary strengths of CA's business model derived from the company's ability to map and exploit the regulatory environment relevant to its operations.
- The primary weakness of CA's business model is that it lacked mechanisms to foster legitimacy. This made the business unsustainable in a liberal democratic environment.

Persuasive Technology

Matteo Farina

CA heavily relied on technology for its operations. Technology was used to harvest, combine and analyse social media and other data. CA also used technology to reach and influence specific groups. Furthermore, it appears that CA relied on technology for its operations because technology has changed how people communicate and access information; it affects politicians and how they interact with voters; it is persistent, ubiquitous and allows anonymity; it can store and process huge volumes of data; it has many modalities and it can scale. Finally, technology is also interactive, personalisable and potentially persuasive. However, whether technology affects peoples' attitudes and behaviours is still open to debate.

Strengths of CA's Persuasive Technologies

CA claimed the ability to covertly manipulate both individuals and groups. CA used online manipulation for its operations. Manipulation might exploit individuals' vulnerabilities (such as dark triad and personality traits) and affect their decision-making processes. CA's manipulation took different forms, including direct intervention using identity-based reasoning to inflame and exploit group dynamics, disinformation spread through Facebook groups that were polarised or primed by CA's intervention, hyperbolae and mobilisation of rage (affect heuristic), and the alleged release of hacked kompromat.

CA had the capacity to access large data sets from multiple sources to profile millions of individuals and target them with personalised content. CA used technology to collect huge amounts of demographic and behavioural data to create psychological profiles of millions of people. It used these profiles to develop messages tailored to specific groups of voters. CA focused especially on persuadables, people who were more likely to be influenced by its political campaigns. CA targeted persuadables with messages that reflected their psychological characteristics. CA relied heavily on social media messages that consisted of visual and textual elements. It seems that these messages were highly effective. They exploited persuadables' cognitive biases and apparently affected their political preferences. CA used psychological profiles because they describe individuals' personalities in terms of a few basic dimensions. Moreover, these dimensions were used to develop tailored content to shift political opinion at scale.

Weaknesses of CA's Persuasive Technologies

The model used by CA for profiling target audiences was simplistic. CA used the Five Factor Model (FFM) to create psychological profiles of millions of individuals. CA was mainly interested in the FFM because it could use Facebook data to computationally and accurately predict personality traits of large groups of individuals. The FFM focuses on five dimensions of human personality: Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness to Experience. Some studies argue that the FFM may be universal. However, others suggest that this might not be the case. In fact, the FFM is not identifiable in all cultures, and seems to mainly describe Western, Educated, Industrialised, Rich and Democratic (WEIRD) populations. More importantly, personality profiles are only a small component of market segmentation (or Target Audience Analysis). Other factors include cognitive processes, social identities, norms, networks and interactions, power dynamics and social movements.

CA's model did not consider that political leanings might not depend on personality traits only. Although personality traits might affect someone's political affiliations, they do not apparently cause them. In other words, both political attitudes and psychological traits might depend on multiple factors.

CA's model did not consider how individuals express themselves online and how information spreads across online social networks. Online profiles are carefully curated and take on a 'performative' component. Whilst apps such as thisisyourdigitallife may yield accurate profiles, other aspects of online behavior are likely to be skewed by performative or polarising processes. Moreover, it is unclear whether CA considered how different types of information spread on social networks.

Key Findings

- CA used large cohorts of online as well as offline data from multiple sources to profile millions of individuals and groups and target them with tailored messages.
- The underpinning influence theories and models used by CA for profiling and manipulating individuals and groups were simplistic which weakened its efficacy.

Cambridge Analytica

Systems and Technology

Matteo Farina

CA used various systems and technology for its campaigns. These systems and technology enabled CA to collect and analyse large amount of data to create psychological profiles of millions of individuals. Using these profiles, CA was able to microtarget voters with personalised messages and political ads. Whether these messages were effective is still unknown.

Strengths of CA's Systems & Technology

CA used large quantities of data for its operations. Although CA collected online and offline data from multiple sources it relied heavily on Facebook for harvesting information about large groups of individuals. CA presumably used Facebook because it has 2.7 billion monthly users, making it the world's most popular social media platform. Through Facebook, CA accessed a diverse and extremely large pool of individuals. In addition, collecting data via Facebook was simple and inexpensive. Moreover, Facebook data was rich in computationable, demographic and behavioural information which appeared to be naturalistic. Finally, and more importantly, CA needed only a single app to collect Facebook data and accurately predict psychological profiles of individuals at scale. The app used by CA for its profiles was thisisyourdigitallife. Although it seems technical information about this app is not publicly available, it appears it was similar to myPersonality, an app developed by researchers at the Psychometric Centre at Cambridge University. Both apps used Facebook Likes to make predictions about individuals' personalities and personal attributes. These predictions were made using the FFM described in the Persuasive Technology section of this report.

CA combined traditional and "quasi-experimental" approaches in its political operations. CA used traditional and "quasi-experimental" techniques in its political operations. On one hand, it utilised market segmentation, interviews, focus groups, surveys, TV commercials, canvassing and direct mails. On the other, it appears it used more sophisticated, data-intensive digital techniques. These included direct and indirect psychological profiling at scale via social media data, kompromat, hacking, mobilisation of rage, and the use of psychologically based hyperbolic narratives that exploited cognitive biases, inflamed group dynamics and generated identity-based reasoning.

CA developed and provided its clients with a single 'one-stop-shop' platform to effectively manage their political campaigns. CA managed all its political campaigns using a single online platform called RIPON. RIPON allowed CA to control all aspects of a political campaign including fundraising, voter profiles, message design, psychographic data, online marketing, campaign scheduling, teams' management and so on.

Weaknesses of CA's Systems & Technology

Potential illegal use of data. Although Facebook allowed academics and developers to collect personal information about users it was, at the time, illegal to use the data for political campaigning. A British Commissioner's Office investigation determined CA illegally used Facebook data for its political operations.

Efficacy of its operations. In its promotional materials CA claimed its operations were highly effective. However, there is evidence it rarely measured the efficacy of the operations and, when it did, apparently used rudimentary techniques, such as click-through rates and unspecified post-election analyses.

Moreover, although some literature suggests that microtargeting, especially through social media platforms, might affect consumer behaviour, whether this is applicable to political campaigning, and therefore CA's operations, is an open question. For example, some studies suggest that political microtargeting might have an adverse effect – rather than triggering support, it may generate reactance (a backlash).

Furthermore, many journal articles argue that CA had a limited impact on elections and that the psychographic model it used was inaccurate. Dr Kogan, the academic who developed the thisisyourdigitallife app used by CA to create its profiles, apparently said the app's personality predictions were not particularly accurate. The limited accuracy of CA's predictions may have been due to several factors: that the effectiveness of Facebook Likes as predictors of personality might weaken over time, self-reported information collected through questionnaires is often unreliable, personalised messages are context dependent. Finally, predicting the outcomes of political elections is an extremely complex task. Election forecasts depend on multiple variables, some of which are known, others unknown. Moreover, it appears that some of these variables might be interconnected and have different values that can change over time. In summary, it is impossible to establish the degree to which CA's operations impacted election results.

Key Findings

- CA utilised traditional and "quasi-experimental" data- intensive digital techniques in its political campaigns.
- CA's influence operations relied on illegal data harvesting and use.

Campaign Awareness and Sensemaking

Melissa-Ellen Dowling

CA needed to map its operating environment because the company's modus operandi involved identifying weaknesses in socio-political systems and leveraging those weaknesses to its advantage. CA was, therefore, unable to identify pressure points without developing and maintaining situational awareness. Through an evaluation of the strengths and weaknesses of CA's sensemaking practices, we find that CA's pre-campaign sensemaking practices were robust and enabled its operations, while post-campaign sensemaking did not occur systematically and this may have compromised the efficacy of subsequent operations.

Strengths of Cambridge Analytica's Sensemaking Practices

CA's pre-campaign sensemaking processes were robust. They enabled the company to identify socio-political vulnerabilities and subsequently exploit those vulnerabilities as per its business objectives. CA mapped a target state's cultural and political traditions as well as its ethnic and economic tensions. Mapping was conducted via surveys, secondary research, interviews and focus groups. The process was fundamental to generating situational awareness of socio-political fissures and pressure points that could be exploited to influence voters' political preferences. CA's emphasis on cultural factors suggests there was a human element in persuasive technologies that was important to its operations.

CA needed to grasp non-technical aspects of the societies in which it operated, as well as technology's role in analysing target audiences and promulgating content to influence voter preferences. The fact it collected mass data on voters suggests it was aware of the important role technology played in voters' lives and its own operations.

CA took large amounts of qualitative and quantitative data and used it to develop psychological profiles that informed the design of targeted content for the purpose of shifting public opinion at scale. CA's use of quantitative and qualitative data was key to its microtargeting strategy. It was able to identify swing voters and their vulnerabilities based on a holistic understanding of the socio-political environment and individual dispositions within that context.

Because CA conducted extensive qualitative research on the culture and traditions of its target audiences it discovered it could access psychographic data from people paid to take an online survey in the US. CA interpreted the data and identified persuadable voters. Identifying persuadable voters was a key component of its early campaign awareness approach and fundamental to its influence operations. It meant CA could reach consensus on its targets, reflecting a functioning sensemaking process. A shared 'cognitive cause map' emerged, and this drove the company's mission and tactics.

The defining features of CA's 'quant-qual' sensemaking process were:

1. Target audience analysis to understand key issues and political dynamics.
    a. Qualitative research – secondary research, interviews, focus groups.
    b. Quantitative research – data collection using a blend of online/offline survey methods.
2. Data analytics to 'segment the population into actionable groups'.

Following these steps, CA was able to target identified individuals and groups, and tailor messages and other content as part of its communications strategy.

CA maintained situational awareness intra-campaign by tracking public opinion and dynamic monitoring of attitude changes via a control group. CA's intra-campaign awareness had two key dimensions: (1) the need to remain aware of a dynamic socio-political environment and (2) the need to assess its role in affecting that environment, and projecting on future developments, in the context of its actions and a fluid situation.

CA had systems to track the socio-political environment of an electorate during its campaigns. For example, it monitored political polls and public opinion to keep abreast of its candidate's electoral chances. Its analysts produced 'intelligence reports' detailing changes to the political environment. These reports focused on popular perception and awareness of candidates and evaluated electorate composition and the political ideology of key voter segments. There is a lack of evidence of intra-campaign sensemaking processes in relation to CA's effect on the environment. Despite this, it appears that CA did have methods to assess its role in relation to the changing political environment. For example, CA compared attitude shifts of a control group (voters not messaged) with those of the group it did message.

Weaknesses of Cambridge Analytica's Sensemaking Practices

During its campaigns, CA was unable to establish causal connections between its operations and changes to public opinion. As detailed in 'Theme 2: Persuasive Technology', CA likely deployed rudimentary statistical modelling that was incapable of processing variable complexities. Despite the use of control groups, it would be difficult to measure with any degree of certainty the effect CA's operations had on the popularity of its candidates due to the large number of variables that might have accounted for fluctuations in public perceptions.

Cambridge Analytica

CA lacked systematic post-campaign sensemaking practices which means it was potentially unable to evaluate the efficacy of its influence methods. Evidence suggests that CA adopted a simplistic approach whereby if its sponsored candidate won an election or even gained a parliamentary seat the CA campaign was deemed a success. As with its intra-campaign mapping, this approach neglected to account for the myriad non-CA variables that would have influenced the outcome. Without stringent post-campaign evaluation and assessment processes, CA may have employed ineffective methods in future campaigns, jeopardising its business.

Key Findings

- CA took large amounts of qualitative and quantitative data and used it to develop psychological profiles that informed the design of targeted content for the purpose of shifting public opinion at scale.
- CA did not make any attempt to measure the impact of its influence operations and likely did not produce the large-scale public opinion effects attributed to them by the company.

Conclusion and Recommendations

This report presented an overview of the CA case study. It focused on the strengths and weaknesses that characterised this private entity and its political operations with a view to deriving insights that could inform future Defence operations. The overarching strength of CA's business model was its ability to efficiently map and exploit the regulatory environment in which it operated, using a combination of traditional and "quasi-experimental" techniques. It was able to accomplish this by gaining access to and exploiting large cohorts of data gathered from multiple sources which were utilised to profile, microtarget and influence individuals as well as public opinion. The report also identified key weaknesses of CA's operations that hindered its capacity to accomplish its aims more effectively. It finds that CA's business model was unsustainable because its methods did not stand up to public scrutiny. In addition, the report shows how the simplicity of its profiling, as well as its inability to measure the efficacy of its operations, cast doubt on the veracity of its claims to change electoral outcomes.

Based on these key findings, the report recommends that the Department of Defence's emerging influence operations capabilities are: (i) founded on a code of practice which protects liberal democratic principles and ensures the legality of data collection procedures; (ii) guided by a strategy for accessing and collecting social media data that evolves with changing digital technologies; (iii) driven by multidisciplinary teams to analyse and develop indicators and metrics for influence operations at the macro, meso and micro levels; and (iv) informed by combined qualitative-quantitative situational awareness strategies for mapping the outcomes of information and influence campaigns.

Case Study 3

Facebook

Facebook

Dr Garry Young and Dr Peter Job

Introduction

Facebook was selected for analysis as a mature and popular social media platform, broadly representative of others in common use now and platforms that may emerge in future. This report summarises research findings under four themes: governance and ethics, persuasive technology, systems and technology, and campaign awareness and sensemaking. Our research identified important considerations for Defence as it seeks to substantively enhance Australia's Information and cyber domain capabilities to counter grey zone threats. The findings and recommendations summarised in this report are discussed in greater detail in a full Case Study available online at: https://dri.unsw.edu.au/groundbreaking_post/understanding-mass-influence/.

