

Dealing with the “Fake News” Problem

Supplementary Submission

to the

Joint Standing Committee on Electoral Matters

19th February, 2019

Dr Carlo Kopp, Dr Kevin B. Korb, Dr Bruce I. Mills

Note

This supplementary submission extends and amplifies answers provided by Dr C. Kopp during the Committee hearing on the *Australian Electoral Commission annual report 2017-18*, held in Canberra on the 13th February, 2019.

Additional Responses to Committee Questions of the 13th February, 2019

Mr DICK: I'm interested in how the fake news aspect intersects with freedom of speech. One person's fake news is another person's right to have their say. I'm interested in your view on whether we could put into place any guidelines or structures in the lead-up to the next election to make sure that accurate information is distributed while freedom of speech is also protected.

Mr DICK: I don't think anyone is suggesting we should have a huge vault of people sitting behind desks looking at every single social media post that comes out during an election campaign. That's not practical. I guess as parliamentarians we want to make sure the information that is being distributed through properly authorised, legal content in an election campaign, whether it be through electronic media, social media, marketing or whatever, in some sense meets community standards. That's the \$64 question: how do we get that balance right?

Additional Response:

In terms of the misuse of technology for meddling in elections and politics, prevention is better than attempting to clean up a mess after the fact. The problem is not inherent in the technology, but rather the consequence of improper or malign human behavior, exploiting what the technology allows.

The exclusion of malign actors, foreign and domestic, from social and mass media does not impact freedom of speech. Censorship of individuals usually does.

Most instances where significant disruption to elections or referenda occurred overseas involved malign nation-state actors, who have the material resources for sustained funding of a large pool of personnel dedicated to trolling activities in social media, including the production of disruptive or biased “fake news”, the management of pools of social media “bots”, and providing the supporting cyber-attack capabilities to steal politically sensitive data and then exploit it for the aforementioned purposes, or to conduct effective denial of service attacks against websites of interest.

The most brazen instance of this kind was observed in Ukraine during the 2014 elections, where the Central Electoral Commission tally computer was hacked and an attempt was made to get the official tallies to be displayed in television and online media replaced by fake tallies. The game was blown as cyber-security personnel were able to disable the malware in time, while Russian media showed on television the rigged tally having not determined that the hack had already been interdicted.

Most of the successful attacks overseas exploited poor cyber-security, known vulnerabilities or “exploits” in consumer software, poor password management practices, a failure to check that social media accounts belonged to real people, and a plethora of other very common bad practices in the management and operation of the digital infrastructure. In the US notable examples include the hacking of the DNC to collect emails that could be used for disruptive propaganda, and the hacking of

state electoral office computers to access voter records. Stolen emails were often mischievously altered to mislead the public as to what was actually stated, a deception tactic employed by the Soviets many times through the Cold War era.

Most of these problems can be pre-empted by proper cyber-security practices and good “cyber-hygiene” and user discipline, and use of firewall systems where appropriate.

We recommend comprehensive and periodic cyber security audits of all Commonwealth digital infrastructure, that could be subject to malign actions during an election campaign, in line with proper practice observed overseas. Appropriate training should be given to personnel using these systems to reduce the risk of successful “phishing”, “spear-phishing” or other attacks reliant on seducing human operators. The recent cyber-attack against the Parliament could most likely have been prevented by robust cyber security practices.

Distribution of propaganda and disruptive “fake news” in previous overseas attacks exploited primarily absent or weak account creation and ownership validation practices by social media platform operators, that allowed accounts to be set up with fake identities that could then be used *en masse* by human trolls and social media “bots”.

Recent media reports indicate practices are beginning to change, with Facebook closing down hundreds of such fake accounts being employed to meddle in the Moldovan and Ukrainian, and a number of other election campaigns now under way (<https://newsroom.fb.com/news/2019/>).

This is a good example of both a denial strategy, to prevent access, and of increasing the cost of malign activity by hindering the use of automation – using individual human operators to manage each fake social media account via proxy email accounts is inherently more expensive than using software robots.

Measures of this kind can significantly reduce opportunities for malign actors to inject toxic propaganda and disruptive “fake news” into social media networks. Some recent reports indicate that malign actors are now employing identity theft or constructing plausible fake identities with networks of plausible fake friends and followers, indicated that the tightening of practices is beginning to produce effect.

The use of fake identities has been a recurring feature of cyber-attacks, spam, and social media hosted attacks. An insistence that only genuine authentic users be allowed to use social media accounts is not unreasonable and in no way restricts freedom of speech.

