PRIMER ON DISINFORMATION

Ian Brown, July 2022

KEY DEFINITIONS AND FRAMING

There is a broad consensus among researchers on definitions of *misinformation* as inaccurate information spread unwittingly, and *disinformation* as inaccurate information spread knowingly, usually for political or financial gain – although some authors argue it is often too difficult for the receiver to determine motivation of the sender to make this distinction meaningful, and that motivation can change along the information distribution chain (Bontcheva and Posetti 2020, pp.24–26).

Where these terms are intended to be used in legislation, they must be defined much more carefully, considering issues of "factual nature, harm, intent, profit and dissemination" (O' Fathaigh, Helberger and Appelman 2021). The UN special rapporteur on freedom of expression declared disinformation is an "extraordinarily elusive concept to define in law" which potentially gives states "excessive discretion to determine what is disinformation, what is a mistake, what is truth" (2020, p.13). In the broader legal context, Marsden, Brown and Veale (2021, p.214) emphasised the importance of "electoral, media and communications and data protection laws. Moreover, disinformation regulation must be responsive to the international human rights law standards of legality, necessity, and proportionality."

Bots are automated systems, often used to share information on social media in large volumes, while *trolls* are human beings sharing information under false pretexts/identities, often paid by governments or political parties. Pomerantsev (2019) notes both have been used to attack independent media (e.g. news agency Rappler in the Philippines) and activists (e.g. by drug trafficking gangs in Mexico).

Benkler, Faris and Roberts (2018, p.29) add *propaganda* as "communication designed to manipulate a target population by affecting its beliefs, attitudes, or preferences in order to obtain behaviour compliant with political goals of the propagandist" – linking disinformation to a century of research on this subject, going back to the emergence of mass media and industrialisation in the late 19th/early 20th century. They found US politics "operating in a propaganda-rich environment and…network propaganda is a much deeper threat to democracy than any out-of-human-control emergent socio-technical process" (p.38) (such as the interaction of social media recommendation algorithms with user reading and sharing behaviour).

Prof. Rasmus Kleis Nielsen has commented: "arguably [the] central disinformation problem we face remains the combination of the age-old one – powerful people who lie – with amplification and advertising from platforms who would really, really like to not be the arbiters of political speech." For example, the *New York Times* found: "As part of their attempt to interfere with the 2020 election, Russians are grabbing screenshots of President Trump's tweets, or quoting his own misleading statements… mostly about the dangers of mail-in ballots" (Sanger and Kanno-Youngs 2020). And Brazilian president Jair Bolsonaro is apparently following a similar strategy to Trump, pre-emptively voicing doubts about the security of Brazil's voting machines before an October 2022 election he is expected to lose (Brito 2021).

Benkler et al. also note similarities between a strategy of *disorientation* practiced by the Russian state and elements of US right-wing media, where repeated reporting of fantastical events (such as the shooting down of Malaysia Airlines Flight 17 over eastern Ukraine by the Ukrainian air force on US instructions, or Hillary Clinton trafficking in Haitian children for the benefit of her husband) leave listeners disconnected from "the core institutionalised mechanisms for defining truth in modernity" (2018, p.36).

Benkler (2019) links this disconnection to much broader structural factors since the 1980s: "Throughout the neoliberal period, elite consensus implemented policies and propagated narratives that underwrote and legitimated the rise of a small oligarchic elite at the expense of delivering economic insecurity to the many. These material drivers of justified distrust were compounded by profound changes in political culture that drove movements, on both the right and the left, to reject the authority structures of mid-twentieth-century high modernism." Similarly, Bennett and Livingston (2018, p.122) concluded:

> *The spread of disinformation can be traced to growing legitimacy problems in many democracies. Declining citizen confidence in institutions undermines the credibility of official information in the news and opens publics to alternative information sources. Those sources are often associated with both nationalist (primarily radical right) and foreign (commonly Russian) strategies to undermine institutional legitimacy and destabilize centre parties, governments and elections. The Brexit campaign in the United Kingdom and the election of Donald Trump in the United States are among the most prominent examples of disinformation campaigns intended to disrupt normal democratic order, but many other nations display signs of disinformation and democratic disruption.*

KEY AREAS OF DISINFORMATION RESEARCH AND KEY INSIGHTS FROM THAT RESEARCH

This section focuses on key problems created by disinformation for vibrant, inclusive democracies, as well as describing evidence relating to interventions that work well, or less so.