Key Findings

Facebook's strengths as a persuasive technology company

- Facebook has a centralised management structure and employs a business model that aggressively acquires and suppresses rivals and provides a highly attractive means of targeted advertising
- Facebook successfully employs psychological techniques of persuasion to facilitate the 'right' kind of user engagement in a manner that is not dependent on the epistemic value (i.e., the veracity) of content
- Facebook adopts a range of strategies to maintain market dominance
- Facebook uses several different means to monitor and measure its effectiveness, most of which are opaque to users.

Facebook's weaknesses as a platform for influence operations

- Facebook is a very efficient platform for the propagation of misinformation
- The methods Facebook uses to persuade users are susceptible to manipulation by malign actors
- Many of Facebook's contractors and agency employees are underpaid and under resourced
- Much of Facebook's transparency is simply about appearing, rather than being, transparent
- Any response to malign activity on the platform should be predicated on the assumption that the present model will continue and, with it, the potential for misuse.

Recommendations

1. Devise a framework and strategy for clear and transparent public communication. This includes guidelines and definitions for what constitutes ethical persuasion and the ethics of the operation, as well as having the capacity to differentiate between authentic and inauthentic patterns of social engagement in the context of true or false content.

2. Develop an understanding of how to operate in a social media landscape with powerful distribution and production networks and concentrated ownership. To achieve this, consider investing in the development of information analytics and leveraging Australian centres of excellence to help develop fit-for-purpose sovereign analytic tools and techniques, including netnographic analysis (a technique for the cultural analysis of social media and online community data) as a potential means to identify malign online actors and their behaviours and vulnerabilities.

3. Recruit a diverse workforce, with the skills to identify nefarious activity disguised as benign social engagement. Natural language algorithms are most effective in conjunction with human fact moderators. However, workers should be provided with support services, including resiliency training and counselling.

4. Effective liaison with platform owners is important in countering malign influence. Examples of government cooperation with Facebook exist and could be used as models. Existing obligations under the Telco Act may provide the basis for the establishment of such a facility. Consideration should be given to whether the liaison facility should consist of a dedicated ADF operation or a whole of government operation.

Background

In February 2004, TheFacebook (as it was then called) was created by Mark Zuckerberg and others to serve the Harvard student community. Due to its popularity it expanded to other universities and, in September 2006, the enterprise, now known as Facebook, became publicly available to people 13 or older. Today, Facebook is the world's most widely used social media platform and, in 2020, claimed a global workforce of 58,604. In addition to the Facebook platform, the company's assets have grown to include Instagram, WhatsApp and the digital gaming company Oculus VR. Facebook also does business through subsidiaries in other countries. See Appendix 1 for key Facebook timelines.

Facebook's US Securities and Exchange Commission (SEC) submission for 2020 reported 1.84 billion worldwide Facebook daily active users (DAUs). Monthly active users (MAUs) numbered 2.80 billion, and family daily active people (DAP) (that is, individuals who visited at least one of the following sites – Facebook, Instagram, Messenger and WhatsApp – daily) totalled 2.60 billion. Social Media News reports that in January 2020 there were 16 million monthly active Facebook users in Australia (see Figure 1 for a breakdown of Australian users by age). However, this does not represent 16 million discrete users, as some individuals have multiple accounts, while other accounts belong to organisations. As a social media business, Facebook is a multisided platform. It has, however, a primary source of revenue: advertising. Facebook is a for-profit corporation (see Figure 2) working to maximise its shareholders' welfare. It does this by paying dividends to its shareholders or increasing its stock value, or both. Shareholder welfare is maximised when advertising revenue is maximised.

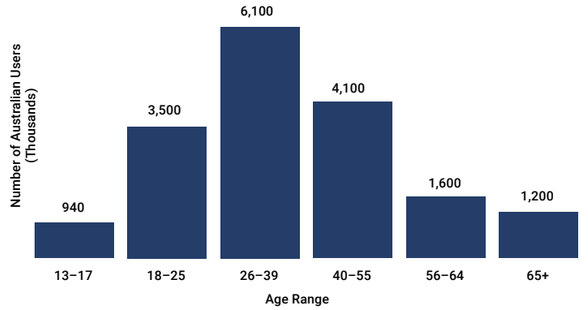


Figure 1: Number of Australian Facebook users by age

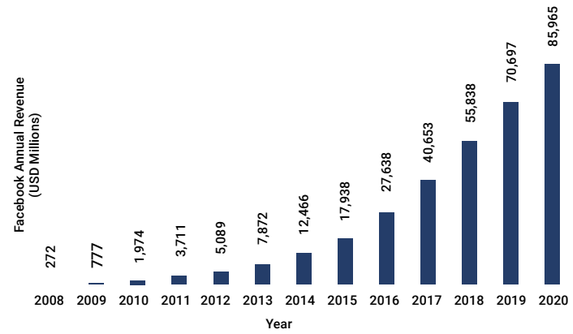


Figure 2: Facebook annual revenue

Findings and Discussion

Governance and Ethics

Key Findings – Governance and Ethics

Strengths

- Facebook employs a business model of aggressively acquiring or suppressing potential rivals.
- The Facebook model of targeted advertising based on user data is highly attractive.
- Decision making in Facebook Inc. is highly centralised.
- Facebook profitability is such that even apparently substantial penalties are not sufficient to provide an incentive for the company to modify substantially its business model.
- Facebook's right under Section 230 of the Communications Decency Act (CDA) to moderate content and users has been mostly affirmed by US courts.

Weaknesses

- Facebook is a very efficient platform for the propagation of misinformation.
- Critics have argued that Facebook practices may put users at risk.
- The practice of using algorithms to target users with ads means that the more users express interest in certain factors, including extreme political positions, the more likely they are to receive information about them.

The Facebook business model, engagement, and profit motive

Facebook will not give up or substantially modify its business model. The Facebook model of targeted advertising based on user data is highly attractive, with no cost of goods sold, no marketing costs and no selling costs, creating what Len Sherman describes as a "trifecta of high scale and high growth and high profit margins unmatched by any high-tech company". This model lies at the heart of Facebook business success and profitability. While Facebook may make certain changes, it will not willingly give up the collection of user data or the use of it to make user profiles for targeted advertising.

Facebook is a very efficient platform for the propagation of misinformation. Users do not need to create content, just share it. The longer users spend on Facebook, the longer Facebook has to collect their data and feed them ads. This is true regardless of the nature of the content keeping them online. In fact, content that resonates emotionally with users, and appeals to their preferences and beliefs, is more likely to be shared, regardless of its provenance or veracity.

Facebook has a strong incentive to keep its users engaged. Its algorithms therefore prioritise content that appears to align with users' interests, rather than the accuracy of veracity of content, as research indicates individuals are more likely to accept information consistent with their existing beliefs than information that contradicts them, even when it is factually inaccurate or otherwise misleading. If inauthentic content increases user engagement, it is as useful as any other in enhancing Facebook profitability. This means that Facebook is not financially incentivised to weed out mis- or disinformation. In fact, quite the reverse, particularly as such content is shared more frequently than authentic content and therefore helps cultivate engagement through 'likes', shares, and by posting comments.

Facebook's lack of a financial motive to remove malicious content, if it is keeping users engaged, is important because there is evidence a significant number of Facebook users use the platform as their primary news source. However, Facebook must balance financial incentives with at least the appearance of embracing social responsibility and integrity regarding content, as failing to do so could have a detrimental impact on business. For this reason, Facebook is open to some reforms regarding privacy and content, if made in a way that does not challenge its business model.

Facebook

The risk to users
The practice of using algorithms to target users with ads means that the more users express interest in certain factors, including extreme political positions, the more they are likely to receive information about them, whether through targeted advertising or by interacting with other users with similar interests. This may in turn increase their response to, and engagement with, such material and positions, thus perpetuating the cycle. There is accordingly a potential conflict of interest between user privacy, social harmony, mis- or disinformation, and Facebook profitability.

Critics have argued that Facebook practices may put users at risk. For example: In October 2018, Facebook linked 540,000 of its users in Saudi Arabia to the ad preference "Homosexuality". In the European Union (EU), Facebook has labelled 73% of users with potentially sensitive interests, possibly in contravention of EU law. These include political opinions, sexual orientation, personal health issues and other matters. The use of potentially sensitive personal data may enable malevolent actors to target ad campaigns attacking specific groups based on it, or to otherwise use the data in a malevolent manner.

Power and profitability at Facebook
Decision-making in Facebook Inc. is highly centralised. A two tier-share structure places effective control of the company in the hands of the Facebook Board of Directors, and the Chairman and Chief Executive Officer, Mark Zuckerberg. Ordinary shareholders hold little power and the board does not face any challenge from them regarding its policy decisions. Facebook employs a business model of aggressively acquiring or suppressing potential rivals. Primary acquisitions are detailed in Timeline 1 (Appendix 1), while a more complete list of companies acquired by Facebook is provided in an annex to the detailed Facebook Case Study.

Facebook profitability is such that even apparently substantial penalties are not sufficient incentive for the company to modify substantially its business model. On 24 July 2019, the US SEC announced a US$100 million fine against Facebook for the Cambridge Analytica debacle. The same day, the Federal Trade Commission (FTC) imposed a US$5 billion fine on the company for violating a 2011 FTC order by deceiving users about the privacy of their data. With shareholders anticipating the possibility of a more severe outcome, the share price jumped after the announcements; the penalties clearly not sufficient to make shareholders lose faith in the Facebook business model. Two of the three FTC Commissioners dissented, arguing the penalty was not sufficient to force change.

Facebook and reform
When faced with regulations or demands for change, Facebook responds by acting in a manner that prioritises its business model over substantial reform. It is likely to continue to behave in this way in the future. This is evident in its response to the FTC mandate in 2011, and to its response to both the General Data Protection Regulation (GDPR) and moves made by Apple to enhance privacy on its devices (discussed below). That said, there is evidence that Facebook is susceptible to public opinion, as this can affect its profitability (as alluded to earlier in relation to social responsibility and integrity). Despite maintaining profitability, the Cambridge Analytica scandal resulted in a slowdown in user growth in 2018. Facebook shares subsequently plunged 19%. This prompted Facebook to promise reform and to make some changes. Zuckerberg, for example, pledged a number of reforms in his testimony to the US Congress in April 2019, including restricting third-party access to Facebook user data; discontinuing the company's purchase of user data from private data mills; investing in AI detection algorithms; employing thousands of new cyber security personnel to prevent the spread of disinformation; requiring developers to get user approval before accessing posts and private data; and requiring advertisers running political campaigns to confirm their identity and location, display their ads publicly and indicate who paid for their promotions.

In March 2019, Zuckerberg posted a blog in which he claimed that Facebook was "pivoting to privacy". This included an emphasis on privacy in personal interactions, a commitment to end-to-end encryption on Facebook apps, and a commitment to end the long-term storage of data. He also committed to interoperability, to allow people to communicate across apps and networks. Critics responded with scepticism, arguing the strategy was centred on Facebook's desire to own the "one-to-one private ephemeral space" of personal communication through apps, which allows Facebook to leverage data from the Facebook platform and provide targeted advertising on privacy-focused platforms.

While it did make some changes, Facebook remained forthright and arguably ruthless in protecting its interests. In an investigative article in November 2018, the New York Times reported that Facebook's Chief Operating Officer Sheryl Sandberg had aggressively lobbied against Facebook's critics, attempting to shift public anger towards rival companies and ward off regulation and even employing a Republican opposition research firm to discredit opponents.

Facebook reforms are also limited to areas that do not threaten the viability of its business model. Observers note that Zuckerberg's apology to Congress fits a pattern of apologising and moving on in the face of criticism: a pattern that can be traced back to Zuckerberg's apology to Harvard University when he was reprimanded for posting pictures of female Harvard students on his website, Facemash, without their consent. Critics argue that the imperative to protect the Facebook model involving targeted advertising based on user data means that almost any fine, scandal or negative publicity poses less of a threat to Facebook's profitability and business model than any substantial reform.

Challenges to the Facebook model
Arguably, Facebook is vulnerable. The Facebook business model based on data collection and profiling is essential to the continued existence of Facebook in its present form, and its ability to deliver profits to its shareholders. Alternative models for an online media company that allows users to share information on a social platform without the retention of data have been floated by various commentators, and even rolled out in an embryonic state. The viability of the Facebook model could also be undermined by regulation, such as legislation by national governments prohibiting the harvesting and retention of user data, possibly supported by international agreements.

The Facebook business model has faced somewhat of a challenge from the EU GDPR. This includes some restrictions on the collection and use of personal data. The situation in the EU is now significantly different to the legal situation in the US regarding the retention and use of data. Facebook has responded to the EU GDPR by focusing its efforts on speeding users through consent processes and gaining user consent, rather than reducing data collection. It has also changed the jurisdiction of users in Africa, Asia, Australia and Latin America from the EU to the more lenient US privacy laws.

The Facebook model is also facing a challenge on Apple platforms. Apple is moving to allow users to opt out of data collection. In the first half of 2021, Apple introduced privacy consent for apps on the Apple platform, including Facebook, collecting their data. Apple will block data collection for those who decline. Facebook attempted to pre-empt this development by introducing a pop-up screen effectively urging users to opt-in to data collection, arguing it will allow Facebook to "continue to give people better experiences". It has denied there is any trade-off between collecting data to provide targeted ads and user privacy, claiming that "in fact we can provide both".

Facebook is presently facing a number of political and legal challenges that may impact its operations and business model. In October 2020, the Democratic majority of the US House Judiciary antitrust subcommittee released a report which concluded that the domination of Apple, Amazon, Facebook and Alphabet is impacting US democracy and the US economy. It suggested parts of these businesses be broken up. In December 2020, the US FTC, in conjunction with a coalition of 46 states and districts, filed parallel anti-trust lawsuits against Facebook, accusing the company of maintaining a social networking monopoly by means of anti-competitive conduct. Outcomes could include: the forced divestiture of Instagram, WhatsApp and other assets; prohibiting Facebook from imposing anti-competitive conditions on software developers; requiring the company to seek approval for future acquisitions; or forcing it to allow users to post material across competing social network platforms, thus facilitating competition.

