THE UNIVERSITY of ADELAIDE

Submission to the Senate Committee on Foreign Influence through Social Media

Defence Security Institute

make history.

Submission Contributors

Professor Michael Webb, Director, Defence and Security Institute, University of Adelaide

Professor Debi Ashenden, Professor of Cybersecurity, School of Computing Science, University of Adelaide

Associate Professor Tim Legrand, Department of Politics and International Relations, University of Adelaide

Professor Dale Stephens, Adelaide Law School, University of Adelaide

Professor Lewis Mitchell, School of Computer and Mathematical Sciences, University of Adelaide

Associate Professor Carolyn Semmler, School of Psychology, University of Adelaide

Dr Keith Ransom; Postdoctoral Research Fellow, School of Psychology, University of Adelaide

Dr Rachel Stephens, Senior Lecturer, School of Psychology, University of Adelaide

Dr Matteo Farina; Postdoctoral Research Fellow, School of Psychology, University of Adelaide

Dr Matthew Kaesler, Lecturer, School of Psychology, University of Adelaide

Contact

Professor Michael Webb
Director
Defence and Security Institute
University of Adelaide

1.1. Overview

The problem of Foreign Influence through Social Media is complex and rapidly evolving. Technology is ubiquitous, adaptive, pervasive and personalisable. It is unlikely that any single policy or legal reform will protect democracies from disruption in the future. The University of Adelaide has a strong expertise in transdisciplinary research into the structure, governance and human behavioural problems presented by Foreign Interference (FI) through Social Media. The teams of experts within the Defence Security Institute have assembled considerable technical and research capability to address major issues of methodology, policy and prevention of FI. The breadth of disciplines captured by this research effort includes; Psychology, Linguistics, Physical Sciences, Politics, Computer Science, Mathematics and Data Science, Bioscience, International Security, Education and Engineering. We address two aspects of the terms of reference for this Committee:

a. the use of social media for purposes that undermine Australia's democracy and values, including the spread of misinformation and disinformation;

b. responses to mitigate the risk posed to Australia's democracy and values, including by the Australian Government and social media platforms;

1.2. University of Adelaide Research Expertise

The University of Adelaide has extensive research capabilities and capacity in the information and cyber domain. At the forefront of defensive and offensive cyber operations is research in the disciplines of data science, psychology, artificial intelligence (AI) and autonomy, and applied mathematics. Examples of the kind of research undertaken include:

- The study of information flow over networks, yielding techniques and algorithms to prediction influence likelihoods of individuals and groups.
- The development of a range of automated approaches to enhancing the security of our cyber infrastructure, particularly in the context on active threats. These approaches are based on advanced meta-graph algebras, game theory (counterfactual regret minimization) and reinforcement learning techniques.
- The use of advanced machine learning techniques to implement cyber deception processes to counter and exploit opportunities raised by the actions of malicious intruders inside our networks.

Two defence funded projects in the last two years have mapped the digital elements of Australia's democratic infrastructure, identified emerging vulnerabilities and threats in the information environment, and their consequences; analysed the social dimensions of digital politics; and designed training scenarios and environments in non-kinetic conflicts.

During 2020 University of Adelaide teams participated in the Joint Influence Assessment task, engaged by the Department of Defence. The teams used the Cambridge Analytica case as a study to show how social media platforms can be manipulated and deployed by individuals and teams in an effort to influence

the outcome of elections in democratic societies. Cambridge Analytica (CA) was a subsidiary of Strategic Communications Laboratories (SCL), a British private behavioural and strategic communication company which specialised in military and intelligence operations that were aimed at influencing public opinion. In the 1990s SCL moved into politics and it was apparently involved in political elections in Australia, Azerbaijan, Bahrain, Colombia, Ecuador, Egypt, Iran, Italy, Mexico, Morocco, Russia, Syria, the United Kingdom, and Venezuela.

CA was involved in the 2016 American elections. During these elections, CA worked for the Republicans. Apparently, CA provided the Republican Party with psychological profiles of approximatively 87 million American voters. These profiles were used to develop personalised political ads for these individuals. However, it seems that many of these psychological profiles were created using Facebook data that were collected without the consent of their owners and could not be used for commercial purposes. In 2018, after the misuse of data was disclosed, CA and SCL ceased to operate.

Influence operations can be carried out by a relatively small workforce (e.g., the Russian Internet Research Agency) and fall outside of the realm of traditional military conflict, making them difficult to combat. In this way, the contemporary strategic environment of Australia is now best described as complex. The Defence Strategic Update (2020) also strongly articulates the need to deter actions against Australia and its national interests, which includes ensuring that the Australian public and its institutions remain free of interference.