Much of the earlier research into online disinformation focused on observation of the online activities of actors attempting to benefit commercially or politically from the spread of false information – such as the notorious Macedonian "clickbait farms" producing false news articles targeted at US readers, attracting advertising revenue from companies such as Google; or the Kremlin's "troll factory" sharing false social media articles in an attempt to interfere in the US 2016 presidential election.

However, such studies frequently make strong claims for societal impacts without convincing evidence, while "[e]vidence across social science indicates that average effects of persuasive messages are small" (Coppock, Hill, and Vavreck 2020). As Karpf (2019) notes: "Generating social media interactions is easy; mobilizing activists and persuading voters is hard."

Even the effect on individuals' factual beliefs is questionable, with one survey study for the European Parliament finding "the [Central European and Italian undergraduate] respondents in more or less all countries have shown resistance to falsehood in scientific communication, casting doubt over false news headlines" (Csepeli et al. 2020). Noting that ease of access has attracted many researchers to analyse Twitter compared to other less open social media services, Benkler et al. (2018, p.384) caution: "It is critical not to confound what is easy to measure (Twitter) with what is significantly effective in shaping beliefs and politically actionable knowledge in society."

Higher-quality research looks for evidence of the societal impact of disinformation, and considers individuals' media consumption as a whole. A broad review of media and communications studies research concluded: "Rigorous empirical work today addresses the interaction and conjunction of multiple devices, channels, and platforms when considering how people access, use, and make sense of their media – rather than studying any medium or use thereof in isolation" (Deuze 2021, p.9). For example, it is important to keep in mind the importance of television in overall US media consumption. Using "a unique multimode dataset that comprises a nationally representative sample of mobile, desktop, and television consumption," Allen et al. (2020) found:

1. "news consumption of any sort is heavily outweighed by other forms of media consumption, comprising at most 14.2% of Americans' daily media diets."

2. "to the extent that Americans do consume news, it is overwhelmingly from television, which accounts for roughly five times as much as news consumption as online."
3. "fake news comprises only 0.15% of Americans' daily media diet. Our results suggest that the origins of public misinformedness and polarization are more likely to lie in the content of ordinary news or the avoidance of news altogether as they are in overt fakery."

Based on their extensive analysis of media coverage of the 2016 US presidential election, Benkler et al. (2018, p.386) concluded disinformation is not a *technical* problem: "There is no echo chamber or filter-bubble effect that will inexorably take a society with a well-functioning public sphere and turn it into a shambles simply because the internet comes to town. The American online public sphere is a shambles because it was grafted onto a television and radio public sphere that was already deeply broken."

Instead, they argue persuasively that "analysis should focus on the long-term dynamic between institutions, culture and technology" (p.384) and how it "shape[s] technological adoption and diffusion patterns" (p.8). Wasserman (2022, p.14) similarly emphasises the importance of "media, social, political, and geopolitical contexts" and notes social media are "seen as the breeding ground for disinformation, racism, xenophobia, and misogyny, a situation which is especially virulent in postcolonial contexts marked by social polarization, ethnic tensions, and economic inequality" (p.11).

The outcomes of the 2016 UK Brexit referendum and US presidential elections are often pointed to as strongly influenced by online disinformation tactics. But Bastos and Mercea (2018) found "there was no evidence supporting the notion that bots had substantively altered the Brexit debate on Twitter. Indeed, by all measures employed in our study, the activity of bot-like accounts was relatively minor with respect to the larger conversation about the referendum. Our findings indicated that bots can potentially amplify a subset of accounts, but that their influence in the network is limited and falls short of a megaphone, a result consistent with literature on cognitive dissonance reporting that political persuasion seems to have little effect in attitude change at the individual level". On state interference, the former head of the UK's foreign intelligence service told the *Financial Times*: "The Russians did not create the things that divide us — we did that. They are adept, albeit in a rather crass manner, at exacerbating those things and I believe that we should prevent that" (Khalaf 2020). Meanwhile, Rid (2020) concluded:

It is unlikely that the trolls convinced many, if any, American voters to change their minds [in 2016]... On Twitter, the IRA's impact practically vanished in the staggering number of election-related tweets... The St. Petersburg troll den generated less than 0.05 percent of all election-related posts. The IRA, according to the data released by Twitter, boosted candidate Donald Trump's retweet count with only 860 direct retweets over the entire campaign.