Facebook's right under Section 230 of the 1996 US Communications Decency Act (see Appendix 1, Timeline 3) to moderate content and users has largely been affirmed by US courts. In response to a lawsuit by the conservative organisation Freedom Watch and YouTube personality Laura Loomer, a US Federal Appeals Court affirmed Facebook's right to ban conservative Facebook users who had violated Facebook's terms of service, finding that Facebook had not in fact violated the US First Amendment, as this prohibits "only government abridgement of freedom of speech". However, legislation signed by President Trump in 2018 removed protection under Section 230 for material related to sex trafficking, making Facebook potentially liable for it. Some US politicians are arguing for further changes.

Persuasive Technology

Key Findings – Persuasive Technology
Strengths
- Facebook exploits people's motivation to connect and share with each other to facilitate prosumerism: the production and consumption of information by its users.
- The big data derived from this activity amounts to a valuable commodity, marketable to third-parties (e.g., advertisers).
- Facebook has been effective at employing a number of psychological techniques of influence.
- Prosumerism is not reliant on the epistemic value (i.e., the veracity) of content to produce its valuable commodity. The current post-truth era therefore benefits Facebook.

Weaknesses
- The method by which Facebook enhances prosumerism – the algorithmic-based preferencing of content which creates filter bubbles and subsequent echo chambers – is susceptible to manipulation by malign actors.
- Facebook's business model is not incentivised to remove mis- or disinformation; but the company is vulnerable to changes in public opinion and legislation, and so needs to maintain a delicate balance between maximising profit through its strategy of promoting prosumerism and at least appearing to embrace its social responsibility when it comes to combating 'fake news' and malign activities.
- Arguably, Facebook manipulates rather than persuades its users to engage as prosumers.

Facebook users can be thought of as prosumers because they are both the consumers and producers of information. They are encouraged to increase their commitment to the platform through a process Fogg and Eckles refer to as a behavioural chain, whereby users proceed through the initial stages of discovery and superficial involvement before truly committing to their role as prosumers (the end-state intended by the Facebook design).

To demonstrate their commitment, Facebook users are encouraged to keep their personal profiles updated, invite friends, respond to others' contributions and return to the platform often. Facebook facilitates these steps by making engagement quick and simple (e.g., they often require only one click), and tracks user activity, such as the signalling of preferences (in the form of 'likes' and 'dislikes'), the adding of connections (i.e., allowing access to one's password protected email list), friending/unfriending, and so on.

Facebook

The long-term viability of Facebook depends on the company maintaining a delicate balance between attracting an audience and exploiting its natural resources (i.e., members' attitudes and behaviour, and propensity to share and connect), while preserving its credibility. Facebook therefore presents as an audience engagement tool (see Figure 3), accommodating and entertaining users through the social connections they develop via its platform. In short, Facebook seeks to influence individuals to participate more widely and more often, and in the right way, and in so doing gather more of the source material (small data points) on which its most prized commodities – big data and predictive algorithms – depend.

The Facebook model is dependent on the authenticity of its users and their identities, as this is the key to accurate profiling and targeting ads. The company's platform integrity relies on users providing their correct identities. As such, Facebook is far less concerned with the veracity of content, which does not impact its profiling or targeted advertising, although it is aware of its need to appear to be socially responsible.

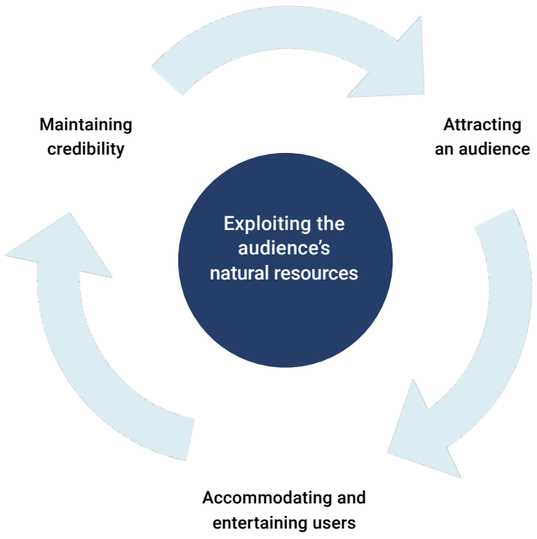


Figure 3: Facebook as an audience engagement tool. Attracting an audience by accommodating and entertaining them while maintaining credibility.

Maintaining engagement
A key aspect of being social on Facebook is sharing. Acts of sharing enable users to develop networks of 'friends' who then become a trusted source when receiving or passing on new information. An individual's 'popularity' on the platform is rooted in the connections they establish as part of their online network (see Figure 4). This connectivity built on trust translates, for Facebook, into a quantifiable commodity (i.e., the amount of engagement). Users can also enhance the saliency of their 'popularity' through increased social presence which is calculable by the number of posts, 'likes', images and 'friends' they have.

Cultivated and trusted sharing can make users more susceptible to mis- or disinformation – commonly known as fake news – especially given that individuals do not always (or often) make entirely rational decisions about sharing information. Facebook capitalises on this because it is designed to facilitate non-rational sharing. It nudges us to share, Waldman claims, by scratching its users' social itches, often through the ease by which we can click 'like' on new content and share it with others. To illustrate: The findings of Facebook's (infamous) 2012 emotional contagion study reveal that emotional content yields higher levels of engagement in terms of comments and shares compared to emotionally neutral content (e.g., a post about food or interior design). Facebook therefore has an added incentive – in terms of increasing prosumer engagement – to prioritise or privilege news or other content that will elicit an emotional response. This incentive remains irrespective of the epistemic value of the content; irrespective, that is, of its truth or falsity.

Cialdini has proposed seven techniques of influence compatible with the notion of non-rational persuasion that can be used to explain Facebook's success at shaping its users' attitudes and behaviour (see Figure 5).

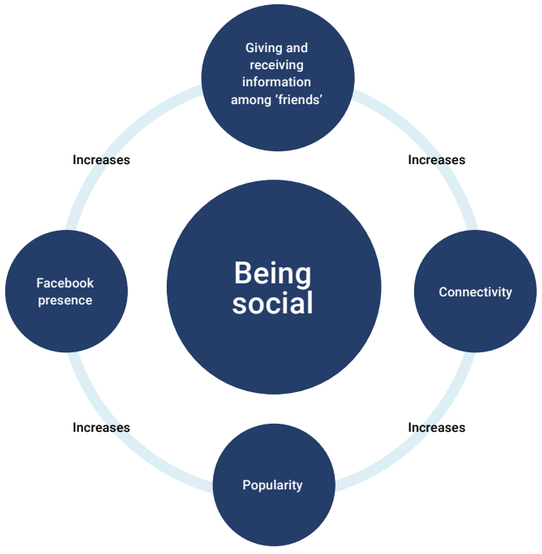


Figure 4: Cultivating a trusted social environment on Facebook

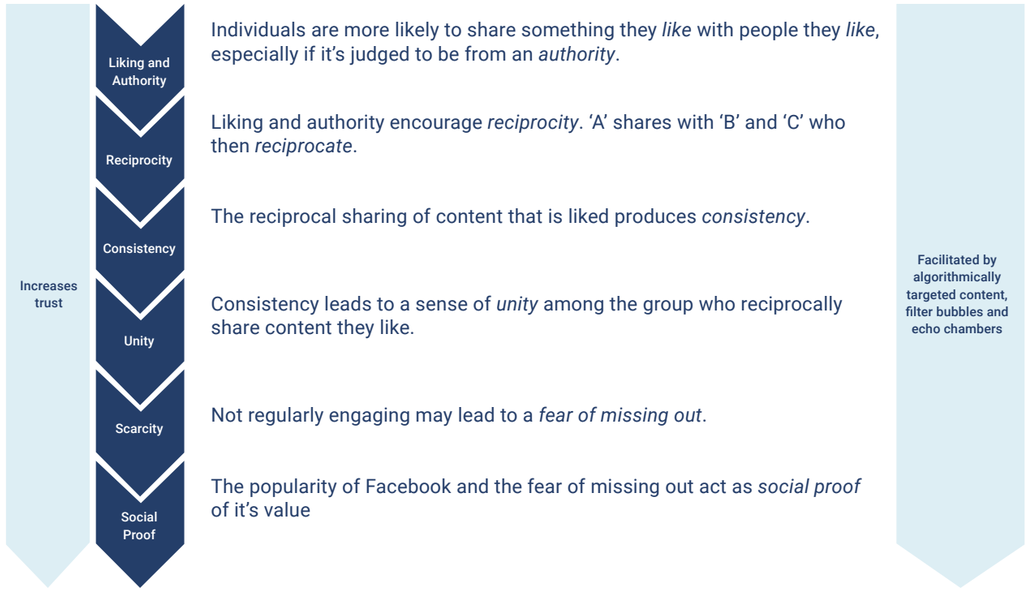


Figure 5: Cialdini's influence techniques applied to Facebook

Facebook

Systems and Technology

Key Findings – Systems and Technology

Strengths

- Facebook has maintained market dominance by utilising a range of strategies.
- Facebook employs a high-quality public relations team to support its public image and respond to critics.
- Examples of government cooperation with Facebook and other social media entities already exist and could provide the basis for a government social media liaison facility.
- The Facebook Oversight Board allows an independent body to act as final arbitrator of content on Facebook.
- Fact checking is now a key part of Facebook operations.

Weaknesses

- Some employees have likened working for Facebook to being part of a cult, owing to the degree of conformity required.
- The benefits and support provided to official full-time Facebook employees are not extended to the many under subcontracting arrangements.
- Critics have expressed scepticism about the ability of the Oversight Board to oversee the enormous task of fact checking the Facebook platform.

Adaptation

We need to be cautious about describing Facebook within a particular era as it is the organisation's power to adapt and evolve that has helped maintain its dominance. It is therefore important to understand the enduring characteristics that have made Facebook's operations successful in different contexts, and the extent to which it has 'future proofed' itself.

Facebook has maintained market dominance through a range of strategies, including ease-of-use, acting to acquire potential competitors, compatibility across platforms, the continuing addition of new features, accommodating video and working to eliminate anonymity. Different aspects of Facebook are upgraded on a continual basis to respond to new developments and improve features. Facebook also employs a high-quality public relations team to support its public image and respond to critics.

Facebook's valued commodity – big data – is acquired from its users. Facebook's customers, on the other hand, are those companies and organisations willing to pay for what Facebook's data and artificial intelligence gives them access to: namely, a target audience. As Lim and Schumann note, Facebook uses an immune system algorithm to control users' mediated experience as they move towards a desired rhythm (in keeping with Facebook's notion of sociality), while filtering out problematic rhythms. These rhythms are a marketable commodity for Facebook because they are of value to advertisers, who bid for the data so that they can intervene to shape people's experience at the most opportune times (that is, in a manner that accords with these rhythms; see Figure 6). Facebook therefore shapes, manages and filters specific rhythms as a means of ordering sociality to make it more valuable.

Workforce

Facebook has a two-tiered workforce – directly employed staff and a large subcontracted workforce. Facebook employees generally enjoy very good conditions in terms of pay, long vacations, health and dental care, parental leave and a range of perks. Facebook's leadership also engages actively and regularly with its workforce. Zuckerberg and other senior personnel provide weekly question and answer sessions for employees. This includes an update on company goals, including confidential matters. In return, employees are expected to display a high degree of loyalty to the company, and to support and promote its mission. Some current and former employees (speaking anonymously) have, however, been critical of the degree of conformity required and have even likened working for Facebook to being in a cult.

In contrast, the large team of subcontracted Facebook moderators work in casualised, insecure, low paid and generally poor conditions. Facebook's reliance on outsourced, third-party fact checking services, staffed by underpaid, under resourced and under supported human operators is arguably a considerable impediment to its capacity to counter disinformation and malign activity. They are also arguably not sufficiently supported in terms of psychological preparation or counselling, given the potentially traumatising and psychologically dangerous nature of their work. There is evidence that some struggle with symptoms of trauma long after they leave their jobs and what counselling is provided during their employment ends when they depart the company.

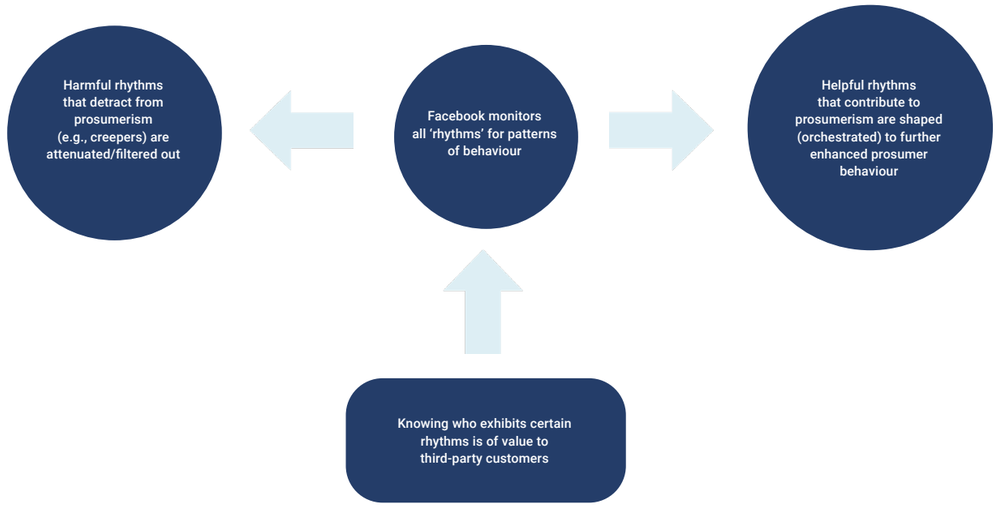


Figure 6: Facebook monitors users' patterns of behaviour or 'rhythms' and seeks to shape these. Access to users (target audiences) that exhibit certain rhythms is of value to third-party organisations.