The aforementioned measures as employed overseas will reduce the scale of the “fake news” problem in social media, by interdicting access by malign actors, and hampering distribution. It does not eliminate the problem, as a determined attacker,

especially if a nation state, will have the resources to set up elaborate fake identities that will be time consuming and thus expensive to unmask.

Purging social media news feeds of known fake media and propaganda sources such as websites is a practice that has been pursued in Europe, for instance by the Ukrainians, who have also insisted that mass media identify what they know or suspect to be propaganda or hostile “fake news” with warning messages to viewers or readers. Cable television and other media operators who have consistently been distributing hostile propaganda have been closed down, and in a number of instances, the operators have been charged with offences related to aiding and abetting wartime enemies, and even treason. While most instances where this regulation has been applied are unambiguous instances of foreign information attack, there has been ongoing controversy over its misuse in some instances.

Where a producer of toxic or malicious “fake news” used to interfere with an election is based in or operates in Australia, they could be charged under Sections 137.1 and 137.2 with producing false or misleading statements or documents.

Australia should also carefully consider the measures proposed in the UK Parliament DCMS Committee Final Report on the “Disinformation and ‘fake news’” inquiry, as many will be directly applicable to Australia, given our Westminster based system, noting that there are differences in many areas of UK law that would need to be addressed should the proposed UK measures be introduced wholly or in part in Australia.

The problem of authentic users redistributing “fake news” content collected from outside social media was addressed in earlier evidence. It is in many ways the most challenging problem to solve, as noted earlier.

The difficulties noted in earlier evidence with mass media redistributing “fake news”, and in some instances creating it, may also be challenging to easily address due to the economics of the digital environment. Fact checking and validating news reports takes time and effort, and also demands suitable skills sets, all of which cost money and this often disadvantages media organisations that care about the integrity of their news, against competitors who care not. There is currently a strong economic disincentive to high quality news reporting.

The problem of bias in the mass media, especially where the bias impairs accuracy, appears to be a consequence of the economic advantages arising from “*playing to the audience*”, as reports that appeal to the prejudices or biases in an audience sector will be more popular, and thus attract more revenue from the audience. This appears to be an underlying cause in the polarization of mass media along political lines in most Western democracies.

Impartiality in reporting and an absence of bias is often a part of regulatory regimes for media licences, but enforcement has been highly variable over time, and approaches have differed across different democracies.

The United States for decades maintained the often very effective *FCC Fairness Doctrine*, which was removed by the Reagan Administration due to a belief that it conflicted with the *First Amendment* and restricted freedom of speech. The United States as noted in earlier evidence now confronts a problem with biased and often wildly inaccurate “*hyper-partisan*” media coverage that has been the subject of many very bitter complaints by legislators on both sides of the political aisle, and has clearly contributed to a loss of public trust in many media organisations.

Fact checkers shaming media organisations for bias or errors has not produced any observable effects. Competing media organisations (or again NGOs and politicians) publicly shaming other media organisations has often produced results.

Both of these problems, as noted in earlier evidence, are essentially the result of traditional media quality assurance mechanisms collapsing to maintain or improve revenue flows.

A change in audience preferences that rejects strong bias and inaccurate reporting would produce an effect, as it would directly impact earnings. Changing audience preferences would require new investments in public education.

As long as bad practices in the media are rewarded by revenue, they will persist.

Mr PITT: I have only one question, just to garner a little bit more knowledge. What's the next iteration? We're looking at fake news now and what those challenges might mean. What do you think is the next issue we will strike and can we address that before it actually occurs? There's a candidate in Queensland now who's using gaming for data collection for political purposes. In your view, where's our next challenge?

Additional Response:

The problem of deep fakes is one of major concern, to the extent that the US *Defense Advanced Research Projects Agency* has generously funded a major research effort to develop forensic tools that are sufficiently robust to detect a deep fake with a high level of confidence (<https://www.technologyreview.com/s/611726/the-defense-department-has-produced-the-first-tools-for-catching-deepfakes/>).

There is an expectation that in the long term, the problem will devolve into a persistent “arms race” between developers of software tools to produce deep fakes, and developers of forensic software tools to detect them. This pattern of “*measure versus countermeasure*” has been a recurring feature of any competitive environment, where the stakes are high. The deep fake problem is also a good example of how most technologies with important legitimate uses (cinema, commercial advertising) can be repurposed for criminal or other destructive purposes.

Another emerging technology following much the same pattern are AI software tools for language processing, capable of understanding and generating natural language content in text documents, and by extension, in spoken voice exchanges.

