Making the ‘Harm Principle’ Central to Approaches Against Information Disorder

Uyiosa Omoregie*

Avram Turing, Guelph, Canada.

Abstract

Misinformation online is an urgent global challenge. The gravity of the challenge, and its effect on global collective behaviour, has led to calls for social media/information disorder to be designated a “crisis discipline” like medicine, conservation biology and climate science. Scholars have generally settled for a definition of ‘information disorder’ that reveals three variants: misinformation, disinformation and malinformation. What should be of paramount importance, in the fight against information disorder, is the potential of falsehood to cause harm. This potential for harm must be the litmus test distinguishing free speech and speech that should not be free. The ‘harm principle’ proposed by John Stuart Mill is more than 150 years old and needs an upgrade for the social media age. One such upgrade is proposed by Cass Sunstein. We summarize different approaches to analysing online information disorder. We conclude that approaches which emphasize analytical and critical thinking are important but have shortcomings. When analysing complex phenomena like conspiracy theories a ‘systems’ approach is more effective to reveal root causes of information disorder, provide actionable insight and long-term solutions.

Key words: Misinformation, Harm Principle, Systems Thinking, Disinformation

1. Introduction

Misinformation propagation in its current form is a global problem that requires urgent solutions. Historically, instances of misinformation publicly propagated can be found as far back as the sixth century AD. Misinformation was propagated publicly when Procopius, the historian, wrote deliberate falsehoods to tarnish the image of Emperor Justinian [1]. In the history of misinformation propagation, three periods are generally recognised by scholars as turning points: World War II, the Cold War, and the 2016 presidential elections in the United States of America [2] [3] [4]. Scholars began studying propaganda as a concept during World War II [2].

I thank Kirsti Ryall for discussions, comments, insightful feedback and manuscript revision.

*Contact: uyiosa.omoregie@avramturing.com

‘Disinformation’ as a concept is a product of the Cold War, it is derived from the Russian word ‘dezinformatsiya’ [4]. During the Cold War state actors on both sides of the Iron Curtain engaged in misinformation propagation [3]. What is distinctive about present day misinformation propagation is the speed, reach and virality, amplified by online social media. During the 2016 US presidential elections an estimated 126 million Americans were exposed to online content, containing misinformation, that was sponsored by a foreign country [31]. The gravity of the information disorder challenge, and its effect on global collective behaviour, has led to calls for social media/information disorder to be designated a “crisis discipline” like medicine, conservation biology and climate science [5].

Scholars and researchers have generally settled for a definition of ‘information disorder’ that reveals three variants: misinformation, disinformation and malinformation. Misinformation is defined as (false or misleading information *not necessarily* intended to deceive), disinformation (false and misleading information with *intention* to deceive) and malinformation (*authentic* information with intent to cause harm) [6]. Lewandowsky et al. [7] define misinformation as “information initially presented as true that is later found to be false.” Ammara et al. [8] (for convenience’s sake and simplicity) referred to all kinds of information disorder as misinformation. What should be of paramount importance, in the fight against information disorders, is the potential of false information to cause harm. Disinformation and malinformation are propagated with intention to cause some sort of harm. Misinformation is usually propagated without intention to cause harm. However, disinformation can be transformed into misinformation (when deliberate falsehood originally propagated to deceive is received by another person as truth and innocently propagated further).

The potential for harm must be the litmus test distinguishing free speech and speech that should not be free. Wardle and Derakhshan [9] (see Figure 1 below) show that malinformation and disinformation are harmful but misinformation is simply falsehood. Can misinformation be harmful, regardless of no intention to cause harm? We think so. People can innocently propagate falsehood, that they sincerely believe to be true, such falsehood can be harmful. A simple example: a sincere belief that children cannot be electrocuted, if propagated (innocently), could cause harm. Aral [30] explains that at the beginning of the COVID-19 pandemic, he rallied his team to work at the intersection of the pandemic and social media technology, with a commitment to ‘no harm’ at the centre of his strategy:

As director of the MIT Initiative on the Digital Economy, one of the largest centres researching the influence of digital technologies on our world, and head of MIT’s Social Analytics Lab, a team of thirty brilliant MIT faculty and students studying the impact of social media on society, I felt we needed “all hands on deck” to address the crisis. . . While stressing the need to first do no harm, I reached out to my contacts at social platforms around the world to see how we could help [30].

