

Fact-Checking Against Disinformation - Empirical Evidence

Davit Kutidze¹

In the modern information environment, the problem of disinformation has become a global phenomenon that affects social, economic, and political life. Digital platforms and social media have greatly simplified the spread of false information, which various actors often use as a tool to manipulate public opinion. The harm caused by disinformation can appear on many levels. One of the most relevant areas is democratic processes, where false beliefs affect election results and political stability. Also harmful is the degradation of information, where constant exposure to false information reduces citizens' ability to possess accurate factual knowledge necessary for making informed decisions. In addition, a dangerous effect of disinformation is the weakening of democratic accountability, especially when the spread of false or manipulative information not only fails to damage the reputation of the person spreading it but, in many cases, even benefits them. This trend weakens the responsibility of politicians and other public figures to society and worsens the quality of public discussion. Finally, among the risks posed by disinformation, one of the most serious is its harmful impact on public health and safety, as it indirectly threatens human lives. This is especially evident during epidemics (for example, the COVID-19 infodemic²), when disinformation spreads in the form of medical myths.

The dangers of disinformation are amplified by the fact that false information often spreads faster than verified facts, especially in times of crisis. This trend can be explained by several factors.³ First, creating and sharing fake news is relatively easy. In the age of online platforms and social media, it can be done quickly and at almost no cost. By contrast, factual information requires verification and analysis of multiple sources. Disinformation also has an advantage because its creators do not follow journalistic standards. They often use sensational, clickbait headlines that attract a large audience. In addition, many social media users consume news for entertainment and are not motivated to think critically about the source of a story or the real motives of its author. As a result, a large number of users read only the headline.

¹ Research Institute Gnomon Wise; e-mail: d.kutidze@ug.edu.ge

² Cambridge Dictionary. Via link: <https://dictionary.cambridge.org/dictionary/english/infodemic>

³ 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>

In parallel with the popularity of social networks and the growth of disinformation flows, fact-checking⁴ has acquired an important role on a global scale. Naturally, it is not a magic wand against fake news, but together with various initiatives it makes its own contribution to the fight against disinformation.⁵ On the other hand, there are opinions that the importance of fact-checking in the fight against disinformation is exaggerated and that it is not very effective. To support this opinion, the concept of the so-called backfire effect is often examined, according to which factual correction of false information or evidence that contradicts a person's existing beliefs, instead of changing them, often has the opposite effect and strengthens these false beliefs even more.⁶

Despite the so-called backfire theory mentioned above, a number of reliable and high-quality studies and their meta-analyses conducted in recent years demonstrate the important role of fact-checking in the fight against disinformation. A study published in 2021,⁷ which was conducted in four different countries (Argentina, Nigeria, South Africa, and the United Kingdom), confirmed that fact-checking significantly reduced media users' erroneous beliefs on various issues. On average, fact-checking increased factual knowledge by 0.59 points on a 5-point scale. The authors of the study conducted experiments in the four countries mentioned above. One group of participants in the experiment was shown only false information, the second, the so-called correction group, was shown disinformation, which was immediately followed by the corresponding verified fact, and the third, control group was not provided with any prior information and answered questions about the possible truthfulness of a particular story as is. In each country, 2,000 people were surveyed. To assess the accuracy of respondents' answers, the researchers used a 5-point scale, and as noted, overall, familiarization with verified facts increased the accuracy of respondents' answers by 0.59 points. It is also worth noting that this result was not short-lived, and most of the positive effect of fact-checking persisted after two weeks. However, the study did not confirm the so-called backfire effect theory, according to which people whose ideological beliefs are

⁴ Graves, L. and Amazeen, M. A. (2019). Fact-Checking as Idea and Practice in Journalism. *Oxford Research Encyclopedia of Communication*. DOI: 10.1093/acrefore/9780190228613.013.808

⁵ Graves, L. and Cunliffe-Jones, P. Misinformation: how fact-checking journalism is evolving – and having a real impact on the world. *The Conversation*. Via link: <https://theconversation.com/misinformation-how-fact-checking-journalism-is-evolving-and-having-a-real-impact-on-the-world-218379>

⁶ Nyhan, B. and Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*. DOI: 10.1007/s11109-010-9112-2

⁷ E. Porter, & T.J. Wood, The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom, *Proc. Natl. Acad. Sci. U.S.A.* 118 (37) e2104235118, <https://doi.org/10.1073/pnas.2104235118> (2021).

contradicted by facts are even more likely to believe lies. On the contrary, fact-checking increased the level of accuracy among representatives of all ideological groups.