These “wicked” problems will not be addressed by applying a single theoretical or methodological lens. Instead, they will invariably require a fusion of complementary methodologies and theoretical perspectives. For example, quantitative and mathematical methods which can manage large volumes of data must be merged with the theoretical and experimental paradigms often used in the more cognate disciplines (e.g., psychology, sociology, anthropology). It is at this interface that we work and bring together the existing communities of psychologists and human behavioural scientists together with data scientists and mathematical modellers, and focus them on the problem of large-scale social influence research in the online and offline worlds. These teams can play a significant role in the development of a national capability to **detect**, **defend**, and **disrupt** social influence operations both online and offline. Our approach is innovative and has a wider disciplinary base than other teams investigating the issue of FI via social media. Our aim is in producing novel digital solutions to detecting malign influence online, countering narratives that threaten individual and public safety and developing evidence for policy and governance solutions that are based upon innovative scientific research. Below we outline the learning from our case study of the Cambridge Analytica affair and what this means for the use of social media to undermine Australia’s democracy and values – including the spread of misinformation and disinformation.

1.3. Learning from CA – extracts from Webb et al. 2022

Our case study (Webb et al. 2022) on CA has shown that it heavily relied on technology for its political campaigns. On one hand, CA utilised technology to harvest data from social media platforms and other

data sources to build profiles of voters. In fact, as clearly stated in one of its political brochures, CA used technology to create:

> "an enriched voter file, developed using a comprehensive range of election, consumer, lifestyle, social media, personality and other datasets. We create advanced models that predict voter behavior in a number of different areas, ranging from likelihood to turn out on Election Day to how they might vote on a specific ballot initiative or their propensity to donate".[1]

On the other hand, CA used technology to reach specific groups or individuals with personalised messages aimed at persuading them.[2] CA claimed that "[it] put the right message in front of the right person at the right time".[3] In other words, it seems that CA used technologies for:

- Analysing psychological data and identifying key target voter groups
- Developing, testing and refining campaign specific messages for the target voter groups
- Assisting deployment of messaging through different media channels to target specific voter groups.[4]

Therefore, it is likely that CA heavily relied on technology for its political operations because: (1) technology has changed the way in which people communicate with each other and access information[5]; (2) technology affects politicians and how they interact with voters.[6] (3) Technology is "persistent, ubiquitous, allow[s] anonymity, can store huge volumes of data, can use many modalities, [and] can scale".[7] (4) Technology is interactive and *personalisable*[8], it can be used to deliver messages that are relevant to individuals and might influence their attitudes and behaviours.[9] In other words, a message generated by a computer might be tailored to suit a person's needs, interests, passions, etc. thus, making it more relevant

[1] Cambridge Analytica, "Ca Political an Overview of Cambridge Analytica's Political Division," (2015).
[2] Emma L. Briant, "Cambridge Analytica and Scl – How I Peered inside the Propaganda Machine," *The Conversation* (2018); Carole Cadwalladr, Emma Graham-Harrison, "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach," *The Guardian* 2018; Angela Chen 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; Ellen Emilie Henriksen, "Big Data, Microtargeting, and Governmentality in Cyber-Times. The Case of the Facebook-Cambridge Analytica Data Scandal" (2019); Matthew Hindman, "How Cambridge Analytica's Facebook Targeting Model Really Worked – According to the Person Who Built It," *The Conversation* 2018; Michael Wade, "Psychographics: The Behavioural Analysis That Helped Cambridge Analytica Know Voters' Minds," ibid.; Christopher Wylie, *Mindf* Ck: Inside Cambridge Analytica's Plot to Break the World* (Profile Books, 2019); Carole Cadwalladr and Emma Graham-Harrison, "How Cambridge Analytica Turned Facebook 'Likes' into a Lucrative Political Tool," *The Guardian* 2018.
[3] Analytica, "Ca Political an Overview of Cambridge Analytica's Political Division."
[4] Unknown, "Cambridge Analytica - Select 2016 Campaign-Related Documents," https://archive.org/details/ca-docs-with-redactions-sept-23-2020-4pm/page/n11/mode/2up.
[5] Paul Foley, "Does the Internet Help to Overcome Social Exclusion," *Electronic Journal of e-government* 2, no. 2 (2004); Wallace Chigona et al., "Can Mobile Internet Help Alleviate Social Exclusion in Developing Countries?," *The Electronic Journal of Information Systems in Developing Countries* 36, no. 1 (2009).
[6] Brian L. Ott, "The Age of Twitter: Donald J. Trump and the Politics of Debasement," *Critical studies in media communication* 34, no. 1 (2017).
[7] B. J. Fogg, *Persuasive Technology Using Computers to Change What We Think and Do*, The Morgan Kaufmann Series in Interactive Technologies (Amsterdam: Morgan Kaufmann Publishers, 2003).
[8] Ryan Calo, "Digital Market Manipulation," *George Washington Law Review* 82, no. 4 (2013).
[9] Fogg, *Persuasive Technology Using Computers to Change What We Think and Do*.