In the next section we summarize the ‘harm principle’ as proposed by John Stuart Mill in 1859. We argue that an update of the ‘harm principle’ is required for the social media age – such an update is provided by Sunstein. We then highlight some of the different approaches used to analyse and combat information disorder online. We conclude by suggesting the ‘systems approach’ as the best approach for analysing complex issues of information disorder.

Figure 1: False and Harmful Information [9]

2. Free speech and the ‘harm principle’

Free speech is the grey area found somewhere between the pillars of freedom and responsibility. It is recognized as one of the most important tenets of any democratic civil society. Nothing else is more important in a democracy after the establishment of the rule of law. There are, of course, obvious limits to free speech. One cannot shout ‘fire!’ on board a plane or during a crowded session of parliament, for example. Free speech should not cover that type of deliberate recklessness.

How do we draw the limits of free speech? *Can* a government legislate into law what is free speech and what is not? A government, if it has a large enough majority, can muster enough votes in support of such legislation. More pertinently, *should* a government legislate into law what is free speech and what is not? Neither governments nor social media platforms should possess such power. What governments and social media platforms can do is follow universal principles regarding free speech and its limits. One universal principle is the ‘harm principle’ proposed by the British philosopher John Stuart Mill [10]. Mill proposed that no government or person should interfere and prevent another person’s liberty to do or say anything. According to Mill, the only reason that permits interference with another person’s liberty is to prevent harm to others.

The ‘harm principle’ is more than 150 years old and needs an upgrade for the social media age. One such upgrade is proposed by Sunstein [11]. Sunstein sides with Mill and believes that free speech should be protected as much as possible for many reasons. Sunstein focuses on

misinformation (falsehoods). He believes that both truth and falsehood play important roles in society: when falsehoods (part of free speech) are censored there would be a growing tendency for government to censor any speech it does not like. Also, if falsehood is censored, then paradoxically, truth would be threatened to some extent because many people would fear being censored for saying what they believe is true. Allowing falsehood may sometimes lead to the discovery of new truths eventually. If you want to know what people really think and not what they pretend to think, you must make allowance for falsehood. According to Sunstein the best response to falsehoods is usually to correct them rather than to punish or censor them. Sometimes punishment or censorship actually fuel more falsehoods.

There are limitations to the kind of falsehoods a society should tolerate. Sunstein presents a framework that builds on Mill's 'harm principle.' He is careful to state that "...many false statements are not lies; people who make or spread them sincerely believe them to be true. Falsehoods are a broad category of which lies are a mere part. Some people say what they know to be false. Others are reckless; it should be obvious that they are spouting falsehoods, but they do not know it is what they are doing. Still other people are simply mistaken; they had reason to say what they did, but they turned out to be wrong. These differences matter." These nuances inform his framework (see Table 1 below). The decision to censor or punish falsehood or regulate it must distinguish between four types of peddlers of falsehood or 'state of mind' of the speakers: the liar, the reckless, the negligent and the reasonable. In terms of harm: the magnitude of harm, the likelihood of harm and the timing of harm are all important considerations. How much damage can the particular falsehood cause (grave or moderate or minor or non-existent)? How likely is it that the falsehood will actually cause harm (for certain or probable or improbable or highly probable)? Is the harm from the falsehood immediate, imminent but immediate or likely to happen in the distant future?

Table 1: Spectrum of harm effects of different types of falsehood [11].