PROMISING INTERVENTIONS

To be most effective, interventions should be context-specific and clearly situated in broad frameworks addressing Benkler et al.'s "institutions, culture and technology" (2018, p.384) across all popular media. Wasserman (2022, pp.13–14) notes: "because information disorder is a multi-levelled, complex phenomenon, responses to the problem also have to be varied, intersectional, and holistic... we understand information disorder not as a universally homogenous phenomenon, but as a contextual one which requires contextually relevant responses."

One good example is Humprecht et al.'s (2020) theoretical framework and indicators for national resilience to disinformation, assessing the following measures as limiting resilience:

1. Political environment – polarisation and populist communication;
2. Media environment – low trust in news, weak public service media, and more fragmented and less overlapping audiences; and
3. Economic environment – large advertising market size and high social media use.

Their cluster analysis found three country groups: high resilience, including northern and western Europe (such as Finland, Denmark, and the Netherlands) and Canada, and two low resilience groups, polarised southern European countries (e.g. Spain, Italy and Greece), and the US:
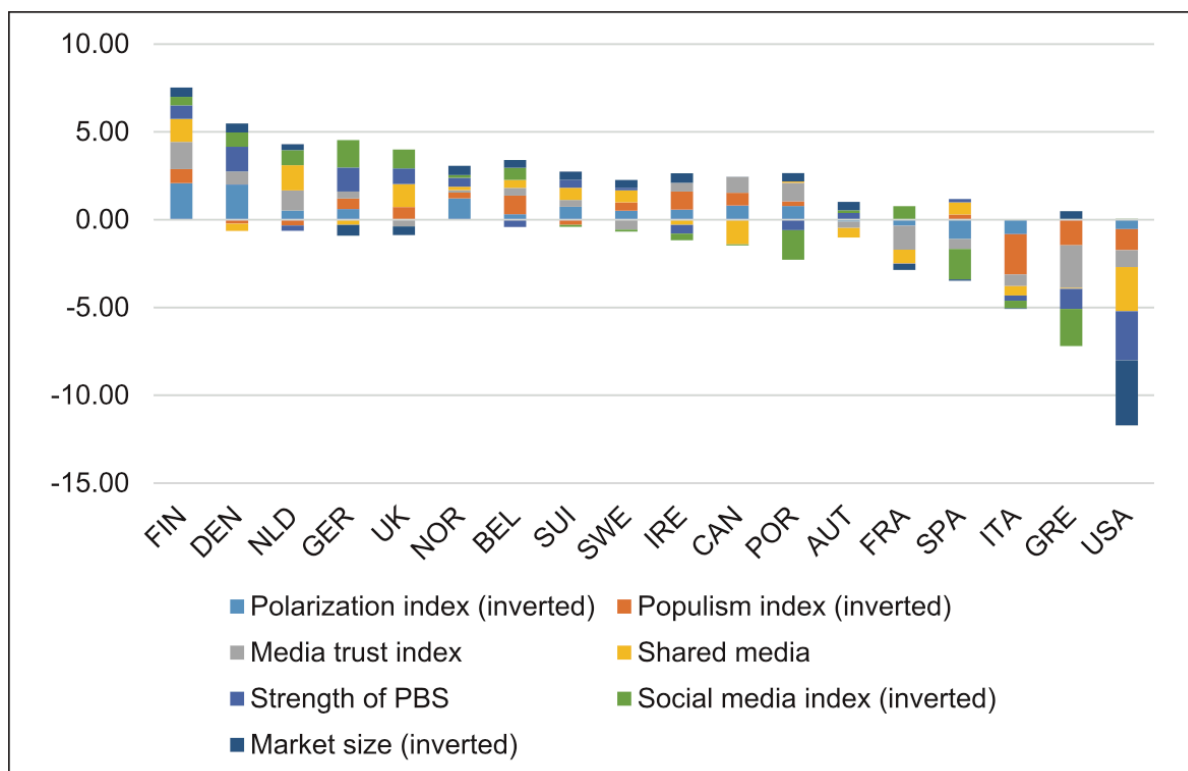