to him or her.[10] (5) Technology might be persuasive, it could be "designed to change a person's attitude, behaviour of both".[11]

Whether technologies (ie. psychological profiling of the type used by CA) affect peoples' attitudes and behaviours is still an open question. Policy responses in this area require evidence and consideration of the possibility that ungoverned areas of the internet will still allow FI operations to occur.

Persuasion, manipulation and coercion

Susser et al.[12] clearly stated that CA used online manipulation for its political operations. More precisely, they claimed that "Since 2016, when the Facebook/Cambridge Analytica scandal began to emerge, public concern has grown around the threat of "online manipulation"".[13] Online manipulation is a specific form of manipulation; manipulation, persuasion, and coercion are part of the influence continuum. Persuasion is positioned at one end of this continuum, coercion at the other end, and manipulation is placed in the middle. As clearly shown by Figure 1 below.

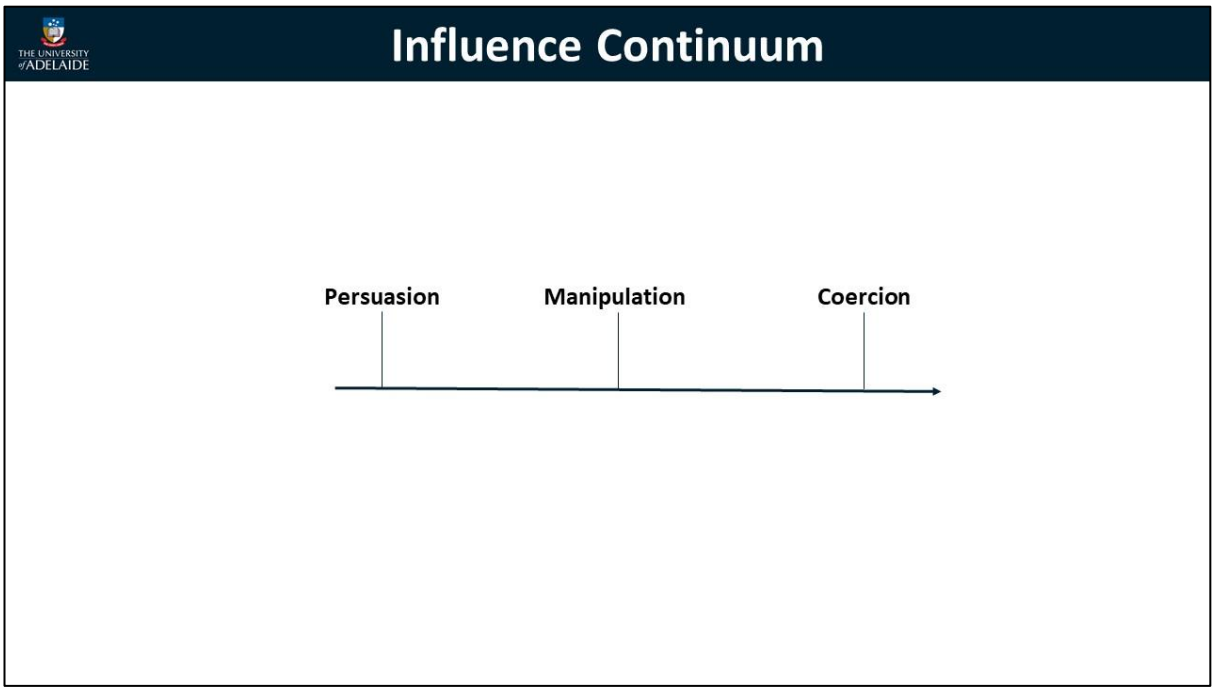

Figure 1. The Influence Continuum.