State of Mind	Magnitude of Harm	Likelihood of Harm	Timing of Harm
Lie	Grave	Certain	Imminent
Reckless	Moderate	Probable	Near Future
Negligent	Minor	Improbable	Reasonably Soon
Reasonable	Non-existent	Highly Probable	Distant Future

3. Approaches against information disorder

The potential for harm should inform the type of intervention into combating information disorder. Disinformation, deliberate production and propagation of falsehood or harm online, can be countered with legislation or strict code of conduct rules: the UK Online Safety Bill and the European Commission Code of Practice on Disinformation are good examples [12][13]. The U.S. Surgeon General issued a health warning about misinformation on social media and the harm it causes to health, a warning similar the past campaigns against tobacco smoking [29].

Zhang et al. [14] emphasize the multifaceted nature of the misinformation challenge and the varied initiatives and solutions proffered against the menace. An ‘infodemiological’ approach to misinformation analysis makes use of network analysis. This approach seeks to understand the spread of misinformation the way an epidemiologist studies disease [15]. Some analysts use a cybersecurity approach: viewing online misinformation as a national security challenge. Equipped with cybersecurity tools, they track misinformation the way they would track malware, to prevent “the hacking of people’s beliefs” [16][17]. Popat et al. [18] employ a linguistic approach to misinformation analysis by assessing the credibility of emerging claims online. They demonstrate the importance of the language style of textual claims online. Increasingly important is the social justice/power structures approach to online misinformation analysis [2] [4] [19]. This approach emphasizes the role power differentials within society play in certain kinds of disinformation that amplify racial stereotypes and increase xenophobia, misogyny and anti-Semitism. O’Connor and Weatherall employ a social network approach to misinformation analysis arguing that misinformation persists because of social factors that inform belief (as opposed to individual psychology) [32].

A Wittgensteinian approach to misinformation analysis [20] [21] presents a model inspired by the framework of the *early* philosophy of Ludwig Wittgenstein. This approach is based on ‘analysis’ or analytical thinking, which could be differentiated from ‘criticism’ or critical thinking. Analysis involves decomposition of information to simpler elements for clarity and better understanding, the opposite of ‘synthesis’ [22]. Critical thinking is more focused on the evaluation of information, interpreting it and making an informed judgment [23]. Analytical and critical approaches complement each other. Misinformation online manifests in different forms, a combination of different approaches to tackle it is required for effective solutions. Cook et al. [24] present a critical thinking approach to deconstruct misinformation. They identify reasoning fallacies, based on argument structure, to refute common misinformation about climate change espoused by denialists. This approach is comprehensive and was successful when applied to 42 claims commonly peddled by denialists.

Fact-checking as a misinformation prevention strategy has limitations [25]. Certain types of misinformation disorder are complex and providing ‘facts’ alone may not change beliefs in adherents, particularly when deeply-held beliefs are involved. Sometimes the opposite effect of strengthening the false belief occurs. Breaking down arguments into ‘logical atoms’ [20] a form of reductionism will not be effective for complex phenomena like conspiracy theories. Analysing conspiracy theories is best done using a systems approach or systems thinking [8]. Systems thinking is a way of viewing the world in all its complexity: a holistic (as opposed to

reductionist) worldview that emphasises learning and the search for high leverage points to facilitate lasting positive change [26] [27]. The systems approach is not a linear approach or “open-loop, event-oriented view of the world” (see Figure 2 below) but rather a “feedback view of the world” (see Figure 3 below) [26].

Figure 2: The open-loop, event-oriented view of the world [26]

Figure 3: The feedback view of the world [26]

The Walters Foundation [28] describes the following “Habits of Systems Thinkers”: 1. Seek to understand the big picture; 2. Observe how elements within the system change over time, generating patterns and trends; 3. Recognize that a system’s structure (elements and interactions) generates behaviour; 4. Identify the circular nature of complex cause-and-effect relationships; 5. Surface and test assumptions; 6. Change perspective to increase understanding; 7. Consider an issue fully and resist the urge to come to a quick conclusion; 8.

Consider how mental models affect current reality and the future; 9. Use understanding of system structure to identify possible leverage actions; 10. Find where unintended consequences emerge; 11. Recognize the impact of time delays when exploring cause-and-effect relationships; 12. Check results and change actions if needed: “successive approximation.”