Figura 1. Humprecht et al.'s measure of national resilience to disinformation (2020, p.505)

An example of a potential specific intervention within this framework would aim to increase public trust in media. In the US, Benkler et al. (2018, pp.356-359) identified a key audience segment as individuals consuming content from both an insular right-wing media ecosystem including Fox News and talk radio filled with false information, and a centre/progressive media ecosystem which was much better at self-correcting false information. They suggested this audience should be more open to "credible reporting, in organizations committed to journalistic norms, but with a heavier emphasis on verifiable and accountable truth and credibility rather than balance and neutrality." Their hypothesis is that a "balanced" approach to US politics which presents "both sides" of an issue without attention to the validity of each side is easily subverted with disinformation, as has previously occurred in areas of overwhelming scientific consensus such as the causes of climate change or lung cancer. An intervention by a journalism school or professional association could include training for journalists on taking such a "verifiable and accountable" approach.

Of course, the global pandemic has "further eroded the revenue base of many news outlets around the world, with subsequent job losses and closures of news outlets often disproportionally impacting on smaller independent and community newspapers… [and] also brought a resurgence of censorship and control by governments in the Global South, which have used disinformation around the pandemic as a smokescreen to intensify their repressive tactics" (Wasserman 2022, p.11).

Other specific interventions would address the problems identified in the "Main Challenges" section below, to improve the quality of evidence underpinning potential interventions. In particular, they would push for legal requirements for research access to platform data while protecting user privacy, which has already been agreed in the European Union's new [Digital Services Act](#) (DSA). They would also introduce transparency requirements for political advertising, including disclosure requirements and a public database of political issue adverts (included more broadly in the DSA), which Benkler et al. (2018) identified as important given platforms are likely to continue developing techniques to "increase their effectiveness at manipulating the preferences of their targets" (p.368).

Public and political pressure following the 2016 US elections led social media companies to develop their own transparency frameworks to report actions they had taken to identify and remove and deprioritise disinformation. However, François and douek concluded these firms were "not continuing to invest in their transparency regimes, and the early promise and momentum behind the creation of these pockets of transparency are being lost as public and regulatory focus turns to other areas of content moderation" (2021, p.1).

More broadly, the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression called in a review for "multidimensional and multi-stakeholder responses that are well grounded in the international human rights framework and urge[d] companies to review their business model and States to recalibrate their responses to disinformation, enhancing the role of free, independent and diverse media, investing in media and digital literacy, empowering individuals and rebuilding public trust" (2021, p.1).

MORE QUESTIONABLE INTERVENTIONS

Benkler et al. (2018) are less positive about frequently-suggested interventions that so far show limited evidence of effectiveness in the US. These include suggestions platforms should develop machine-learning systems that can better identify and deprioritise disinformation (p.366); more institutionalised fact-checking, which is seemingly ignored by many users of right-wing media sites as systemically biased, and has limited effect on other audiences (p.377); and media literacy education, with little evidence it "translates into adoption of critical viewing and listening when students consume media in the real world", and potentially even has "trained media consumers to be distrustful of all media and in a perverse way less discerning about what is credible" (p.378). Wasserman (2022, pp.14) adds: "while media users in Africa often suspect or know that information is false, they share it anyway because of sociocultural reasons such as a community orientation…or satire and humor".