Liaison with government

Examples of government cooperation with Facebook and other social media entities already exist and could provide the basis for a Facebook liaison facility as part of an enhanced Australian counter-influence capability. This includes the Global Internet Forum to Counter Terrorism (GIFCT), founded by Facebook, Google, Microsoft and Twitter in August 2017. The forum is intended to foster cooperation between companies, advance research and engage with other stakeholders, including governments, to counter the spread of terrorism, and extremist and violent content online. Another is the Christchurch Call to Action, formed by governments after the March 2019, mosque shootings in Christchurch, New Zealand. Facebook and other tech companies have signed onto the initiative's nine-point plan designed to coordinate industry efforts to combat violence and extremist material online. Existing Facebook obligations under the Telco Act may provide the basis for the establishment of ADF social media liaison.

The Facebook Oversight Board has been established to allow an independent body to act as a final arbitrator of content on Facebook. Critics have expressed scepticism about the ability of such a board – eventually, to consist of up to 40 members – to oversee the enormous task of fact checking the Facebook platform. The Board's decision to return a final determination on the banning of former President Trump to Facebook itself has caused critics to further question its effectiveness.

Fact checking is now a key part of Facebook's operations. While problematic in some respects, fact checking may nevertheless be effective if done properly. Research indicates that fact checking may not be effective when undertaken in a manner that appears combative or challenging to users' belief systems. Information is most effective in countering mis- or disinformation when presented in a tactful, respectful manner that avoids disparaging the audience. Research also indicates that Facebook's efforts to limit misinformation after the 2016 US presidential elections appears to have had a meaningful impact. While fact checking by humans alone would not be capable of discerning the amount of malign activity on Facebook and the speed at which it spreads, research indicates that fact checking is likely to be most effective when human and machine techniques combine. Although imperfect, an array of automated fake news detection capabilities (examined in the detailed Facebook Case Study) exist, including user-based, post-based, linguistic-based and network-based techniques.

Digital ethnographic (also known as Netnographic) analysis – a technique for the cultural analysis of social media and online community data – is a potential tool to identify malign online actors and their behaviours and vulnerabilities. Digital ethnographic analysis of the online activities of several different groupings and political and social orientations may identify potential vulnerabilities, and guard against their exploitation by malign actors.

Facebook

Campaign Awareness and Sensemaking

Key Findings – Campaign Awareness and Sensemaking

Strengths

Monitoring and Transparency

- A means by which Facebook monitors the effectiveness of its influence activities is also a means by which it is able to enhance prosumerism (i.e., while Facebook provides engagement metrics to its users, the purpose of this transparency, one might surmise, is to make salient an individual's social presence on the platform and their Facebook popularity).
- Facebook's ability to monitor the effectiveness of its influence activities can also be utilised as a means of promoting the effectiveness of its 'big data' and predictive algorithms – in providing access to target audiences – to its customers (e.g., marketing firms and advertisers).

Managing transparency

- Facebook regularly publishes its Community Standards Enforcement Report. Contained within its pages is information on the number of identified cases of hate speech, bullying and harassment, and updates on how Facebook is helping to manage election integrity and combat mis- and disinformation.
- The means and extent of Facebook's transparency is assessed by the Transparency Advisory Group. Such managed visibilities (see also a weakness below) are a means for Facebook to at least give the appearance of embracing its social responsibilities in order to continue its influence activities.

Weaknesses

- Facebook transparency indicators provide some insight into what Facebook employs to monitor its effectiveness (although much is opaque), but these indicators need to be understood in conjunction with the claim that such transparency does not exist simply to provide insight and clarity but to mediate and manage visibilities.
- Facebook's Community Standards do not always align with countries' laws. This has resulted in Facebook having to conform to local regulations even when content may not violate Facebook's own standards.

Monitoring and transparency

The manner in which Facebook monitors the impact of its influence activities is, in large part, opaque. This means that much of what Facebook is able to do has to be inferred from what it decides to make transparent, not only in term of how it monitors user engagement but also why it makes this information available.

Facebook provides engagement metrics on (inter alia) the amount of time a user spends on its platform, the number of 'likes' given, photos uploaded, events attended, groups joined, photos tagged, links and questions posted, and status and location (or check-in) updates. Making these metrics available tells us that Facebook has the capacity to monitor the effectiveness of its influence activities in this way. But it also allows us to speculate that the reason for the information's transparency is to make salient an individual's social presence on the platform, including their Facebook popularity. We can surmise that Facebook provides engagement metrics to users to encourage them to engage further as prosumers, in accordance with Facebook's business model.

Facebook's ability to monitor the effectiveness of its influence activities can also be utilised to promote the effectiveness of its big data and predictive algorithms to its customers. Information is available to advertisers on Facebook via Facebook insights or analytic tools and includes the following tracking metrics: Engagement (the number of actions – 'likes, shares, comments – taken), Reach (the number of people who have seen the ad), and Referral traffic (the number of visits to the advertiser's website via Facebook). By providing metrics to its customers, Facebook can demonstrate how targeting a particular audience is a cost-effective way to advertise. These data also help to identify areas where the ad is not cost-effective, e.g., if users are watching only the first 30 seconds of a three-minute video.

Managing transparency

Facebook's transparency indicators also need to be understood in conjunction with the claim that transparency is designed to mediate and manage visibilities (i.e., give the appearance that Facebook is embracing its social responsibilities and maintaining integrity). To illustrate: in response to scrutiny over the Cambridge Analytica scandal, Facebook's Page Transparency was created to make available disclosure information such as the date a particular Facebook page was created, the primary country the page is managed in and the number of people who manage it, whether the page belongs to a state-controlled media organisation, and so on. The means and extent of Facebook's transparency is also assessed by the Transparency Advisory Group.

Facebook regularly publishes its Community Standards Enforcement Report (which further supports the claim that the company is embracing its social responsibilities and maintaining integrity). The report provides information on the number of identified cases of hate speech, bullying and harassment, as well as updates on how Facebook is helping to manage election integrity and combat misinformation. Facebook's Community Standards does not always align with other countries' laws, however. For example, Facebook's Community Standards is more tolerant than the German Network Enforcement Act in regulating hate speech. This has resulted in Facebook having to conform to the local regulations in Germany, even where content has not violated its own standards.

Conclusion

In conclusion, Facebook demonstrates the power of social media platforms to substantively enable mass influence campaigns, the effects of which may be compounded by the platform's business model, social license, and the legislative and ethical frameworks within which it operates. Facebook's current business model, based on the collection of user data, is essential to the continued existence of Facebook in its present form and to its profitability. It is, and will likely remain, a very efficient platform for the propagation of misinformation. Any response to potentially malign threats must, therefore, be based on countering these threats on the platform on which they thrive, based on the assumption that the Facebook business model as we know it is likely to continue, and with it the societal risks.

Appendix 1

Below are three timelines delineating Facebook's position on key issues. The first provides examples of Facebook's aggressive acquisitions and suppression strategy, designed to help it monopolise the online social communications market. Recent opposition to this is also included. In addition, Timeline 1 shows Facebook's position on user privacy which can be seen to change after the Cambridge Analytica story broke in 2018. The second timeline maps Facebook's changing attitude towards (allegedly) controversial content: from an initial reluctance to police content to a more (public) realisation and acceptance that social media platforms, including Facebook, need to be more accountable when it comes to enforcing standards. The final timeline illustrates Facebook's shifting position on whether it is a platform or a publisher.

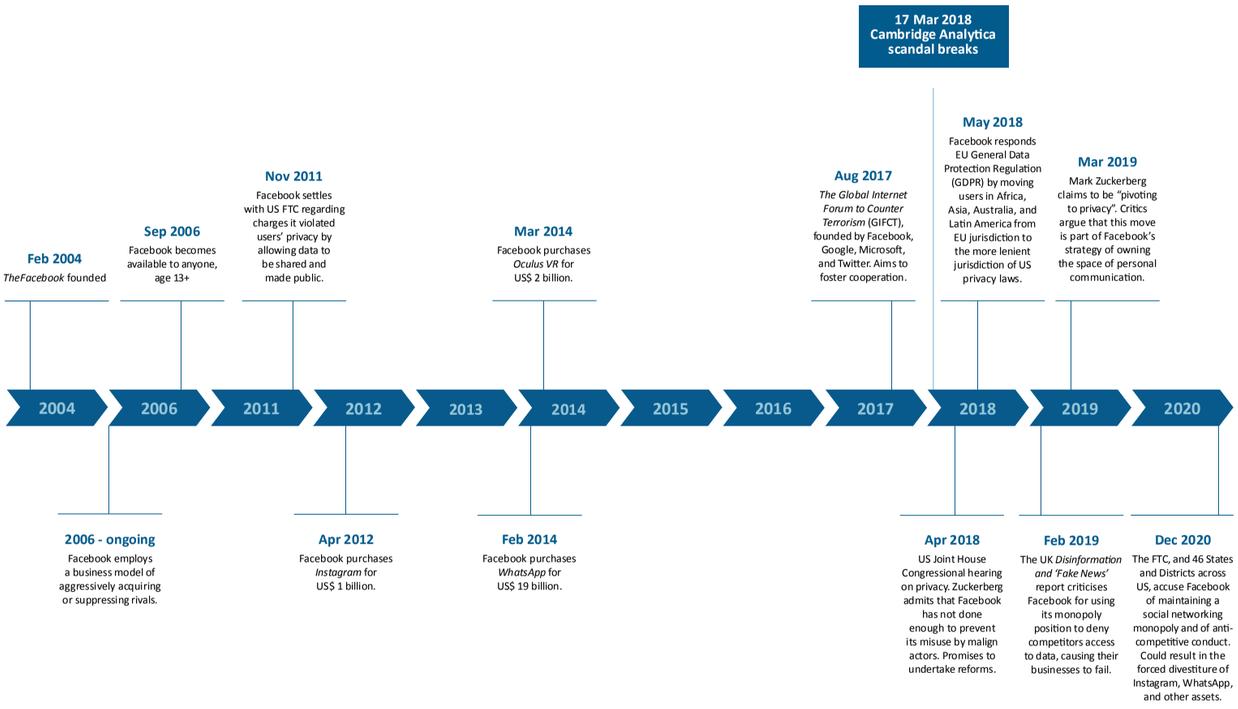


Figure 7: Monopoly and privacy timeline

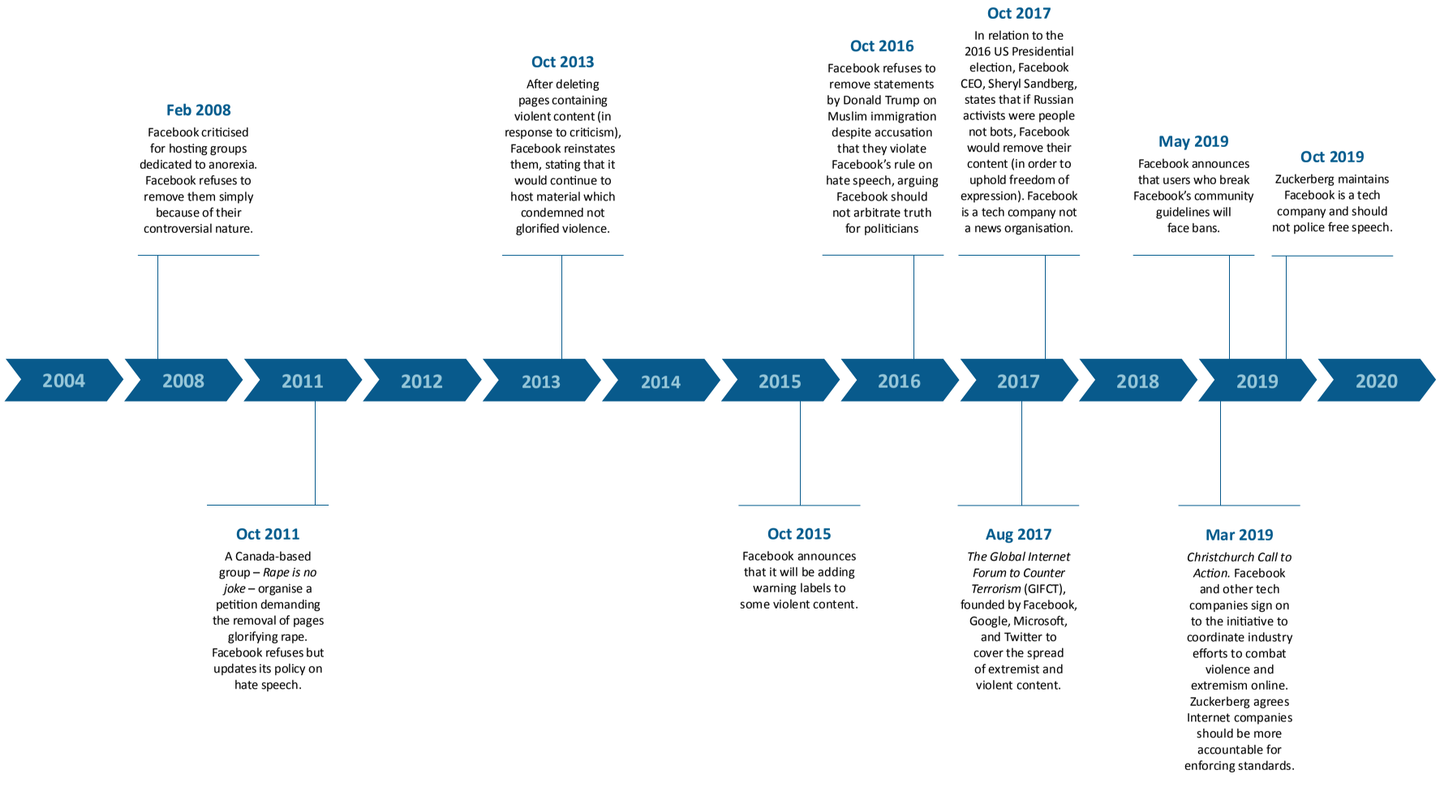
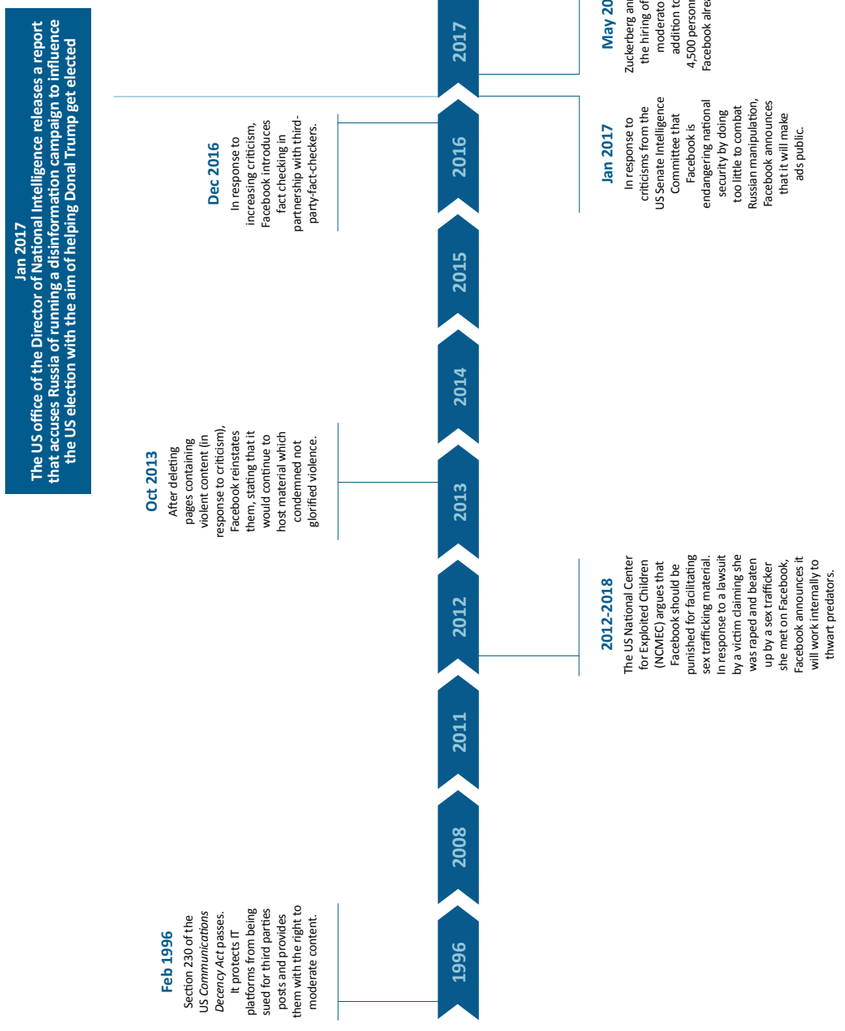