Persuasion, or at least rational persuasion, is defined as an attempt to influence people by providing them with reasons they can evaluate and act on. Moreover, persuasion is intentional and transparent.[14] It is also characterised by freedom of choice and lack of harm. In other words, persuaders do not force their beliefs,

[10] Ibid.
[11] Naomi Jacobs, "Two Ethical Concerns About the Use of Persuasive Technology for Vulnerable People," *Bioethics* 34, no. 5 (2020).
[12] Daniel Susser, Beate Roessler, and Helen Nissenbaum, "Technology, Autonomy, and Manipulation," *Internet policy review* 8, no. 2 (2019).
[13] Ibid.
[14] Ruth R. Faden, Tom L. Beauchamp, and Nancy M. P. King, *A History and Theory of Informed Consent* (New York: Oxford University Press, 1986).

attitudes, values, or intentions on other persons. Persuasion can be rational or non-rational.[15] Rational persuasion is when an individual uses reason and arguments to influence another person, whereas non-rational persuasion takes advantage of emotions, peer pressure, authority, and other means to influence people.[16]

Coercion, on the other hand, limits the options of individuals and forces them to take actions.[17] While persuasion is acceptable, coercion is not. Although placed at the opposite ends of the influence continuum, persuasion and coercion share an important similarity; they are both overt. That is, if someone is trying to persuade or coerce someone else, this person normally knows it. However, when coerced, individuals are not free to act because of threats that might cause them unwanted and avoidable harm. Coercion can also occur among countries. The Australian Defence Glossary uses the term *grey zone* to refer to coercive activities that might be undertaken by a nation to force another country to take actions that could avoid a potential military conflict.

CA apparently used Facebook posts because these messages are more likely to appeal to people's personalities, demographics as well as their priority issues.[18] For example, images, memes and posts that seem to exploit a cognitive bias against illegal immigrants generate emotions, which in turn might affect political discourses.[19] Therefore, FI campaigns like those orchestrated by CA seemed to explicitly target voters interested in immigration and national security.

In summary, social media websites are resource of data for psychologically profiling individuals. Although, it seems possible to use social media data to infer individuals' personalities, it is still unclear whether personality traits are good predictors of political behaviours. In summary, CA clearly set out to disrupt democracy and influence voter behavior – whether they were successful in this is another question that cannot be clearly answered based upon available data.

1.4. Conclusion

After discussing some features of technology which were relevant to CA's operations, this report has focused on persuasion, manipulation and coercion. Unfortunately, CA is not the last in a line of companies that seek to profit from the debasement of democratic principles – more sophisticated operations have now appeared such as 'Aims' a software that is designed to coordinate the spread of misinformation across multiple social media platforms ('Aims': the software for hire that can control 30,000 fake online profiles | Technology | The Guardian). This will continue to occur as long as social media platforms are self or minimally governed spaces.

[15] Richard E. Petty and John T. Cacioppo, *Attitudes and Persuasion: Classic and Contemporary Approaches* (Dubuque, Iowa: W.C. Brown Co. Publishers, 1981).
[16] Jacobs, "Two Ethical Concerns About the Use of Persuasive Technology for Vulnerable People."
[17] Susser, Roessler, and Nissenbaum, "Technology, Autonomy, and Manipulation."
[18] Wylie, *Mindf* Ck: Inside Cambridge Analytica's Plot to Break the World*.
[19] Adler-Nissen, Andersen, and Hansen, "Images, Emotions, and International Politics: The Death of Alan Kurdi."

1.5. Key Recommendations to mitigate risk posed to Australia's democracy and values:

- Consider using multidisciplinary teams of experts to analyse and develop understanding of the effects of persuasive technologies and coordinate responses in terms of policy and public education.
- Consider a 'social media protectorate' that monitors and counters co-ordinated online interference in democratic processes, particularly prior to and during elections.
- Consider a code of practice for the accountability of social media platforms,that addresses where current content moderation practices that fall short of expectations in terms of timeliness, identity verification and democratic values.
- Use evidence of influence and behavior change to inform policy, informed by complementary data science approaches that scale to identify and disrupt malicious actors.

1.6. Summary of Submission

Our extensive examination of just one influence campaign shows the complexity of issues that span many disciplines that are required to gain a comprehensive picture of governing in the new information environment. The strategic capability of Australia's defence requires a careful and guided approach to policy informed by the most comprehensive understanding of online and offline behavior, as well as the features and capabilities of social media technology. The University of Adelaide stands at the forefront of capabilities in this area with partnerships that span defence, cyber and national security contexts. Our extensive capabilities in data science, psychology, cybersecurity, policy and law allow us to understand the complexity of FI operations and provide the evidence required to deliver the capability to limit these threats to our democracy.