Similarly, Wasserman (2022, p.12) notes "While there has been a steep rise in the number of journalistic fact-checking endeavors across the [Global South], such initiatives can only respond to a fraction of publicly circulating mistruths, and verified corrections reach only a tiny proportion of the audience that would have seen the disinformation in the first place" — while Facebook's "fact checking activities are hugely skewed towards English-language content in the Global North" (p.14) and more broadly, "fact-checking organizations that receive support from Western donors often have to contend with perceptions of a lack of independence" (p.16).

Ó Fathaigh et al. (2021) also caution against legal prohibitions on disinformation — both the difficulties in drawing firm boundaries to identify such material, and relatedly the chilling effect it can have on individual freedom of expression. These authors found 11 EU member states have such requirements, in some cases even criminal law provisions carrying potential jail sentences. Wasserman (2022, p.17) found: "In most Arab countries, accusations of 'fake news' or 'disinformation' are used as a pretext to suppress freedom of expression and silence government critics" — and there are related problems in other Global South regions. The UN Human Rights Committee has found the criminalisation of the publication of "false news" is a "clear violation" of the right to freedom of expression (1999, para. 24).

MAIN CHALLENGES WITH CURRENT RESEARCH

DATA ACCESS

Protracted attempts to make the data needed to fully assess the societal impact of disinformation and measures to tackle it available from social media platforms to independent researchers – such as Facebook's Social Science One – have been held up by disputes over user privacy and other concerns (Hegelich 2020). But legislation mandating such access is now being agreed, notably the EU's Digital Services Act.

Benkler et al. (2018, p.384) also suggest: "We need broad, publicly accessible databases for all these different [broadcast] media so that we can begin to apply the emerging array of data science techniques to a sufficiently broad and diverse set of media to actually represent how people get their news and how people come to learn about the world and understand it."

METHODOLOGY/LACK OF PEER REVIEW

Many of the relevant studies, while from credible scientists and institutions, are not peer-reviewed (or are publicised based on university press releases, rather than published studies, where other experts can critique full details) – and are sometimes accused of questionable decisions (Wemple 2018) and embarrassing mistakes (Kreil 2019). While they can still be useful, they take longer to build towards public consensus, and hence legitimacy for state and private actions taken in response to this phenomenon – especially important when disinformation has itself become such a controversial topic, and many responses involve intruding upon individuals' freedom of expression and privacy.

In relation to disinformation from governments, few academic researchers are looking beyond the (partially observable) actions of state disinformation actors to consider the (much harder to access) actors themselves (with rare exceptions, such as Rid (2020), which is based on extensive intelligence archive research and interviews). Most use ground truth data from government investigations about actors, alongside limited information published by large platforms such as Facebook, commercial security firms such as Graphika, and non-governmental organisations. Others use media reports, which contain a widely varying range of quality of evidence, such as interviews with former intelligence officers. Government investigations are expert, but not scientifically peer-reviewed; much of the data used is highly sensitive and cannot be published, even in summarised form. And not all intelligence agencies are politically neutral.

UNEVEN GEOGRAPHIC COVERAGE

Much of the disinformation research done to date has been in relation to the Global North, particularly US-based social media, "despite the fact that information disorder in the Global South is a much older problem, often predating the social media era" (Wasserman 2022, p.7).

Wasserman and also Benkler et al. (2018, pp.53/31) strongly make the case that disinformation can only be understood in relation to a specific context — and that the US is potentially unusual: "the present [US] epistemic crisis is not made of technology; it cannot be placed at the feet of the internet, social media, or artificial intelligence. It is a phenomenon rooted in the radicalization of the right wing of American politics and a thirty-year process of media markets rewarding right-wing propagandists… Trump represents the present state of a dynamic system that has been moving Republican politicians, voters, audiences, and media to the right at least since Rush Limbaugh launched this model of mass media propaganda on talk radio in 1988 and became, as the National Review wrote in 1993, 'the Leader of the Opposition.' In that ecosystem, Trump now operates as catalyst in chief."

Further research in different national contexts is therefore essential, such as that contained in Wasserman (2022), Wasserman & Madrid-Morales (2022) and Sinha (2019), to accurately diagnose disinformation problems, and which interventions are likely to be effective. Wasserman (2022, p.21) suggests "research on information disorder in the Global South should take the deep social, media, political, and geopolitical contexts into account, focus on specificity while seeking comparisons, and avoid treating the South as a monolith."