Figure 8: 'Controversial' content timeline



Facebook has often resisted describing itself as a publisher. This is because, in the US, publishers are liable for the content they publish. Section 230 of the 1996 US Communications Decency Act provides Facebook with some protection, however, insofar as it allows a platform to make editorial decisions about content – to act like a publisher – without having to accept liability for all content published on its platform. Legislation has been introduced recently in the US to limit the application of Section 230. It cannot be used to absolve platforms of liability in cases of content related to sex trafficking, for example. There has also been talk of the current White House administration seeking to limit further, if not altogether abolish, Section 230.

Figure 10: 'Facebook as 'platform or publisher' and mis-/disinformation timeline

Glossary

ACCC	Australian Competition and Consumer Commission
ACMA	Australian Communications and Media Authority
AI	Artificial intelligence
Algorithmic drift	Phenomenon in which a site's algorithmic suggestions funnel a user into viewing more radical material, causing them to drift towards an online environment containing more extreme messages
AR	Augmented Reality. The placement of computer-generated imagery in a user's field of view.
ARPU	Average Revenue Per User (Facebook)
BOSN	Brand Online Social Networking. A means of marketing that enables companies to initiate and cultivate relationships with their customers through social media platforms such as Facebook.
BSA	Broadcasting Services Act, 1992 (Australia)
CDA	Section 230 (c) (1) of the 1996 US Communications Decency Act (CDA). The act protects IT platforms from being sued for third parties posts and provides them the right to moderate content and users.
CEO	Chief Executive Officer
Christchurch Call to Action	Organisation formed by governments after the March 2019 mosque shootings in Christchurch, New Zealand to combat violence and extremist material online
ClaimBuster	A platform that uses machine learning to fact check claims in political discussions
ClaimVerif	A real-time claim verification system
CNOIR	Counter Narratives to Interrupt Online Radicalisation. A project aimed at exploring ways in which to counter online radicalisation
Community Standards	The standards Facebook users are required to adhere to when posting material on Facebook.
CSI	Capture, Score and Integrate. A model composed of Capture, Score and Integrate using Recurrent Neural Network (RNN).
Culture sharing	The exchange or mutual exposure of preferred lifestyles via social ties between users from different cultural backgrounds
DAP	Family Daily Active People (Facebook)
DAU	Daily Active User (Facebook)
DeepFace	A machine learning facial recognition tool claiming a higher accuracy rate than human recognition
DeBot	A system to identify bot accounts on social media
DJINET	Dow Jones Internet Composite Index
DSTS	Dynamic Series-Time Structure. A model to capture the variation of a wide spectrum of social context information over time
DTW	Dynamic Time Warping. An algorithm for measuring similarities between two temporal sequences which may vary in speed.
ELM	Elaboration Likelihood Model. A dual process model that posits central and peripheral routes to persuasion.
Engagement	The number of actions ('likes', shares, comments) to a Facebook advertisement
EU	European Union
Facebook connect	A facility that allows users to use Facebook across applications
Facebook Oversight Board	An independent body of qualified individuals that set standards governing the distribution of harmful content and act as a final arbitrator of content on Facebook
Facemash	A website established by Mark Zuckerberg at Harvard University in 2003 and closed down by Harvard management for non-consensually posting photos and inappropriate content on female Harvard students
FAN	Facebook Audience Network
FCA	Fact Checking and Analysis
FOSTA	Fight Online Sex Trafficking Act, 2018 (US). Along with the SESTA act (below), the act removes the immunity granted under Section 230 of the Communications Decency Act (1996) when dealing with sex trafficking
Freedom Watch	A conservative US organisation that monitors the media and advocates for a position in relation to it
FTC	Federal Trade Commission (US)

GDPR	General Data Protection Regulation (EU). A regulation which enforces a privacy regime on companies operating in the EU, including a right to obtain personal data, to be forgotten, to data portability and a requirement for affirmative consent to use data.
GIFCT	Global Internet Forum to Counter Terrorism. Founded by Google, Microsoft, and Twitter to counter the spread of terrorism and extremist and violent content online.
Grey zone	One of a range of terms used to describe activities, facilitated by technological developments including cyber warfare, designed to coerce countries in ways that seek to avoid military conflict. Examples include using para-military forces, militarisation of disputed features, exploiting influence, interference operations and the coercive use of trade and economic levers.
HITS	Hyperlink-Induced Topic Search. A link analysis algorithm that rates Web pages.
ICO	Information Commissioner's Office (UK)
IRA	Internet Research Agency (Russia)
Impression	The number of times an advertisement is seen and/or acted on by the same person
Information content provider	A person or entity that is responsible for the creation or development of information provided through the internet
Interactive computer service	An information service system or access software provider (such as Facebook) that enables computer access by multiple users to a computer server
MAU	Monthly Active User (Facebook)
MIP	Mass Interpersonal Persuasion. The ability to persuade people of a position on a large scale.
NBC	Naïve Bayes Classifier. A family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naïve) independence assumptions between the features.
NCMEC	The National Center for Missing and Exploited Children (US)
Netnography	A technique for the cultural analysis of social media and online community data
PageRank	An algorithm used by Google Search to rank web pages in their search engine results
Perseverance effect	The phenomenon in which individuals continue to believe fake news or disinformation they are initially exposed to after it has been corrected
Prosumer	An individual who acts as both a consumer and producer of information, usually on a social network
Reach	The number of people who see a Facebook advertisement
Referral traffic	The number of visits to an advertiser's website via Facebook
RFC	Random Forest Classifier. A meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.
RFC	Related Fact Checks. An analysis assistance application for fact checking.
RNN	Recurrent Neural Network. A class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence.
SEC	Securities and Exchange Commission (US)
Section 230	See CDA
SESTA	Stop Enabling Sex Traffickers Act, 2018 (US). Along with the FOSTA act (above), the act removes the immunity granted under Section 230 of the Communications Decency Act (1996) when dealing with sex trafficking.
SGD	An iterative method for optimising an objective function with suitable smoothness properties
Shadow profile	Information about an individual that a social network has obtained indirectly by accessing another user's account
SVC	Support Vector Classifier. Supervised learning models with associated learning algorithms that analyse data for classification and regression analysis.
SVM	Support Vector Machine. See SVC.
TAM	Technology Acceptance Model. An information systems theory that models how users come to accept and use a technology.
WHO	World Health Organisation
WOT	Web of Trust. A service that calculates the reputations of websites and to provide credibility assessment for queries given by users.
WT.Social	A social media platform (launched October 2019) designed for the sharing of information in a similar way to Facebook, but funded through donations rather than data collection

Authors and Contributors

Internet Research Agency

Authors
The University of Melbourne
Emily Ebbott
Dr Morgan Saletta
Richard Stearne

Contributors
Defence Science and Technology Group
Ms Mirela Stjelja

Edith Cowan University
Dr Andrew Dowse
Dr Violetta Wilk

The University of Adelaide
Associate Professor Tim Legrand
Professor Melissa de Zwart
Professor Dale Stephens
Professor Debi Ashenden
Professor Michael Webb

The University of Melbourne
Professor Christopher Leckie
Associate Professor Atif Ahmad
Associate Professor Andrew Perfors
Associate Professor Richard de Rozario
Associate Professor Tim van Gelder
Professor Len Sciacca
Dr Jey Han Lau
Professor Yoshihisa Kashima
Associate Prof Leah Ruppanner
Professor Shanika Karunasekera

University of New South Wales
Professor Monica Whitty

Cambridge Analytica

Authors
The University of Adelaide
Professor Michael Webb
Dr Melissa-Ellen Dowling
Dr Matteo Farina

Edith Cowan University
Dr Stephanie Meek
Dr Carmen Jacques

Contributors
Defence Science and Technology Group
Dr David Matthews

The University of Adelaide
Associate Professor Carolyn Semmler
Associate Professor Tim Legrand
Professor Melissa de Zwart
Professor Dale Stephens
Professor Debi Ashenden
Associate Professor Lewis Mitchell
Associate Professor Martin White

Facebook

Authors
University of New South Wales
Dr Garry Young
Dr Peter Job

Contributors
Defence Science and Technology Group
Ms Laura Carter

Edith Cowan University
Dr Andrew Dowse
Dr Violetta Wilk
Mr Conor McLaughlin

Macquarie University
Professor Ben Schreer
Dr Brian Ballsun-Stanton
Dr Julian Droogan
Dr Lise Waldek
Ms Jade Hutchinson

The University of Melbourne
Professor Yoshihisa Kashima
Dr Jey Han Lau

University of New South Wales
Professor Monica Whitty
Associate Professor Stephen Doherty
Associate Professor Douglas Guilfoyle
Associate Professor Josh Keller
Associate Professor Rob Nicholls
Associate Professor Salih Ozdemir
Dr Raymond Wong

