

Prebunking Disinformation – Psychological Vaccine for Herd Immunity

Davit Kutidze¹

Technological advancements, the internet, and social networks have enabled people to access information swiftly and easily. However, along with these opportunities, challenges have arisen. Firstly, there is the weaponization of information by various malicious actors and attempts to distort reality in this manner. Naturally, disinformation and the manipulation of public opinion have been prevalent in every age, although the simplification of the spread of information showcases this problem as much more serious. Currently, hostile nations, instead of openly waging war, opt to influence the opinions of people through information and interference in other countries' domestic affairs. One of the most noteworthy instances is Russia's information warfare, which poses a significant global challenge. It seeks to sow fear and confusion among target communities, thereby undermining evidence-based, subject-matter discussions.

When discussing the aforementioned challenges, it is important to understand that not only invented stories are used to influence opinions and behavior, but we are mostly dealing with the mixing of truth and lies, distorting context, and, therefore, manipulation of information. It is also noteworthy that this type of information is much more dangerous than so-called fake news because, containing elements of truth, it is harder to analyze correctly and identify real goals. According to Roozenbeek and van der Linden (2021), there are a number of principal techniques used under these circumstances to mislead the audience. These techniques include emotionally manipulative language, polarizing language, conspiratorial reasoning, the so-called trolling, or logical fallacies.²

The issue of manipulative information interference is especially worrisome for democratic communities where freedom of speech is a cornerstone. Certain malicious actors, including hostile foreign adversaries, exploit this for their own gain. There are numerous reports³ and investigations⁴ that highlight

¹ Research Institute Gnomon Wise, The University of Georgia, e-mail: d.kutidze@ug.edu.ge

² Roozenbeek, J. and van der Linden, S. (2021). *Inoculation Theory and Misinformation*. NATO Strategic Communications Centre of Excellence. Online ISBN: 978-9934-564-49-9.

³ Martuscelli, C. (2021). EU warns of disinformation campaign linked to Russia's Sputnik vaccine. *Politico*. Via link: <https://politi.co/3Ba2vVV>

⁴ Select Committee on Intelligence United States Senate. Report 116-XX on Russian Active Measures, Campaigns and Interference in the 2016 U.S. Election. Volume 2: Russia's Use of Social Media with Additional Views. Via link: <https://bit.ly/3jyKWZW>

this concern, delving into the attempts of the Russian Federation to influence elections and other political and social processes⁵ in various countries, including the USA. Western nations faced this problem particularly acutely in 2014 during Russia's "hybrid warfare" against Ukraine, as well as in 2016 after the US Presidential Elections. Therefore, in democratic countries, along with relevant academic communities, there have been continuous debates over the years on how to counter manipulated, malignant information.⁶

Simultaneously with the growth of the disinformation burden, so-called Fact-Checking organizations began to emerge in many countries across the world. Their goal is to counter malicious campaigns or actors that aim to manipulate, polarize, and confuse society with fact-based evidence.⁷ Naturally, their attempts are important and valuable, although the question remains whether it is enough to be only responsive to malicious information.

Drawing on different academic studies, Roozenbeek and van der Linden outline four major reasons why manipulated information is one step ahead of verified facts. 1. False rumours can spread further, faster and deeper through social networks than information that was later fact-checked. In addition, fact-checkers are often unlikely to reach the same people as the original misinformation. 2. People who have been exposed to misinformation may continue to rely on it, even if it has been debunked – a phenomenon known as the “continued influence effect”. Therefore, we cannot expect fact-checks to reliably and comprehensively undo the damage done by misinformation exposure. 3. Repeated exposure to disinformation in the information environment contributes to higher likelihood that eventually it will be perceived as truth – a phenomenon known as the “illusory truth effect” 4. There is some evidence that people do not like being fact-checked and therefore, they may respond negatively to debunking attempts.⁸

In addition to the aforementioned challenges, purveyors of manipulated information also enjoy several other advantages. Firstly, it is noteworthy that the invention of myths or conspiracy theories

⁵ Marineau, S. (2020). Fact Check US: What is the impact of Russian interference in the US presidential election? *The Conversation*. Via link: <https://bit.ly/3ORLzqX>

⁶ Note – This paragraph was taken from the following publication: Kutidze, D. (2021). The Modern Dilemma – the Fight Against Disinformation Through Regulation VS Freedom of Speech. Research Institute *Gnomon Wise*. Accessible at: <https://gnomonwise.org/ge/publications/analytics/58>

⁷ Verified signatories of the International Fact-Checking Network (IFCN) code of principles. Via link: <https://ifcncodeofprinciples.poynter.org/signatories>

⁸ Roozenbeek, J. and van der Linden, S. (2021). Inoculation Theory and Misinformation. *NATO Strategic Communications Centre of Excellence*. Online ISBN: 978-9934-564-49-9.