Ammara et al. [8] use a systems approach to analyse misinformation: an emphasis on identifying roots causes, provide actionable insight and anticipating long-term consequences. They reveal that incorrect labelling of news by fact-checkers could lead to more scepticism of news credibility as a whole. They also discover that a focus on improvements in artificial intelligence (AI) for *detection* of fake news could lead to similar technology being used to *generate* more fake news.

4. Conclusion

We argue that a viable definition of information disorder, and approaches to counter this global scourge, must place at the centre the potential for harm caused by falsehood. Information disorder is multifaceted and effective intervention must consist of varied complementary approaches. Fact checking, analysis or breaking down of arguments and censorship approaches are important but have limitations. For long-term solutions, a systems approach is required that looks to reveal root causes of information disorder. Upholding freedom of speech online is crucial and clearly defining the potential for harm from falsehood is the other side of this coin.

References

- [1] Darnton, R. The true history of fake news, *The New York Review of Books*, <https://www.nybooks.com/daily/2017/02/13/the-true-history-of-fake-news/> (2017, accessed 28 April 2021).
- [2] Kreiss, D. “Social Media and Democracy: The State of the Field, Prospects for Reform,” edited by Nathaniel Persily and Joshua A. Tucker. *Int. J. Press/Politics*. 2021; 26(2):505-512.
- [3] Freedland, J. Disinformed to death. *The New York Review of Books*, <https://www.nybooks.com/articles/2020/08/20/fake-news-disinformed-to-death/> (2020, accessed 22 January 2021).
- [4] Marwick, A, Kuo, R, Cameron, S and Weigel, M. Critical disinformation studies: a syllabus. Center for Information, Technology, & Public Life (CITAP), University of North Carolina at Chapel Hill, <http://citap.unc.edu/critical-disinfo> (2021, accessed 28 April 2021).
- [5] Bak-Coleman, JB, Alfano, M, Barfuss, W, Bergstrom, CT, Centeno, MA, Couzin, ID, Donges, JF, Galesic, M, Gersick, AS, Jacquet, J, Kao, AB. Stewardship of global collective behavior. *Proc. Natl. Acad. Sci. U.S.A.* 2021; Jul 6;118(27).

- [6] Turcio, L and Obrenovic, M. Misinformation, disinformation, malinformation: causes, trends, and their influence on democracy, https://hk.boell.org/sites/default/files/importedFiles/2020/11/04/200825_E-Paper3_ENG.pdf (2020, accessed 22 January 2021).
- [7] Lewandowsky, S, Ecker, UK, Seifert, CM, Schwartz, N and Cook, J. Misinformation and its correction continued influence and successful debasing. *J. Appl. Res. Mem. Cogn.* 2018; 6: 353-69.
- [8] Ammara, U, Bukari, H and Qadir, J. Analyzing misinformation through the lens of systems thinking. *Proceedings of the 2020 Truth and Trust Online (TTO 2020)*; 55-63.
- [9] Wardle, C, Derkhshan, H. Information disorder: Toward an interdisciplinary framework for research and policymaking. *Council of Europe Report*, <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c> (2017, accessed 10 July 2021).
- [10] Mill, JS. *On Liberty*. London: Penguin Classics, 1859/2010.
- [11] Sunstein, C. *Liars: Falsehood and Free Speech in an Age of Deception*. New York: Oxford University Press, 2021, pp. 3-4.
- [12] GOV.UK. Draft online safety bill, <https://www.gov.uk/government/publications/draft-online-safety-bill> (2021, accessed 13 July 2021).
- [13] European Commission. Commission presents guidance to strengthen the Code of Practice on Disinformation, https://ec.europa.eu/commission/presscorner/detail/en/IP_21_2585 (2021, accessed 10 July 2021).
- [14] Zhang, A, Ranganathan, A, Metz, S, et al. A structured response to misinformation: defining and annotating credibility indicators in news articles. *The 2018 Web Conference Companion*. <https://dl.acm.org/doi/10.1145/3184558.3188731> (2018, accessed 17 July 2021).
- [15] Freedman, D. As QAnon conspiracy theories draw new believers, scientists take aim at misinformation pandemic, <https://www.newsweek.com/2020/10/23/qanon-conspiracy-theories-draw-new-believers-scientists-take-aim-misinformation-pandemic-1538901.html>, (2020, accessed 28 April 2021).
- [16] Kehrt, S. One data scientist's quest to quash misinformation. *Wired*, <https://www.wired.com/story/data-scientist-cybersecurity-tools-quash-misinformation/> (2020, accessed 27 September 2020).
- [17] Schapals, A. Fake news. *J. Pract.* 2018;12, 976-985.
- [18] Popat, K, Mukherjee, S, Strötgen, J, Weikum, G. Where the truth lies: explaining the credibility of emerging claims on the web and social media. In *Proceedings of the 26th International Conference on World Wide Web Companion* 2017 Apr 3 (pp. 1003-1012). <https://dl.acm.org/doi/10.1145/3041021.3055133>, (2017 accessed 28 April 2021).
- [19] Noble, S. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press. 2018.
- [20] Omoregie, U. Online misinformation analysis and information quality theory. *SocArXiv*, <https://osf.io/preprints/socarxiv/82swk/> (2021, accessed 10 July 2021).
- [21] Omoregie, U. Misinformed about the information age: The existential crisis of online social media. *Business Day*, <https://businessday.ng/opinion/article/misinformed-about-the-information-age-the-existential-crisis-of-online-social-media/> (2021, accessed 10 July 2021).