References

Internet Research Agency

1. Ben Nimmo, and Aric Toler. “How They Did It: The Real Russian Journalists Who Exposed the Troll Factory in St. Petersburg.” Global Investigative Journalism Network, March 26, 2018. https://gijn.org/2018/03/26/real-russian-journalists-exposed-troll-factory-st-petersburg/.
2. Maria Snegovaya and Kohei Watanabe, “The Kremlin’s Social Media Influence inside the United States: A Moving Target,” February 10, 2021.
3. Maria Snegovaya and Kohei Watanabe, “The Kremlin’s Social Media Influence inside the United States: A Moving Target,” February 10, 2021.
4. Maria Snegovaya and Kohei Watanabe, “The Kremlin’s Social Media Influence inside the United States: A Moving Target,” February 10, 2021.
5. Maria Snegovaya and Kohei Watanabe, “The Kremlin’s Social Media Influence inside the United States: A Moving Target,” February 10, 2021.
6. “U.S. v. Internet Research Agency LLC, et al (U.S. District Court for the District of Columbia; 1:18-Cr-32),” February 16, 2018
7. Prigozhin also funds and controls various media outlets that form part of Russia’s propaganda ecosystem. He also helps fund and operate the Wagner Group, whose mercenaries operate as proxies for the Russian government in the physical, kinetic environment and has close ties to the Intelligence Community.
8. Intelligence Community Assessment: Assessing Russian Activities and Intentions in Recent US Elections” (Office of the Director of National Intelligence, 2017), https://www.dni.gov/files/documents/ICA_2017_01.pdf.
9. Maria Snegovaya and Kohei Watanabe, “The Kremlin’s Social Media Influence inside the United States: A Moving Target,” February 10, 2021.
10. “U.S. v. Internet Research Agency LLC, et al (U.S. District Court for the District of Columbia; 1:18-Cr-32),” February 16, 2018
11. Renee DiResta, Kris Shaffer, Becky Ruppel, David Sullivan, Robert Matney, et al., “The Tactics & Tropes of the Internet Research Agency,” 2019; Philip N Howard et al., “The IRA, Social Media and Political Polarization in the United States, 2012-2018,” 2019.
12. U.S. v. Internet Research Agency LLC, et al (U.S. District Court for the District of Columbia; 1:18-Cr-32),” February 16, 2018
13. David Holt, “US District Court Criminal Complaint against Elena Khusyaynova”; Aric Toler, “Inside the Kremlin Troll Army Machine: Templates, Guidelines, and Paid Posts,” Global Voices (blog), March 14, 2015, https://globalvoices.org/2015/03/14/russia-kremlin-troll-army-examples/.
14. U.S. v. Internet Research Agency LLC, et al; Nathaniel Reynolds, “Putin’s Not-So-Secret Mercenaries: Patronage, Geopolitics, and the Wagner Group,” Carnegie Endowment for International Peace, 2019, https://carnegieendowment.org/2019/07/08/putin-s-not-so-secret-mercenaries-patronage-geopolitics-and-wagner-group-pub-79442.
15. Maria Snegovaya and Kohei Watanabe, “The Kremlin’s Social Media Influence inside the United States: A Moving Target,” February 10, 2021.
16. U.S. v. Internet Research Agency LLC, et al (U.S. District Court for the District of Columbia; 1:18-Cr-32),” February 16, 2018
17. “Report of the Senate Select Committee on Intelligence United States Senate on Russian Active Measures Campaigns and Interference. Volume 2: Russia’s Use of Social Media with Additional Views”; DiResta et al., “The Tactics & Tropes of the Internet Research Agency.”
18. Nathaniel Gleicher and David Agranovich. “Removing Coordinated Inauthentic Behavior from France and Russia.” About Facebook (blog), 2020. https://about.fb.com/news/2020/12/removing-coordinated-inauthentic-behavior-france-russia/.
19. Report of the Senate Select Committee on Intelligence United States Senate on Russian Active Measures Campaigns and Interference. Volume 2: Russia’s Use of Social Media with Additional Views”; DiResta et al., “The Tactics & Tropes of the Internet Research Agency.”
20. Maria Snegovaya and Kohei Watanabe, “The Kremlin’s Social Media Influence inside the United States: A Moving Target,” February 10, 2021.
21. Christopher Paul and Miriam Matthews, The Russian “Firehose of Falsehood” Propaganda Model: Why It Might Work and Options to Counter It (RAND Corporation, 2016), https://doi.org/10.7249/PE198.
22. Maria Snegovaya and Kohei Watanabe, “The Kremlin’s Social Media Influence inside the United States: A Moving Target,” February 10, 2021.
23. U.S. v. Internet Research Agency LLC, et al (U.S. District Court for the District of Columbia; 1:18-Cr-32),” February 16, 2018
24. Stephan Lewandowsky et al., “Misinformation and Its Correction: Continued Influence and Successful Debiasing,” Psychological Science in the Public Interest, September 17, 2012, https://doi.org/10.1177/1529100612451018.
25. U.S. v. Internet Research Agency LLC, et al (U.S. District Court for the District of Columbia; 1:18-Cr-32),” February 16, 2018
26. Report of the Senate Select Committee on Intelligence United States Senate on Russian Active Measures Campaigns and Interference. Volume 2: Russia’s Use of Social Media with Additional Views”; DiResta et al., “The Tactics & Tropes of the Internet Research Agency.”
27. “How Is Fake News Spread? Bots, People like You, Trolls, and Microtargeting | Center for Information Technology and Society - UC Santa Barbara.” Accessed February 18, 2021. https://www.cits.ucsb.edu/fake-news/spread.
28. Dennis Assenmacher et al., “Demystifying Social Bots: On the Intelligence of Automated Social Media Actors,” Social Media + Society 6, no. 3 (July 2020): 205630512093926, https://doi.org/10.1177/2056305120939264.
29. Dennis Assenmacher et al., “Demystifying Social Bots: On the Intelligence of Automated Social Media Actors,” Social Media + Society 6, no. 3 (July 2020): 205630512093926, https://doi.org/10.1177/2056305120939264.
30. Dennis Assenmacher et al., “Demystifying Social Bots: On the Intelligence of Automated Social Media Actors,” Social Media + Society 6, no. 3 (July 2020): 205630512093926, https://doi.org/10.1177/2056305120939264.
31. U.S. v. Internet Research Agency LLC, et al (U.S. District Court for the District of Columbia; 1:18-Cr-32),” February 16, 2018
32. U.S. v. Internet Research Agency LLC, et al (U.S. District Court for the District of Columbia; 1:18-Cr-32),” February 16, 2018
33. U.S. v. Internet Research Agency LLC, et al (U.S. District Court for the District of Columbia; 1:18-Cr-32),” February 16, 2018
34. U.S. v. Internet Research Agency LLC, et al (U.S. District Court for the District of Columbia; 1:18-Cr-32),” February 16, 2018
35. U.S. v. Internet Research Agency LLC, et al; Ivan Nechepurenko and Michael Schwirtz, “What We Know About Russians Sanctioned by the United States,” The New York Times, February 17, 2018, sec. World, https://www.nytimes.com/2018/02/17/world/europe/russians-indicted-mueller.html.

Cambridge Analytica

Adler-Nissen, Rebecca, Katrine Emilie Andersen, and Lene Hansen. “Images, Emotions, and International Politics: The Death of Alan Kurdi.” *Review of International Studies* 46, no. 1 (2020): 75-95.

Ahmad, Nadeem, and Jawaid Siddique. “Personality Assessment Using Twitter Tweets.” *Procedia Computer Science* 112 (2017): 1964-73.

Allen, Jonathan, and Jason Abbruzzese. “Cambridge Analytica’s Effectiveness Called into Question Despite Alleged Facebook Data Harvesting.” NBC News, https://www.nbcnews.com/politics/politics-news/cambridge-analytica-s-effectiveness-called-question-despite-alleged-facebook-data-n858256.

Analytica, Cambridge. “Ca Political an Overview of Cambridge Analytica’s Political Division.” 2015.

Anderson, Helen. “Piercing the Veil on Corporate Groups in Australia: The Case for Reform.” Melb. UL Rev. 33 (2009): 333.

Auletta, Ken. *Frenemies: The Epic Disruption of the Advertising Industry (and Why This Matters)*. London: Harper Collins, 2018.

Australian Information Commissioner v Facebook Inc (No 2), No. 1307 (FCA September 14, 2020).

Back, Mitja D., Juliane M. Stopfer, Simine Vazire, Sam Gaddis, Stefan C. Schmukle, Boris Egloff, and Samuel D. Gosling. “Facebook Profiles Reflect Actual Personality, Not Self-Idealization.” *Psychological Science* 21, no. 3 (2010): 372-74.

Bai, Shuotian, Tingshao Zhu, and Li Cheng. “Big-Five Personality Prediction Based on User Behaviors at Social Network Sites.” *arXiv preprint* 1204, no. 4809 (2012).

Bailey, Michael A., Daniel J. Hopkins, and Todd Rogers. “Unresponsive and Unpersuaded: The Unintended Consequences of a Voter Persuasion Effort.” *Political behavior* 38, no. 3 (2016): 713-46.

Bakir, Vian. “Psychological Operations in Digital Political Campaigns: Assessing Cambridge Analytica’s Psychographic Profiling and Targeting.” *Frontiers in Communication* 5 (2020): 67.

Baldassarri, Delia, and Andrew Gelman. “Partisans without Constraint: Political Polarization and Trends in American Public Opinion.” *American Journal of Sociology* 114, no. 2 (2008): 408–46.

Bartlett, Jamie. *The People Vs Tech: How the Internet Is Killing Democracy (and How We Save It)*. London: Ebury Press, 2018.

BBC. “Cambridge Analytica Parent Firm SCL Elections Fined Over Data Refusal.” BBC, January 10, 2019. https://www.bbc.com/news/technology-46822439.

Berghel, Hal. “Malice Domestic: The Cambridge Analytica Dystopia.” *Computer* 51, no. 5 (2018): 84–89.

Bevir, Mark. *Democratic Governance*. Princeton University Press, 2010.

Birch, A.H. *The Concepts and Theories of Modern Democracy*. London: Routledge, 2001.

Blumenthal-Barby, J. “Between Reason and Coercion: Ethically Permissible Influence in Health Care and Health Policy Contexts.” *Kennedy Institute of Ethics journal* 22, no. 4 (2012): 345–66.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. “A 61—Million—Person Experiment in Social Influence and Political Mobilization.” *Nature (London)* 489, no. 7415 (2012): 295-98.

Briant, Emma L. “Cambridge Analytica and Scl – How I Peered inside the Propaganda Machine.” *The Conversation* (2018).

Briant, Emma, and Kaiser, Brittany. “Propaganda Machine: The Military Roots of Cambridge Analytica’s Psychological Manipulation of Voters,” January 2020. https://www.democracynow.org/2020/1/7/cambridge_analytica_data_manipulation_john_bolton.

Briant, Emma L. “We Need Tougher Action against Disinformation and Propaganda.” *Brookings* (blog), July 15, 2020.
https://www.brookings.edu/blog/techtank/2020/07/15/we-need-tougher-action-against-disinformation-and-propaganda/.

Broockman, David E., and Donald P. Green. “Do Online Advertisements Increase Political Candidates’ Name Recognition or Favorability? Evidence from Randomized Field Experiments.” *Political Behavior* 36 (2014): 2.

Cadwalladr, Carole. “Cambridge Analytica’s Ruthless Bid to Sway the Vote in Nigeria.” *The Guardian*, 2018.

References

———. “Revealed: Graphic Video Used by Cambridge Analytica to Influence Nigerian Election.” *The Guardian*, 2018.

Cadwalladr, Carole, Emma Graham-Harrison. “Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach.” *The Guardian*, 2018.

Cadwalladr, Carole, and Emma Graham-Harrison. “How Cambridge Analytica Turned Facebook ‘Likes’ into a Lucrative Political Tool.” *The Guardian*, 2018.

Cadwalladr, Carole, and Emma Graham-Harrison. “Cambridge Analytica: Links to Moscow Oil Firm and St Petersburg University.” the Guardian, March 17, 2018. http://www.theguardian.com/news/2018/mar/17/cambridge-academic-trawling-facebook-had-links-to-russian-university.

Cambridge Analytica. “CA Intelligence Report,” 2015. https://twitter.com/hindsightfiles?lang=en.

Cambridge Analytica,. “CA Political,” 2016. https://twitter.com/hindsightfiles?lang=en.

Cambridge Analytica. “SCL Elections: John Bolton Super PAC,” n.d. https://twitter.com/hindsightfiles?lang=en.

“Cambridge Analytica and Facebook: The Scandal and the Fallout So Far - The New York Times.” Accessed April 17, 2021. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.

Wayback Machine. “Cambridge Analytica: Better Audience Targeting,” 2016. https://web.archive.org/web/20210320164149/https://cambridgeanalytica.org/.

“‘Cambridge Analytica Planted Fake News’ - BBC News.” Accessed April 17, 2021. https://www.bbc.com/news/av/world-43472347.

Calo, Ryan. “Digital Market Manipulation.” *George Washington Law Review* 82, no. 4 (2013): 995- 1051.

Casadesus-Masanell, Ramon, and Joan Ricart. “How to Design a Winning Business Model.” *Harvard Business Review* January-February (2011).

Chambers, Simone. “Truth, Deliberative Democracy, and the Virtues of Accuracy: Is Fake News Destroying the Public Sphere?” *Political Studies* 69, no. 1 (2020): 147–63.

Chen, Adrien. “Cambridge Analytica and Our Lives inside the Surveillance Machine.” *The New Yorker* 21 (2018): 8–10.

Chen, Angela, and Alessandra Potenza. “Cambridge Analytica’s Facebook Data Abuse Shouldn’t Get Credit for Trump.” The Verge, https://www.theverge.com/2018/3/20/17138854/cambridge-analytica-facebook-data-trump-campaign-psychographic-microtargeting.

Chester, Jeff, and Kathryn C Montgomery. “The Role of Digital Marketing in Political Campaigns.” *Internet Policy Review 6*, no. 4 (2017): 1–20.

Chigona, Wallace, Darry Beukes, Junaid Vally, and Maureen Tanner. “Can Mobile Internet Help Alleviate Social Exclusion in Developing Countries?”. *The Electronic Journal of Information Systems in Developing Countries* 36, no. 1 (2009): 116.

Chou, Hui-Tzu Grace, and Nicholas Edge. ““They Are Happier and Having Better Lives Than I Am”: The Impact of Using Facebook on Perceptions of Others’ Lives.” *Cyberpsychology, Behavior and Social Networking* 15, no. 2 (2012): 117–21.

Clement, J. “Number of Social Network Users Worldwide from 2017 to 2025.” https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/.

Dahl, Robert A. *On Democracy*. Yale university press, 2008.

Dahl, Robert A. “What Political Institutions Does Large-Scale Democracy Require?” *Political Science Quarterly* 120, no. 2 (2005): 187–97. http://www.jstor.org/stable/20202514.

DCMS. “Disinformation and Fake News: Final Report.” Vol. HC1791. London, 2019. https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf.

———. “Disinformation and ‘Fake News’: Interim Report.” London: House of Commons, July 24, 2018. https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/363.pdf.

———. Oral Evidence: fake news Alexander Nix HC 363, Pub. L. No. HC 363, § Digital, Culture, Media and Sport Committee (2018). http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/disinformation-and-fake-news/oral/79388.pdf.

———. Oral Evidence: fake news Alexander Nix HC 363, Pub. L. No. HC 363, § Digital, Culture, Media and Sport Committee (2018). http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/disinformation-and-fake-news/oral/84838.pdf.

———. Oral evidence: fake news Brittany Kaiser HC363, § Digital, Culture, Media and Sport Committee (2018). http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/disinformation-and-fake-news/oral/81592.pdf.

———. Oral Evidence: fake news Christopher Wylie, § Digital, Culture, Media and Sport Committee (2018). http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/fake-news/oral/81022.html.

De Zwart, Melissa. “Contractual Communities: Effective Governance of Virtual Worlds.” *UNSWLJ* 33 (2010): 605.

Dine, J., and M. Koutsias. *The Nature of Corporate Governance*. Corporations, Globalisation and the Law Series. Edward Elgar Publishing, 2013.

Dowling, Melissa-Ellen. “Democracy under Siege: Foreign Interference in a Digital Era.” *Australian Journal of International Affairs*, 2021, 1–5.