Recent reports from the US on the results of research by the *OpenAI* research institute are a good example – the software is capable of producing large volumes of fake news if repurposed, some of which could be good enough to fool less adept audiences (<https://www.technologyreview.com/s/612960/an-ai-tool-auto-generates-fake-news-bogus-tweets-and-plenty-of-gibberish/>). *OpenAI* have also raised concerns that the technology could be employed to improve the effectiveness of deceptive “chatbots”, programs that pretend to be a human and engage in an email or social media conversation intended to trick the victim into giving away sensitive information.

The current technology being used by cybercriminals in “chatbots”, and political players in social media e.g. “twitterbots”, is a repurposing of software that was developed for stress testing computer systems over two decades ago, by running simple software robots that emulated human activity on a keyboard. This is yet another example of technology with important legitimate uses being repurposed for criminal or other illegitimate purposes, like improperly influencing election outcomes. Current jargon for this problem is to label it as a “*Black Mirror*” scenario (<https://www.technologyreview.com/s/610321/the-black-mirror-scenarios-that-are-leading-some-experts-to-call-for-more-secrecy-on-ai/>).

History tells us that this is a problem that will likely persist indefinitely. There is archaeological evidence showing that millenia ago, stone axes with legitimate uses in gathering food were quickly repurposed as weapons. The intensive technological competition of the *Great War*, *World War Two*, and the *Cold War*, displayed exactly the same pattern. The repurposing of “drones” with legitimate commercial uses for use as weapons in the Syrian and Ukraine wars is a very good contemporary example. The notion that such competitive behaviours will not persist is optimistic.

Mrs WICKS: You've thrown me with this whole deep fake thing. I'm absolutely fascinated now. Are there any sort of regulatory solutions to some of these things you're seeing on the horizon? What can a government or a parliament do and is there anything that can be done? Obviously, awareness is one of the things. You've raised that, but—

Additional Response:

In many Western nations, government bureaucracies appear to have a firmly held conviction that all problems involving advanced technology should be solved by unilateral regulation. The ongoing matter of the *Defence Trade Controls Act* and its adverse impacts is the Australian case study. If taken to its logical conclusion, the unilateral application of intensive and aggressive regulatory regimes such as the *Defence Trade Controls Act* resolves the problem of improper use or disclosure of advanced technology in Australia by killing off all research and development in the

regulated technology, and related public discourse. If there is no such technology being researched and developed, or publicly debated in Australia, it cannot be acquired by foreign opponents or competitors. Equally so, most or all benefits from the legitimate use of the technology are ceded to other nations that pursue a more rational and less dogmatic approach to the regulation of technology and related science than Australia has followed in recent years. If Australia wishes to be competitive in an increasingly technological world, it will need to completely rethink how it approaches this problem.

A subject not previously discussed in our submission or the hearing is Australia's level of national investment in developing and maintaining the skills base and technology base to be able to deal with current and future challenges in this area. The United States was able to respond, a little late due to hesitant executive decision-making, to large scale Russian meddling in elections, by making use of a robust skills set in government agencies, and a well funded research community in universities, think tanks and industry. This is a pattern observed earlier in Europe, and Estonia is often cited as a case study of mobilizing national capabilities to good effect.

While Australia has a reasonably strong research capability in areas such as computer security and cyber, and AI, other research capabilities necessary to tackle problems such as election interference via digital media are much weaker. This research problem area has traditionally fallen under the umbrella of "*Information Warfare*", or more narrowly "*Information Operations*". While Australia had a small but very active research community in this area a decade ago, Commonwealth policy on research, intended to improve international benchmark rankings, selectively prioritized funding and other rewards for large research communities, over small research communities.

Research in the "*Information Warfare*" domain became "*collateral damage*", and is far less active now than a decade ago. The problem is further compounded by the reality that research in this area is inherently cross/inter-disciplinary, spanning areas that include *computer security and AI, information systems, human factors and psychology*, and *strategy*. This style of cross/inter-disciplinary research is not a good fit for a highly structured Commonwealth system designed to strongly reward research results in highly active single-discipline research areas, and the result is that many inter-disciplinary and niche research areas are being actively killed off to improve international benchmark numbers.

Short-term popularity is not necessarily a good measure of the long-term strategic importance of a research area, and historically, short-term predictions of the long-term value of research have often been very inaccurate (there are countless examples of now highly valuable technology and science being undervalued when initially discovered).

If Australia wishes to protect itself effectively in the longer term against malign actors in the digital domain, it will need to rethink how it develops and maintains a national

skills base to understand the problem, develop capabilities and provide advice on policy, and education and training for government agency personnel. Smaller European nations have built up robust capabilities in this area, so this is an achievable outcome for Australia, if the Commonwealth accords appropriate priority to research funding and less aggressive regulation of many key component research areas. Australia's current weakness in this area could have been avoided had a less dogmatic approach been adopted in manipulating priorities in research over the last decade.

