**SSRC-ESRC workshop on Trust in Democratic Institutions. March 2022**

**Manipulating Trust: what do we know about the circulation of false and affective information and what should be done about it?**

**Vian Bakir**

**1. Any investigation into the spread of false information online should be mindful of many types of actors and communicative processes at work**.

These processes are philosophical and epistemological (e.g. the rise of relativism); cultural (e.g. the decline of deference and trust in experts and elites; diverse levels of digital literacy); political (e.g. the rise of nationalism, populism and the alt-right; ruling cultures of deception, bullshit and corruption); economic (e.g. increasing competitive pressures on news outlets and a resulting tendency to produce content designed to attract users and generate revenue); and regulatory (e.g. absence of data protection cultures).

*Scholarship examining these broader processes concentrates on the USA and Western Europe, with few studies on regions likely swamped with false information: e.g. Africa, India, China, Russia, the Middle East, and small nations not central to the business model or public relations concerns of dominant digital platforms.*

**2. Alongside these wider contexts, it is important to understand the affordances of digital networked environments.**

We have a good understanding of the digital affordances that encourage the spread of **false information online,** *but platforms’ design choices and algorithms constantly change, so this knowledge needs continuously updating.*

We have a small, if growing, understanding of the social media affordances that encourage **conspiracy theories**. For instance, the attention economy of social media is designed to maximise user attention and affect. Conspiracies online are concentrated within communities who already agree with them. Some US and UK studies show that reliance on social media for news correlates with higher conspiracy beliefson COVID-19, while studies in other countries are less conclusive.

*For false information and conspiracy theories, more studies across more platforms (not just Facebook and Twitter) and countries (not just polarised countries like the USA and Italy), would be helpful.*

**3. Empirical links between false information and trust need more study.**

Considering online disinformation (i.e. deliberate attempts to deceive) and misinformation (i.e. inadvertent spreading of false information), it is *hard to assess its scale* as social media platforms have sole, proprietary access to their data; and algorithms and users shift over time. Nonetheless, big data studies suggest that false information online is prevalent.

Studies show thatfalse information during election campaigns worldwide is spread by bots and humans. Studies from the USA and UK show that false information is shared mainly by politicians, partisan media, partisans and the politically engaged. *More user-based studies across different countries are needed*.

*Why people share false information online has not been studied extensively* but reasons uncovered include partisanship; a desire to spread rumours; a belief that the information is true; ineffectiveness at recognising deception; for fun; reliance on mental short cuts to evaluate information and source credibility online; a mismatch between what people are interested in and what news outlets provide; and congruence with individuals’ pre-existing worldviews.

*Whether this is a newly problematic situation is open to question.* Trust in the news to tell the truth has been under strain for decades worldwide. Also, trust in experts has long been in global decline: as far back as 2005, trust shifted from authorities to peers. Yet, it is a particularly harmful situation in times of crisis (such as pandemic) when governments require large-scale, significant behaviour change to quell the virus. Recent multi-country research on COVID-19 vaccine hesitancy finds that those with poor information hygiene are less likely than those with good information hygiene to say that they would take the vaccine; and lack of vaccine acceptance is associated with conspiratorial thinking, and lack of trust in authorities and scientists.

**4. The role of affect in circulating false information requires further study.**

Many studies point to the often highly negative and positive emotions that social media platforms circulate worldwide. Big data studies repeatedly show that expression of emotion is socially contagious on social media (meaning that a perceiver’s emotions become more similar to others’ emotions as a result of exposure to these emotions), with caveats that such *causality is difficult to prove*. This is evident in studies on Facebook, Twitter and on non-US based social media platforms such as Chinese microblogging site Weibo. Such emotional contagion is not an accident, but the result of social media algorithms that optimise engagement, these exploited by the architects of disinformation. Increasing critical scholarly attention is also being paid to links between emotions, attention and revenue, and how these have been algorithmically optimised to monetise deception.

*We need to know more about different platforms’ proprietary algorithms that enable sorting, recommending and judging of emotional and false content. This includes understanding how false, emotive information is propelled, but also how platforms engage in content moderation over time and across countries. We especially need to know more about algorithmic choices beyond US and UK, on other platforms (beyond Facebook and Twitter) and its impact on important groups, e.g. political parties, conspiracy theories, the persuadables etc.*

**5. We need more studies into why people do not trust different media forms, content, sources***.*

Social media platforms are prime sites for circulating false information and conspiracy theories, propelled by the attention economy. *But we need to know to what extent people buy into these ideas, why, and under what circumstances.* The empirical reality will be messy. Fault lines will include active v. passive users; national differences; demographic differences, with older people most vulnerable; and platform differences. *Not all types of false information will be of equal concern.* Deception, especially by domestic politicians, is more concerning to populations worldwide than poor journalism (factual mistakes, dumbed-down stories, clickbait). We also *do not know enough about why people buy into conspiracy theories*.

1. **We need to address what works in combatting false information****online***.*

Over five years of intensive governance and multi-disciplinary academic interest in tackling false information online has not yet fixed the problem, nor even agreed on how best to do so.

In a forthcoming book (Bakir & McStay 2022, *Optimising Emotions, Incubating Falsehood*, Springer), we assess seven solution areas: namely, government action; cyber-security; digital intermediaries/platforms; advertisers; professional political persuaders/public relations; media organisations; and education. We conclude that each area has an important role in protecting against false information online but faces challenges. Censorship in the form of coercive governmental intervention damages freedom of expression and speech, but lack of governmental action also harms the civic body in many ways given the unrelenting scale, speed and spread of false, emotive information online. The unpreparedness of automation to detect and address all false information online; the lack of transparency and global unevenness in platforms’ content moderation practices; the lack of transparency and ethics in digital political advertising and wider strategic political communications; the unhealthy media ecology dominated by global digital platforms, decreasing trust in journalism, and under-resourced fact-checking and journalism; and the practical, psychological and sociological limits to increasing people’s digital literacy, truly make this a wicked problem.

Yet, leaks from the world’s biggest social media platform (Facebook, now rebranded as Meta) lead us to conclude that the ultimate solution would be to change the business models of social media platforms, so that they do not seek maximal user engagement at all costs, and so that they do not design algorithms that make emotional, deceptive content go viral to flood certain types of users’ social media feeds. However, this is unlikely to happen without either (a) a mass exodus of users (which is unlikely given how strongly imbricated dominant digital platforms are in people’s daily lives); or (b) strong governmental and coordinated inter-governmental intervention to regulate the algorithms that promote emotive, false, viral information (this could be possible, but care would be needed not to sacrifice the benefits of free speech). While most governments express reluctance to stifle innovation in the technology industry, self‐regulation is unsuitable where harms are also indistinguishable from the commercial interests of industry players.