Submission to the Senate Select Committee on Foreign Interference through Social Media

Dr Jake Wallis and Mr Thomas Uren are Senior Analysts at the Australian Strategic Policy Institute (ASPI). We research disinformation and the deliberate manipulation of the information environment to achieve strategic goals—what we will refer to here as influence operations. The views expressed here are our personal opinions as ASPI does not take corporate positions on any issues.

Deliberate manipulation of the information environment

Access to accurate unbiased information is a pre-condition for effective decision-making, yet malign actors are engaged in organised and concerted efforts to manipulate the information environment to achieve their strategic goals. Authoritarian states have identified influence operations as a cheap yet effective mechanism for influencing and weakening liberal democratic societies.

Russian meddling in the 2016 US election has become perhaps the most well-known and best documented case study of foreign interference through social media and it is a striking, but not representative, example. These influence operations are not limited to nation-states—there are a range of actors, with diverse motivations, who are willing and able to manipulate social media audiences at scale. And these influence operations are not limited to elections—they are persistent, ongoing, and are used to pursue a range of strategic goals.

Elections have created a new business opportunity for financially motivated malign actors, due to the heighted levels of public sentiment and engagement in online political discourse. During the 2019 Australian federal election financially-motivated actors from Kosovo, Albania and the Republic of North Macedonia used nationalistic and Islamophobic content to target and manipulate Australian Facebook users.[1] A combined audience of 130,000 Facebook users across four Facebook pages were steered off the platform towards content farms that generated advertising revenue from each page view. The Guardian uncovered a similar operation run from Israel that used similarly divisive Islamophobic content, again to steer Facebook audiences to revenue-generating content farms.[2] These activities have the potential to skew Australia's political discourse, influence voting behaviour and affect electoral outcomes.

Beyond elections, research at ASPI has found influence operations relating to Indonesia's West Papua independence movement[3], Kashmir[4], and PRC operations targeted at various political dissidents and

[1] Workman, Michael, and Stephen Hutcheon. 2019. "Popular Australian Facebook Pages Manipulated by Trolls from the Balkans." *ABC News*. https://www.abc.net.au/news/2019-03-15/trolls-from-kosovo-are-manipulating-australian-facebook-pages/10892680 (March 13, 2020).

[2] Knaus, Christopher, Nick Evershed, Michael McGowan, and Oliver Holmes. 2019. "Inside the Hate Factory: How Facebook Fuels Far-Right Profit." *The Guardian*. https://www.theguardian.com/australia-news/2019/dec/06/inside-the-hate-factory-how-facebook-fuels-far-right-profit (March 13, 2020).

[3] Strick, Benjamin, and Elise Thomas. 2019. Investigating Information Operations in West Papua. Bellingcat. https://www.bellingcat.com/wp-content/uploads/2019/10/Investigating_Information_Operations_in_West_Papua.pdf (October 31, 2019).

[4] Elise Thomas, unpublished research.

the anti-extradition protests in Hong Kong[5]. In the case of the Hong Kong protests, social media actions have ranged from vitriolic attacks on Twitter,[6] to targeted harassment of key protest organisers including posting their personal details online to intimidate and deter.[7]

Actors operating on behalf of Middle Eastern states including Iran, the UAE and Saudi Arabia have used influence operations for a variety of purposes: to influence US politics[8], to justify the 2017 blockade of Qatar[9], and to seed divisive stories that seek to sow discord amongst allies[10].

Actions outside of elections causes real harms

Communications networks are increasingly geographically unbounded, creating opportunities for malign actors to exploit this disintermediated reach into Australian audiences. As we note above these actors have different motivations and their activities are persistent and ongoing. Russian meddling in the 2016 US presidential election has increased the focus of democratic governments on the threats posed by foreign interference exploiting social media as an attack surface but propaganda and disinformation have a significant historical legacy.

Extremist groups understand the need for the oxygen of publicity. Social media allows these groups a vehicle for communicating directly with mainstream audiences; to promote their ideas, recruit, finance and mobilise ongoing support. The internet facilitates decentralised networks of followers and supporters. These networks can lie dormant until activated in response to a particular call for mobilisation, operationalising a model of leaderless resistance espoused by US white supremacist Louis Beam.[11]

The perpetrator of the Christchurch shootings distributed a manifesto on 8chan and livestreamed the attack on Facebook. Everything about the choreographed nature of the attack (and even his initial court appearance) was designed to mobilise others. The shooter wanted to become a meme that other extremists might reproduce online and through kinetic violence. His actions appear to have