Dowling, Melissa-Ellen, and Tim Legrand. “Countering Foreign Interference Series.” Canberra: Department of Defence, 2021.

Ekdale, Brian, and Melissa Tully. “African Elections as a Testing Ground: Comparing Coverage of Cambridge Analytica in Nigerian and Kenyan Newspapers.” *African Journalism Studies* 40, no. 4 (2019): 27–43.

Endsley, Mica R. “Toward a Theory of Situation Awareness in Dynamic Systems.” *Human Factors* 37, no. 1 (1995): 32–64.

Endsley, Mica R., and Daniel J. Garland. *Situation Awareness Analysis and Measurement*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2000.

Faden, Ruth R., Tom L. Beauchamp, and Nancy M. P. King. *A History and Theory of Informed Consent*. New York: Oxford University Press, 1986.

Farina, Matteo. *Facebook and Conversation Analysis*. London: Bloomsbury, 2018.

———. “Facebook First Post Telling.” *Journal of pragmatics* 90 (2015): 1-11.

Finlay, Lorraine, and Christian Payne. “The Attribution Problem and Cyber Armed Attacks.” *AJIL Unbound* 113 (2019): 202–6.

Fogg, B. J. *Persuasive Technology Using Computers to Change What We Think and Do*. The Morgan Kaufmann Series in Interactive Technologies. Amsterdam: Morgan Kaufmann Publishers, 2003.

Foley, Paul. “Does the Internet Help to Overcome Social Exclusion.” *Electronic Journal of e-government* 2, no. 2 (2004): 139-46.

Gelman, Andrew, Jessica Hullman, Christopher Wlezien, and George E. Morris. “Information, Incentives, and Goals in Election Forecasts.” *Judgment and Decision Making* 15, no. 5 (2020): 863-80.

Gorton, William A. “Manipulating Citizens: How Political Campaigns’ Use of Behavioral Social Science Harms Democracy.” *New Political Science* 38, no. 1 (2016): 61–80.

Gosling, Samuel D., Adam A. Augustine, Simine Vazire, Nicholas Holtzman, and Sam Gaddis. “Manifestations of Personality in Online Social Networks: Self-Reported Facebook-Related Behaviors and Observable Profile Information.” *Cyberpsychology, Behavior, and Social Networking* 14, no. 9 (2011): 483-88.

Green v SCL Group Ltd and others [2019] EWHC 954 (Ch) (April 17, 2019).

Guess, Andrew M., Brendan Nyhan, and Jason Reifler. “Exposure to Untrustworthy Websites in the 2016 Us Election.” *Nature human behaviour* 4, no. 5 (2020): 472-80.

Gupta, Yayati, Akrati Saxena, Debarati Das, and Iyengar. S. R. S. “Modeling Memetics Using Edge Diversity.” *Complex Networks VII Springer*, Cham (2016): 187-98.

Gurven, Michael, Christopher von Rueden, Maxim Massenkoff, Hillard Kaplan, and Marino Lero Vie. “How Universal Is the Big Five? Testing the Five-Factor Model of Personality Variation among Forager-Farmers in the Bolivian Amazon.” *Journal of Personality and Social Psychology* 104, no. 2 (2013): 354-70.

Hausman, Daniel M, and Brynn Welch. “Debate: To Nudge or Not to Nudge.” *Journal of Political Philosophy* 18, no. 1 (2010): 123–36.

Heawood, Jonathan. “Pseudo-Public Political Speech: Democratic Implications of the Cambridge Analytica Scandal.” *Information Polity* 23, no. 4 (2018): 429–34.

Henriksen, Ellen Emilie. “Big Data, Microtargeting, and Governmentality in Cyber-Times. The Case of the Facebook-Cambridge Analytica Data Scandal.” 2019.

Hindman, Matthew. “How Cambridge Analytica’s Facebook Targeting Model Really Worked – According to the Person Who Built It.” *The Conversation*, 2018.

Hinds, Joanne, Emma J Williams, and Adam N Joinson. “‘It Wouldn’t Happen to Me’: Privacy Concerns and Perspectives Following the Cambridge Analytica Scandal.” *International Journal of Human-Computer Studies* 143 (2020): 102498.

Hirsh, Jacob B., Colin G. DeYoung, Xu Xiaowen, and Jordan B. Peterson. “Compassionate Liberals and Polite Conservatives: Associations of Agreeableness with Political Ideology and Moral Values.” *Personality and Social Psychology Bulletin* 36, no. 5 (2010): 655–64.

Hirsh, Jacob B., Sonia K. Kang, and Galen V. Bodenhausen. “Personalized Persuasion: Tailoring Persuasive Appeals to Recipients’ Personality Traits.” *Psychological science* 23, no. 6 (2012): 578-81.

Hu, Margaret. “Cambridge Analytica’s Black Box.” *Big Data & Society* 7, no. 2 (2020): 1–6.

ICO. “SCL Elections Prosecuted for Failing to Comply with Enforcement Notice,” January 9, 2019. https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/01/scl-elections-prosecuted-for-failing-to-comply-with-enforcement-notice/.

In the Matter of Cambridge Analytica, No. 9383 (Federal Trade Commission USA 2019).

In the Matter of Cambridge Analytica LLC, No. 9383 (Federal Trade Commission December 18, 2019).

Information Commissioner’s Office. “Investigation into the Use of Data Analytics in Political Campaigns.” UK, 2018. https://ico.org.uk/media/action-weve-taken/2259371/investigation-into-data-analytics-for-political-purposes-update.pdf.

Jacobs, Naomi. “Two Ethical Concerns About the Use of Persuasive Technology for Vulnerable People.” *Bioethics* 34, no. 5 (2020): 519-26.

Kaiser, B. Targeted: *The Cambridge Analytica Whistleblower’s Inside Story of How Big Data, Trump, and Facebook Broke Democracy and How It Can Happen Again*. Harper, 2019.

Karlsson-Vinkhuyzen. “Legitimacy.” *In Handbook on Theories of Goverannce*, edited by Christopher Ansell and Jacob Torfing, 2016.

Kefford, G. *Political Parties and Campaigning in Australia: Data, Digital and Field*. Political Campaigning and Communication. Springer International Publishing AG, 2021.

References

Kosinski, Michal, Yoram Bachrach, Pushmeet Kohli, David Stillwell, and Thore Graepel. “Manifestations of User Personality in Website Choice and Behaviour on Online Social Networks.” *Machine learning* 95, no. 3 (2013): 357-80.

Kosinski, Michal, Sandra C. Matz, Samuel D. Gosling, Vesselin Popov, and David Stillwell. “Facebook as a Research Tool for the Social Sciences.” *American Psychologist* 70, no. 6 (2015): 543-56.

Kosinski, Michal, David Stillwell, and Thore Graepel. “Private Traits and Attributes Are Predictable from Digital Records of Human Behavior.” *Proceedings of the national academy of sciences* 110, no. 15 (2013): 5802-05.

Krogerus, Mikael, and Hannes Grassegger. “The Data That Turned the World Upside Down.” Motherboard Tech by Vice, https://www.vice.com/en/contributor/hannes-grassegger-and-mikael-krogerus.

Lau, Richard R, and Ivy Brown Rovner. “Negative Campaigning.” *Annual Review of Political Science* 12 (2009): 285–306.

Lampe, Cliff, Nicole B. Ellison, and Charles Steinfield. “A Face (Book) in the Crowd: Social Searching Vs. Social Browsing.” Paper presented at the Computer supported cooperative work, New York, NY, USA, 2006.

Lee-Won, Roselyn J., Minsun Shim, Yeon Kyoung Joo, and Sung Gwan Park. “Who Puts the Best “Face” Forward on Facebook?: Positive Self-Presentation in Online Social Networking and the Role of Self-Consciousness, Actual-to-Total Friends Ratio, and Culture.” *Computers in Human Behavior* 39 (2012): 413–23.

Lomas, Natasha. “Cambridge Analytica’s Nix Said It Licensed ‘millions of Data Points’ from Acxiom, Experian, Infogroup to Target US Voters.” *Techcrunch*, June 7, 2018. https://techcrunch.com/2018/06/06/cambridge-analyticas-nix-said-it-licensed-millions-of-data-points-from-axciom-experian-infogroup-to-target-us-voters/.

Mackay, A., S. Tatham, and L. Rowland. *Behavioural Conflict: Why Understanding People and Their Motivations Will Prove Decisive in Future Conflict*. Military Studies, 2011.

Maddox, G. *Australian Democracy in Theory and Practice*. Sydney: Pearson Education Australia, 2005.

Manokha, Ivan. “Surveillance: The DNA of Platform Capital—The Case of Cambridge Analytica Put into Perspective.” *Theory & Event* 21, no. 4 (2018): 891–913.

Mares, Radu. “Liability within Corporate Groups: Parent Company’s Accountability for Subsidiary Human Rights Abuses.” In *Research Handbook on Human Rights and Business*, 446–71. Cheltenham: Edward Elgar Publishing, 2020.

Marlowe, Ann, and Wendy Siegelman. “SCL Group Companies and Shareholders,” May 9, 2017. https://wsiegelman.medium.com/scl-companies-shareholders-e65a4f394158.

Marriott, Tamsin C., and Tom Buchanan. “The True Self Online: Personality Correlates of Preference for Self-Expression Online, and Observer Ratings of Personality Online and Offline.” Computers in Human Behavior 32 (2014): 171-77.

Matz, Sandra C., Michal Kosinski, Gideon Nave, and David J. Stillwell. “Psychological Targeting as an Effective Approach to Digital Mass Persuasion.” *Proceedings of the national academy of sciences* 114, no. 48 (2017): 12714-19.

McCrae, Robert R., and Paul T. Costa. “Personality Trait Structure as a Human Universal.” *The American Psychologist* 52, no. 5 (1997): 509-16.

———. “Toward a New Generation of Personality Theories: Theoretical Contexts for the Five-Factor Model.” In *The Five-Factor Model of Personality: Theoretical Perspectives*, edited by Jerry S. Wiggins. New York: Guilford Press, 1996.

McCrae, Robert R., and John P. Oliver. “An Introduction to the Five-Factor Model and Its Applications.” *Journal of Personality* 60, no. 2 (1992): 175-215.

Narayan, Deepa. “Bonds and Bridges: Social Capital and Poverty.” World Bank, 1999. https://documents1.worldbank.org/curated/en/989601468766526606/107507322_20041117172515/additional/multi-page.pdf.

Nickerson, David W., and Todd Rogers. “Political Campaigns and Big Data.” *The Journal of economic perspectives* 28, no. 2 (2014): 51-73.

Office, Information Commissioner’s. “Investigation into the Use of Data Analytics in Political Campaigns.” 2018.

O’Neil, Cathy. Weapons of *Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Allen Lane an imprint of Penguin Books, 2016.

Ott, Brian L. “The Age of Twitter: Donald J. Trump and the Politics of Debasement.” *Critical studies in media communication* 34, no. 1 (2017): 59-68.

Park, Gregory, H., Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Michal Kosinski, David J. Stillwell, Lyle H. Ungar, Martin E. P. Seligman, and no. 6 (2015): 934. “” 108. “Automatic Personality Assessment through Social Media Language.” *Journal of personality and social psychology* 108, no. 6 (2015).

Pateman, Carole. “Participatory Democracy Revisited.” *Perspectives on Politics* 10, no. 1 (2012): 7–19.

Persson, Per-Arne, and James M Nyce. “Technology and Sensemaking in the Modern Military Organization,” 2002.

Petty, Richard E., and John T. Cacioppo. *Attitudes and Persuasion: Classic and Contemporary Approaches*. Dubuque, Iowa: W.C. Brown Co. Publishers, 1981.

Ramley, Ziad. “Cambridge Analytica: A Timeline of Events.” https://medium.com/@ziadramley/cambridge-analytica-a-timeline-of-events-326ab3ef01a9.

Reuters. “Cambridge Analytica and British Parent SCL Elections Shut down after Facebook Scandal Drives Clients Away.” First Post, May 3, 2018. https://www.firstpost.com/world/cambridge-analytica-and-british-parent-scl-elections-shut-down-after-facebook-scandal-drives-clients-away-4454535.html.

Reuters. “Whistleblower Says Canadian Company Worked on Software to Find Republican Voters.” Reuters, https://www.reuters.com/article/us-facebook-cambridge-analytica-election-idUSKBN1H31CK.

Rosenberg, Matthew, Nicholas Confessore, and Carole Cadwalladr. “How Trump Consultants Exploited the Facebook Data of Millions.” *The New York Times*, March 17, 2018. https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html.

Rosenberger, Laura, and Lindsay Gorman. “How Democracies Can Win the Information Contest.” *The Washington Quarterly* 43, no. 2 (2020): 75–96.

Salmon, P.M., N.A. Stanton, and D.P. Jenkins. *Distributed Situation Awareness: Theory, Measurement and Application to Teamwork*. Human Factors in Defence. CRC Press, 2017.

Sandberg, Jörgen, and Haridimos Tsoukas. “Making Sense of the Sensemaking Perspective: Its Constituents, Limitations, and Opportunities for Further Development.” *Journal of Organizational Behavior* 36, no. S1 (2015): S6–32.

Schmidt, Vivien A. “Democracy and Legitimacy in the European Union Revisited: Input, Output and ‘Throughput.’” *Political Studies* 61, no. 1 (2013): 2–22.

Schwaba, Ted, Mijke Rhemtulla, Christopher J. Hopwood, and Wiebke Bleidorn. “A Facet Atlas: Visualizing Networks That Describe the Blends, Cores, and Peripheries of Personality Structure.” *PloS one* 15, no. 5 (2020).

Schwartz, Andrew H., Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah, et al. “Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach.” *PloS one* 8, no. 9 (2013).

Schwartz, Hansen Andrew, Johannes C. Eichstaedt, Lukasz Dziurzynski, Margaret L. Kern, Eduardo Blanco, Michal Kosinski, David Stillwell, Martin EP Seligman, and Lyle H. Ungar. “Toward Personality Insights from Language Exploration in Social Media.” 2013 *AAAI Spring Symposium Series* (2013).