In terms of the scale of the study population, we should single out another study (van Erkel et al., 2024),⁸ in which the effect of fact-checking was examined across 16 European countries. These countries represent both Western and Eastern Europe and differ significantly in their media consumption contexts. The study draws on the theory of motivated reasoning. The authors investigate two opposing phenomena: accuracy goals, in which people seek correct information, and directional goals, in which people seek to maintain consistency with their existing beliefs, accept information they prefer, and avoid cognitive dissonance. The study also examines how the reputation of the information provider affects the perceived reliability of the information. As noted, the study was conducted simultaneously in 16 European countries. In each country, one thousand respondents were surveyed, for a total of 16,000 participants. The experiment measured short-term effects by providing respondents with correct information immediately after exposing them to the fake news item. The results confirmed the effectiveness of fact-checking in all countries studied, along with a clear reduction in misperceptions on the specific topic (an average improvement of 1.29 points on a 5-point scale). In addition, the study further questioned the existence of the so-called “backfire effect.” However, it also found that the effect of fact-checking is relatively smaller among those who support a politician spreading misinformation. The effectiveness of fact-checking is supported not only by individual studies but also by meta-analyses of academic papers on the topic. According to one such meta-analysis⁹ published in 2019, which reviewed 30 individual studies, fact-checking has a significant positive impact on correcting misperceptions.

Fact-checking is particularly important within the context of social media. In 2024, researchers at the Massachusetts Institute of Technology (MIT) demonstrated¹⁰ that warning labels (about the falsity of a specific piece of news), used on social platforms such as Facebook, Instagram, and Twitter (now X), significantly reduced users’ intention to believe and share false information. According to the study, on average, belief in marked fake news decreased by 27.6%, and the intention to share it decreased by 24.7%.

⁸ van Erkel, P. F. A., van Aelst, P., de Vreese, C. H., Hopmann, D. N., Matthes, J., Stanyer, J., & Corbu, N. (2024). When are Fact-Checks Effective? An Experimental Study on the Inclusion of the Misinformation Source and the Source of Fact-Checks in 16 European Countries. *Mass Communication and Society*, 27(5), 851–876. <https://doi.org/10.1080/15205436.2024.2321542>

⁹ Nathan Walter, Jonathan Cohen, R. Lance Holbert & Yasmin Morag. (2019). Fact-Checking: A Meta-Analysis of What Works and for Whom, *Political Communication*, Via link: <https://doi.org/10.1080/10584609.2019.1668894>

¹⁰ Martel, C., Rand, D.G. Fact-checker warning labels are effective even for those who distrust fact-checkers. *Nat Hum Behav* 8, 1957–1967 (2024). <https://doi.org/10.1038/s41562-024-01973-x>

It is also worth noting that this effect was observed to some extent even among social media users who had low trust in fact-checking organizations. More specifically, belief in fake news among such users decreased by 12.9%, and their willingness to share such news decreased by 16.7%. The effectiveness of so-called "warning labels" on social platforms has been confirmed by numerous studies and reports conducted over the past decade. According to theoretical and synthetical reviews,¹¹ these labels effectively reduce belief in disinformation and its spread. However, at the same time, the effect of "warning labels" depends on how wide an audience they reach, how specific and visible they are to the user, and whether they come from highly reliable sources. This is where another important issue comes in: the influence of so-called moderators on the effectiveness of fact-checking. Moderators' roles include selecting appropriate text or visuals for the fact being checked, verifying reliable sources, ensuring clear communication, and managing a range of other components, all of which may further increase the effectiveness of detecting and correcting disinformation.