REFERENCES

Allen, Jennifer, Baird Howland, Markus Mobius, David Rothschild & Duncan J. Watts (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances* 6.

Bastos, Marco & Dan Mercea (2018). The public accountability of social platforms: lessons from a study on bots and trolls in the Brexit campaign. *Phil. Transactions of the Royal Society A*, 376(2128).

Benkler, Yochai (2019) Cautionary Notes on Disinformation and the Origins of Distrust. *MediaWell*, 22 October.

Benkler, Yochai, Robert Faris & Hal Roberts (2018). *Network Propaganda*. Oxford University Press.

Bennett, W Lance & Steven Livingstone (2018) The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication* 33(2), pp.122-139.

Bontcheva, Kalina & Julie Posetti (eds. 2020). *Balancing act: countering digital disinformation while respecting freedom of expression*. UNESCO.

Brito, Ricardo (2021) Bolsonaro directly spread disinformation on Brazil's voting system, police report says, *Reuters*, 17 December.

Coppock, Alexander, Seth. J. Hill & Lynn Vavreck (2020). The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments. *Science Advances* 6(36).

Csepeli, György et al. (2020). Disinformation and Science: Report of an investigation into gullibility of false science news in central European countries. European Parliamentary Research Service Scientific Foresight Unit (STOA) PE 656.300, p.I.

Deuze, Mark (2021). "On the 'grand narrative' of media and mass communication theory and research: a review". *Profesional de la información* 30(1).

François, Camille & evelyn douek (2021). The Accidental Origins, Underappreciated Limits, and Enduring Promises of Platform Transparency Reporting about Information Operations. *Journal of Online Trust & Safety*, 1(1).

Hegelich, Simon (2020) Facebook needs to share more with researchers, *Nature* 579(473).

Humprecht, Edda, Frank Esser & Peter Van Allston (2020). Resilience to Online Disinformation: A Framework for Cross-National Comparative Research. *The International Journal of Press/Politics* 25(3) pp. 493-516.

Human Rights Committee (1999). Concluding Observations of the Human Rights Committee: Cameroon. CCPR/C/79/Add.116.g

Karpf, David (2019). On Digital Disinformation and Democratic Myths. *MediaWell*, 10 December.

Khalaf, Roula (2020). Alex Younger: 'The Russians did not create the things that divide us — we did that'. *The Financial Times*, 30 September.

Kreil, Michael (2019). The Army that Never Existed: The Failure of Social Bots Research, OpenFest Conference, Sofia, 2 November.

Marsden, Christopher T., Ian Brown & Michael Veale (2021). Responding to Disinformation: Ten Recommendations for Regulatory Action and Forbearance. In *Regulating Big Tech: Policy Responses to Digital Dominance* (eds. Martin Moore and Damien Tambini), Oxford University Press.

Ó Fathaigh, Ronan, Natali Helberger & Naomi Appelman (2021). The perils of legally defining disinformation. *Internet Policy Review* 10(4).

Pomerantsev, P. (2019) The disinformation age: a revolution in propaganda, *The Guardian*, 27 July.

Rid, T. (2020). *Active Measures: The Secret History of Disinformation and Political Warfare*. Macmillan.

Sanger, David E. & Zolan Kanno-Youngs (2020). The Russian Trolls Have a Simpler Job Today. Quote Trump. *The New York Times*, 22 September.

Sinha, A. (2019). *The Networked Public: How Social Media is Changing Democracy*. Rupa Publications.

UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2020). Disease pandemics and the freedom of opinion and expression. A/HRC/44/49.

UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2021). Report on disinformation. A/HRC/47/25.

Wasserman, H. (ed. 2022). *Meeting the challenges of information disorder in the Global South*. IDRC.

Wasserman, H. & D. Madrid-Morales (eds. 2022). *Disinformation in the Global South*. Wiley-Blackwell.

Erik Wemple (2018). Study bashes Trumpites for promoting 'junk' news. But what's that? *The Washington Post*, 7 February.