[5] Uren, Tom, Elise Thomas, and Jacob Wallis. 2019. "Tweeting through the Great Firewall." https://www.aspi.org.au/report/tweeting-through-great-firewall (September 18, 2019).
[6] ibid
[7] Blundy, Rachel, and Esther Chan. 2019. "'Bulletproof' China-Backed Doxxing Site Attacks Hong Kong's Democracy Activists." Hong Kong Free Press HKFP. https://www.hongkongfp.com/2019/11/01/bulletproof-china-backed-doxxing-site-attacks-hong-kongs-democracy-activists/ (March 12, 2020).
[8] Collier, Kevin. 2018. "How Persian Gulf Rivals Turned US Media Into Their Battleground." *BuzzFeed News*. https://www.buzzfeednews.com/article/kevincollier/qatar-uae-iran-trump-leaks-emails-broidy (December 3, 2019).
[9] DeYoung, Karen, and Ellen Nakashima. 2017. "UAE Orchestrated Hacking of Qatari Government Sites, Sparking Regional Upheaval, According to U.S. Intelligence Officials." *Washington Post*. https://www.washingtonpost.com/world/national-security/uae-hacked-qatari-government-sites-sparking-regional-upheaval-according-to-us-intelligence-officials/2017/07/16/00c46e54-698f-11e7-8eb5-cbccc2e7bfbf_story.html (December 9, 2019).
[10] Lim, Gabrielle et al. 2019. "Burned After Reading: Endless Mayfly's Ephemeral Disinformation Campaign." *The Citizen Lab*. https://citizenlab.ca/2019/05/burned-after-reading-endless-mayflys-ephemeral-disinformation-campaign/ (March 12, 2020).
[11] Berger, J. M. 2019. "The Strategy of Violent White Supremacy Is Evolving." *The Atlantic*. https://www.theatlantic.com/ideas/archive/2019/08/the-new-strategy-of-violent-white-supremacy/595648/ (March 13, 2020).

directly inspired at least one copycat attack (the Poway synagogue shooting in the US).[12] His manifesto itself was inspired by extremist ideas that originated in France, circulating online and gaining increasing traction since 2012, seeping across both encrypted messaging and mainstream social media environments.[13]

Islamic State used social media and digital communications as components of its psychological warfare operations in the conflict zones of Syria and Iraq but also exploited the reach of digital communications networks further. Social media provided a vehicle for the group to radicalise, finance and recruit foreign fighters (including from Australia), inspire and mobilise acts of domestic terror in coalition countries, all from a distance. Its use of social media as a component of its media ecosystem rapidly afforded the group a presence on the international stage that other extremist groups have long sought after.[14] The grotesque choreographed violence that featured in its extended media productions, and the threats directed at coalition leaders and populations were a form of psychological warfare that have led governments to rethink their communications strategies for countering violent extremism.[15]

The issues that malign actors use to drive division, to influence and manipulate audiences at scale may not even be overtly political. As hierarchical models of information distribution (from government, from national broadcasters, from mainstream media) are replaced by a proliferation of information flows, trusted networks become increasingly important as sources of reliable content. This creates vulnerabilities in population-level sense-making, for example trusted networks may share anti-vaccination material or fake news. Russia takes advantage of this dynamic, amplifying conspiracy-theory narratives and 'useful idiots' in order to degrade public trust in authoritative sources of information, reducing capacity for consensus decision-making guided by expert-informed, evidence.[16] This is a long-term project to inhibit the effective functioning of deliberative democracy in targeted states. These activities can also have specific foreign policy objectives that diverge from Australian interests, particularly around the ongoing strength of the NATO alliance,[17] the downing of

[12] Ebner, Julia. 2019. "How Do We Beat 8chan and Other Far-Right Sites? The Same Way We Beat Isis | Julia Ebner." *The Guardian*. https://www.theguardian.com/commentisfree/2019/aug/07/8chan-far-right-sites-white-supremacists-governments (March 13, 2020).

[13] Davey, Jacob, and Julia Ebner. 2019. "'The Great Replacement': The Violent Consequences of Mainstreamed Extremism." *ISD*. https://www.isdglobal.org/isd-publications/the-great-replacement-the-violent-consequences-of-mainstreamed-extremism/ (March 13, 2020).

[14] Winter, Charlie. 2018. "Apocalypse, Later: A Longitudinal Study of the Islamic State Brand." *Critical Studies in Media Communication* 35(1): 103–21.

[15] Votel, General Joseph L. et al. 2017. "#Virtual Caliphate." https://www.cnas.org/publications/reports/virtual-caliphate (March 13, 2020).