SCL Group. “Kenya: Proposal for TNA/Jubilee Alliance,” 2014. https://twitter.com/hindsightfiles?lang=en.

———. “Kenya: Proposal for TNA/Jubilee Alliance,” 2014. https://twitter.com/hindsightfiles?lang=en.

Smith, Stone & Knight Ltd vs Birmingham Corporation [1939] 4 All ER 116. (n.d.).

Tarran, Brian. “What Can We Learn from the Facebook—Cambridge Analytica Scandal?,” 2018.

Sosnowska, Joanna, Peter Kuppens, Filip De Fruyt, and Joeri Hofmans. “New Directions in the Conceptualization and Assessment of Personality—a Dynamic Systems Approach.” *European Journal of Personality* 34, no. 6 (2020): 988-98.

Susser, Daniel, Beate Roessler, and Helen Nissenbaum. “Technology, Autonomy, and Manipulation.” *Internet policy review* 8, no. 2 (2019).

Tankovska, H. “Number of Monthly Active Facebook Users Worldwide as of 4th Quarter 2020.” https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/.

Tilly, C. *Democracy*. Cambridge: Cambridge University Press, 2007.

Tricco, Andrea C, Jesmin Antony, Wasifa Zarin, Lisa Strifler, Marco Ghassemi, John Ivory, Laure Perrier, Brian Hutton, David Moher, and Sharon E Straus. “A Scoping Review of Rapid Review Methods.” *BMC Medicine* 13, no. 1 (2015): 1–15.

Tsagourias, Nicholas. “Cyber Attacks, Self-Defence and the Problem of Attribution.” *Journal of Conflict and Security Law* 17, no. 2 (2012): 229–44.

Unknown. “Cambridge Analytica - Select 2016 Campaign-Related Documents.” https://archive.org/details/ca-docs-with-redactions-sept-23-2020-4pm/page/n11/mode/2up.

Verhulst, Brad, Lindon J. Eaves, and Peter Hatemi, K. “Correlation Not Causation: The Relationship between Personality Traits and Political Ideologies.” *American Journal of Political Science* 56, no. 1 (2012): 34-51.

Wade, Michael. “Psychographics: The Behavioural Analysis That Helped Cambridge Analytica Know Voters’ Minds.” *The Conversation*, 2018.

Ward, Ken. “Social Networks, the 2016 US Presidential Election, and Kantian Ethics: Applying the Categorical Imperative to Cambridge Analytica’s Behavioral Microtargeting.” *Journal of Media Ethics* 33, no. 3 (2018): 133–48.

Watt, Holly, and Hilary Osbourne. “Tory Donors among Investors in Cambridge Analytica Parent Firm.” *The Guardian*, March 22, 2018. https://www.theguardian.com/politics/2018/mar/21/tory-donors-among-investors-in-cambridge-analytica-parent-firm-scl-group.

Weaver, Matthew. “Facebook Scandal: I Am Being Used as Scapegoat – Academic Who Mined Data.” *The Guardian*, https://www.theguardian.com/uk-news/2018/mar/21/facebook-row-i-am-being-used-as-scapegoat-says-academic-aleksandr-kogan-cambridge-analytica.

Witt, Jesse, and Alex Pasternack. “Before Trump, Cambridge Analytica Quietly Build Psyops for Militaries.” Fast Company, September 2019. https://www.fastcompany.com/90235437/before-trump-cambridge-analytica-parent-built-weapons-for-war.

Witting, C.A. *Liability of Corporate Groups and Networks*. International Corporate Law. Cambridge University Press, 2018.

Woolley, Samuel. The Reality Game: *How the Next Wave of Technology Will Break the Truth and What We Can Do About It*. Great Britain: Endeavour, 2020.

Wright, Glen. “Risky Business: Enterprise Liability, Corporate Groups and Torts.” *Journal Of European Tort Law* 8, no. 1 (2017): 54–77.

Wylie, Christopher. *Mindf* Ck: Inside Cambridge Analytica’s Plot to Break the World*. London: Profile Books, 2019.

References

———. “Written Statement to the United States Senate Committee on the Judiciary: In the Matter of Cambridge Analytica and Other Related Issues.” Senate Judiciary Committee, May 16, 2018. https://www.judiciary.senate.gov/download/05-16-18-wylie-testimony.

York, Geoffrey. “Cambridge Analytica Parent Company Manipulated Nigeria’s 2007 Election, Documents Show.” *The Globe and Mail*, 2018.

Youyou, Wu, Michal Kosinski, and David Stillwell. “Computer-Based Personality Judgments Are More Accurate Than Those Made by

Facebook

Allcott, Hunt, Gentzkow, Matthew & Yu, Chuan. “Trends in the Diffusion of Misinformation on Social Media.” Stanford Institute for Economic Policy Research, October 2018. fake-news-trends.pdf (stanford.edu)

Balfour, Virginia H. “Likes, comments, action! An examination of the Facebook audience engagement strategies used by strategic impact documentary.” *Media International Australia* 176, no. 1 (2020): 34-51. Share, like and achieve: the power of Facebook to reach health-related goals - de la Peña - 2015 - International Journal of Consumer Studies - Wiley Online Library

Bond, Shannon. “In 1st Big Test, Oversight Board Says Facebook, Not Trump, Is The Problem.” *NPR*, May 7, 2021.
https://www.npr.org/2021/05/07/994436847/what-we-learned-about-facebook-from-trump-decision

Cabañas, González, Cuevas, Àngel, Arrate, Aritz, & Cuevas, Rubén. “Does Facebook use sensitive data for advertising purposes?” Communications of the ACM, 64, no. 1 (January 2021): 62-69. Does Facebook Use Sensitive Data for Advertising Purposes? | January 2021 | Communications of the ACM

Cain, Áine. “The 11 coolest perks at Facebook, the best place to work in America.” *Business Insider*, 7 December 2017. https://www.businessinsider.com.au/facebook-employee-perks-benefits-2017-11?r=US&IR=T#a-bike-repair-shop-for-menlo-park-employees-2

Carlson, Nicholas. “At Last – the Full Story of How Facebook Was Founded.” T*ech Insider*, March 5, 2010.
https://www.businessinsider.com.au/how-facebook-was-founded-2010-3?r=US&IR=T#we-can-talk-about-that-after-i-get-all-the-basic-functionality-up-tomorrow-night-1

Christchurch Call to eliminate terrorist & violent extremis content online. New Zealand Ministry of Foreign Affairs and Trade.
https://www.christchurchcall.com/call.html

Christofides, Emily, Amy Muise, and Serge Desmarais. “Information disclosure and control on Facebook: Are they two sides of the same coin or two different processes?” *CyberPsychology and Behavior* 12, no. 3 (2009): 341-345. Information disclosure and control on Facebook: are they two sides of the same coin or two different processes? - PubMed (nih.gov)

Cialdini, Robert, B. *Influence: The psychology of persuasion*. New York: Harper Collins, 1984.

Cialdini, Robert, B. *Pre-suasion: A revolutionary way to influence and persuade*. New York: Random House, 2016.

Constine, Josh. “A flaw-by flaw guide to Facebook’s new GDPR privacy changes.” *Tech Crunch*, April 18, 2018.

Doody, Amy. “Can anyone challenge Facebook’s dominance in social?” *The Irish Times*, 7 February 2020., p.1 Can anyone challenge Facebook’s dominance in social? (irishtimes.com)

“Facebook to Pay $100 Million for Misleading Investors about the Risks It Faced from Misuse of User Data.” *U.S. Securities and Exchange Commission*, July 25, 2019: https://www.sec.gov/news/press-release/2019-140

Federal Trade Commission (US). “FTC Imposes a $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook.” *US Federal Trade Commission, press releases*, July 24, 2019.

Federal Trade Commission (US). “FTC Sues Facebook for Illegal Monopolisation.” *Federal Trade Commission press releases*, December 9, 2020.

Frenkel Sheera, et al. “Delay, Deny, Deflect: How Facebook Leaders Leaned Out in Crisis.” *New York Times*, November 15, 2018.

Fisher, Christine. “Wikipedia co-founder wants to give you an alternative to Facebook and Twitter.” *Engadget*, November 14, 2019.

Fogg, Brian J., and Dean Eckles. “The behavior chain for online participation: How successful web services structure persuasion.” Paper presented at the *International Conference on Persuasive Technology*, Berlin, Heidelberg, 2007. The Behavior Chain for Online Participation: How Successful Web Services Structure Persuasion | SpringerLink

Fox, Gary. “Facebook Business Model: How Does Facebook Make Money.” *Kerry Fox*, March 8, 2020.

Gillett, Rachel. “7 reasons Facebook is the best place to work in America.” *Business Insider*, 7 December 2017. https://www.businessinsider.com.au/facebook-best-place-to-work-in-america-2017-12?r=US&IR=T

Global Internet Forum to Counter Terrorism. https://gifct.org/about/

House of Representatives, (US). “Investigation of Competition in Digital Markets: Majority Staff Report and Recommendations.” *Subcommittee on Anti-Trust, Commercial and Administrative Law of the Committee on the Judiciary*, (Washington: US House of Representatives, 2020).

Ingram, David. “Exclusive: Facebook to put 1.5 billion users out of reach of new EU privacy law.” *Reuters*, April 19, 2018.

Kozinets, Robert, Dolbec, Pierre-Yann, & Earley, Amanda. “Netnographic Analysis: Understanding Culture through Social Media Data.” In *Sage Handbook of Qualitative Data Analysis*, ed. Uwe Flick (London: Sage, 2014), 262-272.

Kramer, Adam, Guillory, Jamie, & Hancock, Jeffrey. “Experimental evidence of massive-scale emotional contagion through social networks.” *Proceedings of the National Academy of Sciences* 111, no. 24 (2014): 8788-8790. Experimental evidence of massive-scale emotional contagion through social networks | PNAS

Larson, Eric. “Twitter, Facebook Win Appeal in Anticonservative Bias Suit.” *Bloomberg*, May 28, 2020.

Lim Heejin, & Schummann, David. “Employing a Dramatugrical Lens to the Interpretation of Brand Online Social Networking: Evidence of Augmented Self.” *Qualitative Market Research: An International Journal 22*, no. 3 (2019): 278-300.
https://www.emerald.com/insight/content/doi/10.1108/QMR-09-2017-0127/full/html

Neate, Rupert. “Over $119bn wiped off Facebook’s market cap after growth shock.” *The Guardian*, July 27, 2018.

Newton, Casey. “The Trauma Floor: The secret lives of Facebook moderators in America.” *The Verge*, 25 February 2019. https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona

Tillman, Maggie. “10 reasons why Facebook has thrived for 15 years.” Pocket-lint, 4 February 2019. 10 reasons why Facebook has thrived for 15 years (pocket-lint.com)

“Open Facebook.” *Forbes*, September 11, 2006. https://www.forbes.com/2006/09/11/facebook-opens-up-cx_rr_0911facebook.html?sh=77b79defa23c

Prier, Jarred. “Commanding the trend: Social media as information warfare.” Strategic Studies Quarterly 11, no. 4 (2017): 50-85.

Security and Exchange Commission. “Facebook to Pay $100 Million for Misleading Investors about the Risks It Faced from Misuse of User Data.” *U.S. Securities and Exchange Commission*, July 25, 2019.

Securities and Exchange Commission. Form 10-K. Annual report pursuant to section 13 or 15(D) of the Securities Exchange Act of 1934. For the fiscal year ended December 31, 2020. Facebook, Inc, (Washington: United States Securities and Exchange Commission, 2020). https://www.sec.gov/Archives/edgar/data/1326801/000132680116000043/fb-12312015x10k.htm

Sherman, Len. “Why Facebook Will Never Change Its Business Model.” *Forbes*, April 16, 2018.
https://www.forbes.com/sites/lensherman/2018/04/16/why-facebook-will-never-change-its-business-model/?sh=1520c38764a7

Sherman, Len. “Zuckerberg’s Broken Promises Show Facebook Is Not Your Friend.” *Forbes*, May 23, 2018.

Social Media News. *Social Media Statistics Australia*, January 2020. https://www.socialmedianews.com.au/social-media-statistics-australia-january-2020/

Taylor, Josh. “Facebook v Apple: the looming showdown over data tracking and privacy.” *The Guardian*, February 14, 2021.

Thompson, Ben. “Facebook’s privacy cake.” *Stratechery*, March 7, 2019.

van Dijck, José. “Facebook as a tool for producing sociality and connectivity.” *Television & New Media* 13, no. 2 (2012): 160-176. Facebook as a Tool for Producing Sociality and Connectivity - José van Dijck, 2012 (sagepub.com)

Waldman, Ari. “Manipulating trust on Facebook.” *Loyola Consumer Law Review* 29, no. 1 (2016): 175-198. Manipulating Trust on Facebook by Ari Ezra Waldman :: SSRN

Wong, Julia Carrie. “Facebook’s underclass: as staffers enjoying lavish perks, contractors barely get by.” The Guardian, 26 September 2017. https://www.theguardian.com/technology/2017/sep/26/facebook-workers-housing-janitors-unique-parsha

Wohn, Donghee Yvette, and Brian J Bowe. “Micro agenda setters: The effect of social media on young adults’ exposure to and attitude toward news.” Social Media + Society 2, no. 1 (2016): 1-12.

Minda, Zetlin, “Here’s Why Facebook’s Former Employees Describe the Company as Cult-Like.” *Inc*, Undated. https://www.inc.com/minda-zetlin/facebook-culture-cult-like-former-employees-report.html

Zuckerberg, Mark. “A Privacy Focused Vision for Social Networking.” *New York Times*, March 6, 2019.


THE UNIVERSITY
of ADELAIDE

THE UNIVERSITY OF
MELBOURNE

UNSW

AUSTRALIA
ECU
EDITH COWAN
UNIVERSITY

MACQUARIE
University
SYDNEY·AUSTRALIA