- [22] Beaney, M. Analysis. *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/analysis/> (2014, accessed 28 April 2021).
- [23] The Peak Performance Center. Analytical thinking and critical thinking, <https://thepeakperformancecenter.com/educational-learning/thinking/critical-thinking/analytical-thinking-critical-thinking/> (2021, accessed 28 April 2021)
- [24] Cook, J, Ellerton, P, Kinkead, D. Deconstructing climate misinformation to identify reasoning errors. *Environ. Res. Lett.* 2018; 13(2):024018.
- [25] Cook, J, Lewandowsky, S and Ecker, UK, et al. Neutralizing misinformation through inoculation: exposing misleading argumentation techniques reduces their influence. *Plos One* 2017; 12:5 e0175799.
- [26] Stermann, JD. Systems dynamics: system thinking and modelling for a complex world. *Massachusetts Institute of Technology Engineering Systems Division, Working Paper Series ESD-WP-2003-01.13-ESD Internal Symposium 2002*, <https://dspace.mit.edu/bitstream/handle/1721.1/102741/esd-wp-2003-01.13.pdf?sequence=1&isAllowed=y> (2002, accessed 14 July 2021).
- [27] Omoregie, U. Management thinking for complex issues: review of two contrasting perspectives. *FMR*. 2017; 1:96-105. http://www.isaacpub.org/images/PaperPDF/FMR_100020_2017091216050950335.pdf
- [28] Waters Foundation. Habits of a systems thinker, <https://ttsfilestore.blob.core.windows.net/ttsfiles/habits-single-page-2020.pdf> (2020, accessed 10 July 2021).
- [29] HHS.gov Office of the Surgeon General. Confronting health misinformation: The U.S. Surgeon General’s advisory on building a healthy information environment, <https://www.hhs.gov/sites/default/files/surgeon-general-misinformation-advisory.pdf> (2021, accessed 17 July 2021).
- [30] Aral, S. *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health – and How We Must Adapt*. New York: Currency, 2020, p.xi.
- [31] Ingram, D. Facebook says 126 million American may have seen Russia-linked political posts. *Reuters*, <https://www.reuters.com/article/us-usa-trump-russia-socialmedia-idUSKBN1CZ2OI> (2017, accessed 19 July 2021).
- [32] O’Connor, C. and Weatherall, J. *The Misinformation Age: How False Beliefs Spread*. New Haven: Yale University Press, 2019.

Author Information

Uyiosa Omoregie is the principal analyst at Avram Turing, a research firm focused on misinformation analysis and prevention. He is a member of the Association of Internet Researchers (AoIR) and the Canadian Academy of Independent Scholars (CAIS).