[16] Kirk, Katherine. 2019. "How Russia Sows Confusion in the U.S. Vaccine Debate." *Foreign Policy*. https://foreignpolicy.com/2019/04/09/in-the-united-states-russian-trolls-are-peddling-measles-disinformation-on-twitter/ (March 13, 2020).

[17] Galeotti, Dr. Mark. 2019. "Russian Intelligence Operations Shifting Tactics Not Goals." *NATO Review*. https://www.nato.int/docu/review/articles/2019/04/26/russian-intelligence-operations-shifting-tactics-not-goals/index.html (March 13, 2020).

MH17,[18] the credibility of international governance bodies such as the Organisation for the Prohibition of Chemical Weapons[19] and the World Anti-Doping Agency,[20] events in the middle east[21] where Australian Defence Force personnel are deployed.

China similarly is leaning into Western social media platforms in order to shape the information environment. It's ambassadors, embassies, state media but also state-owned enterprises exploit the affordances of Western social media platforms - Twitter in particular - to which the Chinese population does not have access as a result of direct censorship.[22] This allows the Chinese Communist Party (CCP) to proactively shape the information environment in the West, whilst tightly controlling that of their own population. The CCP's censorship of coronavirus content on Chinese-language social media platforms may have limited the population's capacity for disease prevention[23] yet the reach of CCP officials and state media on Western platforms allows the CCP to shape the narrative around its response to the outbreak in ways that favour its model of political power[24], obfuscate the origins of the virus[25] and critique the responses of other governments.[26]

In traditional media markets news producers could afford to invest in high-quality journalism because their monopoly or oligopoly position in local advertising markets allowed them to collect 'rivers of

[18] Hawley, Samantha. 2020. "Russia Engaged in 'textbook' Disinformation Campaign, MH17 Trial Told." *ABC News*. https://www.abc.net.au/news/2020-03-11/mh17-trial-told-of-russian-disinformation-to-hack-investigation/12044384 (March 13, 2020).

[19] Andriukaitis, Lukas et al. 2018. *Breaking Ghouta*. http://www.publications.atlanticcouncil.org/breakingghouta/wp-content/uploads/2018/09/20180924_breakingghouta_web.pdf (March 13, 2020).

[20] "WADA Decision against Russia Is an Unfounded Punishment." 2019. *EU vs DISINFORMATION*. https://euvsdisinfo.eu/report/wada-decision-against-russia-are-unfounded-punishments/ (March 13, 2020).

[21] Czuperski, Maksymilian et al. 2016. *Distract, Deceive, Destroy: Putin at War in Syria*. http://publications.atlanticcouncil.org/distract-deceive-destroy/assets/download/ddd-report.pdf (March 13, 2020).

[22] "China Finds a Use Abroad for Twitter, a Medium It Fears at Home." *The Economist*. https://www.economist.com/china/2020/02/20/china-finds-a-use-abroad-for-twitter-a-medium-it-fears-at-home (March 13, 2020).

[23] Ruan, Lotus, Jeffrey Knockel, and Masashi Crete-Nishihata. 2020. "Censored Contagion: How Information on the Coronavirus Is Managed on Chinese Social Media." *The Citizen Lab*. https://citizenlab.ca/2020/03/censored-contagion-how-information-on-the-coronavirus-is-managed-on-chinese-social-media/ (March 13, 2020).

[24] Allen-Ebrahimian, Bethany. 2020. "Beijing's Coronavirus Propaganda Blitz Goes Global." *Axios*. https://www.axios.com/beijings-coronavirus-propaganda-blitz-goes-global-f2bc610c-e83f-4890-9ff8-f49521ad6a14.html (March 13, 2020).

[25] "Lijian Zhao 赵立坚 on Twitter: '2/2 CDC Was Caught on the Spot. When Did Patient Zero Begin in US? How Many People Are Infected? What Are the Names of the Hospitals? It Might Be US Army Who Brought the Epidemic to Wuhan. Be Transparent! Make Public Your Data! US Owe Us an Explanation! Https://T.Co/VYNZRFPWo3' / Twitter." 2020. *Twitter*. https://twitter.com/zlj517/status/1238111898828066823 (March 13, 2020).

[26] "Hua Chunying 华春莹 on Twitter: '@CDCDirector Dr. Robert Redfield: Some Cases That Were Previously Diagnosed as Flu in the US Were Actually #COVID19. It Is Absolutely WRONG and INAPPROPRIATE to Call This the Chinese Coronavirus. Https://T.Co/Mk4RB7XYq0' / Twitter." 2020. *Twitter*. https://twitter.com/SpokespersonCHN/status/1238003509510856704 (March 13, 2020).

gold'.[27] The business was two-sided in that journalism, a cost centre, attracted an audience, and it was that audience to which lucrative advertising was sold.