requires less time compared to the time needed to respond to them. Reliable and fact-based information necessitates the analysis and scrutiny of numerous sources. Time, however, is one of the most critical factors in countering disinformation. The longer it takes to correct fake news from the time of its spread, the less likely verified information will undermine the effect of the initial falsehood.⁹ ¹⁰ Rapid dissemination of myths and manipulations is further abetted by the prevalence of "clickbait," which operates without constraints of journalistic or ethical standards. It appears to be "gossiping like a friend," while quality journalism involves more details and formal language, complicating its perception and emotional connectivity with the audience.¹¹

Given all the aforementioned factors, scholars have concluded that merely being responsive when dealing with conspiracy theories or manipulated information is insufficient for the triumph of truth. Therefore, they have devised a strategy that involves pre-emptively debunking— "prebunking"— disinformation. This approach is rooted in the inoculation theory developed in the 1960s, initially conceptualized by the social psychology scientist William McGuire.¹² According to Roozenbeek and van der Linden, during the Vietnam War, the US government was concerned about the potential influence of foreign propaganda on US troops, fearing they could be "brainwashed." McGuire explored the concept of a "psychological vaccine" against "brainwashing." According to his hypothesis, instead of bombarding people with accurate facts, it could be more effective to pre-emptively inoculate them with a certain number of manipulative arguments. This ensures that individuals are familiar with manipulative messages about a specific issue and are more cautious regarding information operations expected in the future. Numerous studies have validated the relevance of this approach.¹³

A meta-analysis¹⁴ conducted by Banas and Rains (2010), which examined 54 different studies, provides solid evidence supporting the effectiveness of the inoculation theory. It was on the basis of such

⁹ Walter, N and Tukachinsky, R. (2019). A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It? *Communication Research*. Via link: <https://journals.sagepub.com/doi/10.1177/0093650219854600>

¹⁰ Tsfaty, Y. Et al. (2020). Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis. *Annals of the International Communication Association*. Via link: <https://www.tandfonline.com/doi/pdf/10.1080/23808985.2020.1759443>

¹¹ Kutidze, D. (2020). Why the Real and Verified News Find It Difficult to Combat Disinformation. Research Institute Gnomon Wise. Accessible at: <https://gnomonwise.org/ge/publications/analytics/22>

¹² W. J. McGuire, A vaccine for brainwash. *Psychol. Today*. **3**, 36–64 (1970).

¹³ Roozenbeek, J. and van der Linden, S. (2021). Inoculation Theory and Misinformation. *NATO Strategic Communications Centre of Excellence*. Online ISBN: 978-9934-564-49-9.

¹⁴ John A. Banas & Stephen A. Rains. (2010). A Meta-Analysis of Research on Inoculation Theory, *Communication Monographs*, 77:3, 281-311, DOI: 10.1080/03637751003758193

academic findings that scholars began incorporating inoculation theory into their strategies following the 2016 US Presidential Elections, particularly in the context of modern online disinformation. Initially, pre-bunking approaches were employed in relation to specific issues, such as climate change denial campaigns. However, over time, scholars shifted their focus from specific issues to manipulation techniques in general. This shift enables the audience to critically assess any piece of manipulated information, regardless of its topic.

Inoculation based on techniques reveals and clarifies for the audience the manipulation methods typically characteristic of almost all types of disinformation narratives. These methods include conspiracy theories (attributing all of society's problems to a specific and isolated group), emotionally manipulative language that elicits anger and fear, communication designed to polarize and divide society, and the use of so-called trolls and bots in social networks, among others. Roozenbeek and van der Linden (2021) believe that if people were to be inoculated against these techniques, then they might become better able to recognize the use of such techniques in the content they see online.¹⁵

To assess the validity of the aforementioned approach, Roozenbeek, van der Linden, and Nygren (2020)¹⁶ developed an online game called Bad News,¹⁷ featuring a fictitious social media environment. In this setting, players assume the role of creators of manipulative information and go through six steps, each covering a specific disinformation technique. Concurrently, players become acquainted with the aforementioned techniques and ultimately gain the skills to recognize them in various situations. This effect was demonstrated through an experiment involving players who spoke five different languages (English, German, Greek, Polish, Swedish). The study revealed that after completing the game, individuals, irrespective of their age, gender, level of education, or political ideology, significantly improved their ability to identify manipulative techniques and, consequently, detect disinformation.

Another online game, Harmony Square,¹⁸ created later with the aim of reducing disinformation and polarization, also demonstrated the same effect. The game-based experiment revealed a decrease in the trustworthiness of disinformation and an enhancement of individuals' abilities to identify it.

¹⁵ Roozenbeek, J. and van der Linden, S. (2021). Inoculation Theory and Misinformation. *NATO Strategic Communications Centre of Excellence*. Online ISBN: 978-9934-564-49-9.