Ultimately, success in this game is about out-smarting and pre-empting malign actors. Without the necessary skills base in Australia, we will end up in a perpetual game of "catch-up" and suffer accordingly.

References:

1. Kopp C, Korb KB, Mills BI "Information-theoretic models of deception: Modelling cooperation and diffusion in populations exposed to 'fake news'". PLoS ONE 13(11), November, 2018: e0207383, URI: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0207383>
2. Kopp, C, Korb, K.B, "We made deceptive robots to see why fake news spreads, and found its weakness", The Conversation, November, 2018, URI: <https://theconversation.com/we-made-deceptive-robots-to-see-why-fake-news-spreads-and-found-a-weakness-104776>
3. Kopp, C, Korb, K.B, Mills, B.I., "Understanding the Inner Workings of "Fake News"", Science Trends, November, 2018, URI: <https://sciencetrends.com/understanding-the-inner-workings-of-fake-news/>
4. Kopp, C., "Understanding the Deception Pandemic", Presentation Slides, Australian Skeptics Seminar, 16th July, 2018, Melbourne, Australia, URI: <http://users.monash.edu/~ckopp/Presentations/Understanding-The-Deception-Pandemic-2018-B.pdf>
5. Carlo Kopp, Kevin B. Korb, Bruce I. Mills, Written Evidence, "Inquiry on Disinformation and 'fake news'", House of Commons Digital, Culture, Media and Sport Committee, 12 December 2018, URI: <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/disinformation-and-fake-news/written/93672.html>
6. Wineburg, Sam and McGrew, Sarah and Breakstone, Joel and Ortega, Teresa. (2016). "Evaluating Information: The Cornerstone of Civic Online Reasoning." Stanford Digital Repository. Available at: <http://purl.stanford.edu/fv751yt5934>
7. Samantha Bradshaw, Lisa-Maria Neudert, Philip N. Howard, "COUNTERING THE MALICIOUS USE OF SOCIAL MEDIA: GOVERNMENT RESPONSES TO MALICIOUS USE OF SOCIAL MEDIA", NATO STRATCOM COE, Riga, November 2018, 11b Kalciema Iela, Riga LV1048, Latvia, URI: <https://www.stratcomcoe.org/download/file/fid/79655>
8. Kopp C., "Considerations on deception techniques used in political and product marketing", Proceedings of the 7th Australian Information Warfare and Security Conference, 4 December 2006 to 5 December 2006, School of Computer Information Science, Edith Cowan University, Perth WA Australia, pp. 62-71. URI: <http://users.monash.edu/~ckopp/InfoWar/Lectures/Deception-IWC7-2006-BA.pdf>
9. Kopp C., "Classical Deception Techniques and Perception Management vs. the Four Strategies of Information Warfare", in G Pye and M Warren (eds), Conference Proceedings of the 6th Australian Information Warfare & Security Conference (IWAR 2005), Geelong, VIC, Australia, School of Information Systems, Deakin University, Geelong, VIC, Australia, ISBN: 1 74156 028 4, pp 81-89. URI: <http://users.monash.edu/~ckopp/InfoWar/Lectures/Deception-IWC6-05.pdf>
10. Kopp C., "The Analysis of Compound Information Warfare Strategies", in G Pye and M Warren (eds), Conference Proceedings of the 6th Australian Information Warfare & Security Conference (IWAR 2005), Geelong, VIC, Australia, School of Information Systems, Deakin University, Geelong, VIC, Australia, ISBN: 1 74156 028 4, pp 90-97. URI: <http://users.monash.edu/~ckopp/InfoWar/Lectures/Method-IWC6-05.pdf>
11. Kopp C., "Shannon, Hypergames and Information Warfare", in W Hutchinson (ed), Proceedings of the 4th Australian Information Warfare & Security Conference 2003 (IWAR 2003). Perth WA Australia, 28 - 29 November 2003, Edith Cowan University, Churchlands WA Australia, ISBN: 0-7298-0524-7. URI: <http://users.monash.edu/~ckopp/InfoWar/Lectures/JIW-2002-1-CK.pdf>
12. Kopp C. and Mills B.I., "Information Warfare and Evolution", in W Hutchinson (ed), Proceedings of the 3rd Australian Information Warfare & Security Conference 2002 (IWAR 2002). Perth WA Australia, 28 - 29 November 2002, Edith Cowan University, Churchlands WA Australia, ISBN: 0-

7298-0524-7, pp 352-360. URI: http://users.monash.edu/~ckopp/InfoWar/Lectures/_JIW-2002-2-CK-BIM.pdf

End of Supplementary Submission