But social media companies have a different business model. Social media companies are not exchanging quality content for audience and rely instead on user generated content to attract audiences for advertising. This has resulted in changed incentives for 'news' and content producers. Online, financial incentives are linked to audience size—views, eyeballs, or clicks—and sensationalist and provocative content gathers more engagement,[28] so content producers are *de facto* encouraged to produce sensationalist content, not necessarily high-quality journalism or even journalism of any sort.

The governance models and ethics that previously applied to traditional journalism have been replaced on social media; absent restraining forces, the default profit-maximising behaviour for social media platforms is to allow sensationalist, provocative content. In this social media ecosystem foreign interference and malign actors can flourish.

But despite this change in underlying incentives, social media platforms can powerfully influence behaviour in many ways including: terms and conditions, content moderation policies, algorithms that limit increase the exposure of any individual content, and adding or removing 'friction' to online actions such as on-sharing content.

Facebook have started to adjust policies and algorithms to discourage provocative content,[29] but effective and transparent content policies are a societal issue, not an entirely voluntary issue that should be left to the whims of the management of social media platforms.

Policy responses

Social media companies and governments have diverging interests. Government and civil society need to proactively engage to remove the space for malign actors to thrive in the social media ecosystem. We recommend policy responses that fall into these three categories:

1. Transparency
2. Oversight
3. Public awareness

Transparency

Social media companies should be required to make their content moderation policies and enforcement actions transparent. This would include publishing their content moderation guidelines and regular transparency reports that describe the behaviours and harms they see on their

[27] Clark, Andrew. 2018. "The End for Fairfax Began a Decade Ago." *Australian Financial Review*. https://www.afr.com/companies/media-and-marketing/the-end-for-fairfax-began-a-decade-ago-20180726-h136m8 (March 12, 2020).

[28] Zuckerberg, Mark. 2018. "A Blueprint for Content Governance and Enforcement | Facebook." https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/ (March 12, 2020).

[29] Ibid.

platform and the enforcement and content moderation actions they have taken. These reports should encompass all forms of harms and the responses taken.

Oversight

Additionally, **we suggest an independent Statutory authority** that is empowered to observe and report on how the incentives, policies, algorithms and enforcement actions of social media platforms are operating, with the ultimate goal being to maximise benefits and reduce harm for society and its citizens. This authority would be granted explicit insight into how content is filtered, blocked, amplified or supressed, both from a moderation and algorithmic amplification point of view.

Crucially, these obligations should be placed on *all* social media operating in Australia, including those companies that originate from authoritarian regimes and those fringe platforms servicing niche communities—not just the dominant Western platforms such as Facebook, Twitter, Instagram and Snapchat.

These transparency and oversight measures would go some way towards countering the default incentive towards sensational, provocative and potentially polarising content.

Public awareness

Beyond the incentives and policies of social media platforms, government and civil society need to focus on the groups that seek to damage and harm liberal democracies and their citizens. Adversaries are seizing the asymmetric advantage that the absence of organised resistance and deterrence affords.

Foreign interference is a national security problem where every possible weak point in society, both online and offline, may be attacked to weaken society and liberal democracy. Although social media is an attractive and cost-effective means of achieving influence, foreign actors operate across the entire information environment and will conduct co-ordinated influence operations across many platforms simultaneously.

Focussing narrowly on altering the incentives and behaviour of individual social media companies, therefore, misses the bigger picture of how malicious actors operate. Society also needs more transparency and information about how these malicious actors behave, their tactics, techniques, and how they conduct influence operations and undermine and exploit societal fractures.

We suggest the funding of independent civil society that can provide the in-depth publicly-accessible research and tools to discover, track and make transparent—and therefore deter—malign influence operations. These malign operations aim to alter public opinion and the public's awareness of how they are being manipulated is a key element of resilience that only civil-society bodies can credibly deliver.

In the Australian context, government agencies are also appropriately reluctant to perform this function as they want to avoid the perception of government manipulation of the information environment. Independent bodies would allow better engagement with social media companies; they are typically reluctant to engage directly with government agencies that research threat actors

because they are concerned that many governments would seek to interfere in social media to their own advantage or to the detriment of their citizens' human rights.

Conclusion

This submission outlines the threat from various forms of social media interference and makes concrete suggestions for policy responses. As this is a persistent ongoing and diverse threat, we recommend rapid implementation to prevent a significant disruptive event such as a manipulated election.

Dr Jake Wallis and Thomas Uren

Australian Strategic Policy Institute

13 March 2020