¹⁶ Roozenbeek, J. van der Linden, S. Nygren, T. (2020). Prebunking interventions based “inoculation” theory can reduce susceptibility to misinformation across cultures. *The Harvard Kennedy School (HKS) Misinformation Review*, Volume 1, Issue 2. Via link: <https://api.repository.cam.ac.uk/server/api/core/bitstreams/46689225-6647-48ee-a3ae-1a5bed445d92/content>

¹⁷ See the game's link: <https://www.getbadnews.com/books/english/intro>

¹⁸ See the game's link: <https://harmonysquare.game/>

Furthermore, those who played this game were much less inclined to share unverified information on their social media profiles compared to the control group.¹⁹

The game "Go Viral!,"²⁰ created by a group of scholars (Basol et al., 2021)²¹ focusing on COVID-19 topics, also illustrated the effectiveness of prebunking disinformation. Empirical data collected from a five-minute play of this game demonstrated an improvement in COVID-19 disinformation detection skills at a multicultural level, an increase in individuals' self-confidence in countering manipulative messages, and a reduction in the desire to share disinformation with others. Additionally, according to the study, the first two effects persisted for at least one week after the end of the game.

Among the games based on the inoculation theory, it's important to highlight the Cranky Uncle²² game. Covering fourteen manipulation techniques, it is designed for educational purposes in schools and comes with a guide for teachers. Additionally, noteworthy are videos created for similar purposes, requiring less time than games while still having a noticeable impact on dealing with manipulative techniques.²³ Research indicates that running an advertising campaign on YouTube using inoculation videos significantly enhances the ability of users to correctly identify manipulative content. Consequently, political decision-makers can implement similar campaigns on various social media platforms using inoculation videos.²⁴

Roozenbeek, van der Linden, and Nygren (2020) advocate a similar recommendation for democratic governments, social media platforms, and educational institutions. They contend that large-scale "vaccination programs" based on the inoculation theory can be far more effective in combating online disinformation compared to the classic fact-checking approach.²⁵ The governments of the USA and UK have already taken steps in this direction, actively collaborating with researchers and practitioners in

¹⁹ Roozenbeek, J. and van der Linden, S. (2021). *Inoculation Theory and Misinformation*. NATO Strategic Communications Centre of Excellence. Online ISBN: 978-9934-564-49-9.

²⁰ See the game's link: <https://www.goviralggame.com/en>

²¹ Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W.P., van der Linden, S. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society*. January-June: 1-18. DOI: 10.1177/20539517211013868

²² See the game's link: <https://crankyuncle.com/>

²³ See the videos' URL: <https://inoculation.science/inoculation-videos/>

²⁴ Traberg, C.S. et al. (2023). Prebunking Against Misinformation in the Modern Digital Age. In: Purnat, T.D., Nguyen, T., Briand, S. (eds) *Managing Infodemics in the 21st Century*. Springer, Cham. https://doi.org/10.1007/978-3-031-27789-4_8

²⁵ Roozenbeek, J. van der Linden, S. Nygren, T. (2020). Prebunking interventions based "inoculation" theory can reduce susceptibility to misinformation across cultures. *The Harvard Kennedy School (HKS) Misinformation Review*, Volume 1, Issue 2. Via link: <https://api.repository.cam.ac.uk/server/api/core/bitstreams/46689225-6647-48ee-a3ae-1a5bed445d92/content>

the field to develop an optimal strategy for dealing with disinformation through verification and early exposure (inoculation theory approach). Twitter (now X) also employed the inoculation approach to counter the spread of manipulative information during the 2020 presidential election. However, the exposure approach to disinformation, being relatively novel, is still in the process of development.²⁶

Based on the academic literature reviewed in this document, it can be argued that relying solely on a responsive (verifying) approach against misleading and malicious information to the public is insufficient. It is crucial to strengthen "herd immunity" and pre-bunk disinformation. In this process, an approach based on the inoculation theory aims to provide individuals with misleading messages, disseminated on a specific topic, in advance, with a small dose. Additionally, it familiarizes them with widespread techniques of information manipulation, making them more immune to disinformation. The effectiveness of such a "psychological vaccine" has been demonstrated by numerous studies. Furthermore, the inoculation process itself is not difficult or boring, as it is based on entertaining games and short videos. It has also been proven that such interventions enhance an individual's ability to identify manipulative information and reduce the desire to share disinformation with others. It is noteworthy that the disinformation pre-bunking approach is still in the development phase, and there are some unanswered questions at this point. According to scholars in the field, the main question pertains to time—how long the "psychological vaccine" can endure. It also needs to be seen whether this approach will work similarly successfully outside academic research institutions in real life, and where the boundary lies after which we can say that a particular society has obtained the so-called "herd immunity" in the fight against manipulative information.²⁷

²⁶ van der Linden, S., Roozenbeek, J., Maertens, R., Basol, M., Kácha, O., Rathje, S., & Traberg, C. S. (2021). How can psychological science help counter the spread of fake news? *The Spanish Journal of Psychology*, 24. e25. <https://doi.org/10.1017/SJP.2021.23>

²⁷ Roozenbeek, J. and van der Linden, S. (2021). *Inoculation Theory and Misinformation*. NATO Strategic Communications Centre of Excellence. Online ISBN: 978-9934-564-49-9.