The effectiveness of fact-checking is not static and depends on the structure of the message, the characteristics of the source, and individual preferences. Studies show that properly selected conditions can significantly increase the effect of fact-checking. First of all, the form in which verified information is delivered to the individual is of crucial importance. In particular, fact-checking is most effective when it not only refutes falsehoods but also provides the reader with an alternative, accurate explanation.¹² Visual elements are equally important. Numerous studies indicate that graphic images that clearly demonstrate the falsity of specific information improve comprehension when they are properly integrated with the text. In addition, when exposing fake news, linguistic simplicity is important. Texts written in clear, understandable language and not overloaded with complex vocabulary demonstrate higher effectiveness, especially among groups with low media literacy.¹³

For the success of accurate information, moderators directly connected to its dissemination are of great importance. In particular, the effect of fact-checking is stronger when respondents have a high level of trust in the organization responsible for dissemination. In this regard, independent fact-checking organizations play a special role, as they are often trusted more by people who are distrustful of traditional

¹¹ Cameron Martel, David G. Rand, Misinformation warning labels are widely effective: A review of warning effects and their moderating features, *Current Opinion in Psychology*, Volume 54, 2023, 101710, ISSN 2352-250X, <https://doi.org/10.1016/j.copsyc.2023.101710>.

¹² E. Porter, & T.J. Wood, The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom, *Proc. Natl. Acad. Sci. U.S.A.* 118 (37) e2104235118, <https://doi.org/10.1073/pnas.2104235118> (2021).

¹³ Ibid, pp. 5, <https://doi.org/10.1073/pnas.2104235118>

media.¹⁴ In addition to these factors, the effect of fact-checking also depends significantly on the audience's preconceptions or ideological factors. It is stronger when it confirms a person's existing views than when it contradicts them.¹⁵

When discussing the impact of fact-checking organizations, we should not limit ourselves to emphasizing that they only check individual facts and do nothing else to combat disinformation. According to the Spanish fact-checking organization Maldita.es, their work goes beyond reacting to already widespread falsehoods and moves into a proactive phase called pre-bunking. This strategy involves preparing the public for disinformation before it reaches them. Instead of simply debunking a false story after the fact, the organization explains in advance the manipulative mechanisms and strategies that disinformation purveyors typically use. This process is similar to psychological inoculation. The user receives a small "dose" of disinformation along with an analysis of how it works, which builds a kind of cognitive immunity. As a result, when people encounter real disinformation, they are less vulnerable because they already have tools to identify falsehoods. Therefore, fact-checking organizations work not only to correct specific facts but also to increase public awareness, helping people critically evaluate suspicious information based on its form or distribution technique.¹⁶

In conclusion, one of the most significant challenges in the modern information environment is disinformation, which undermines democratic processes and citizens' awareness in various ways. In this context, fact-checking emerges as an important tool. Although doubts about its effectiveness have existed in the past, particularly due to the so-called "backfire" theory, the studies discussed above clearly demonstrate that fact-checking reduces false beliefs and enhances factual knowledge across different societies. Furthermore, its effects may be sustained over time and not limited to the short term.

¹⁴ van Erkel, P. F. A., van Aelst, P., de Vreese, C. H., Hopmann, D. N., Matthes, J., Stanyer, J., & Corbu, N. (2024). When are Fact-Checks Effective? An Experimental Study on the Inclusion of the Misinformation Source and the Source of Fact-Checks in 16 European Countries. *Mass Communication and Society*, 27(5), 851–876. <https://doi.org/10.1080/15205436.2024.2321542>

¹⁵ Nathan Walter, Jonathan Cohen, R. Lance Holbert & Yasmin Morag (2019): Fact-Checking: A Meta-Analysis of What Works and for Whom, Political Communication. <https://doi.org/10.1080/10584609.2019.1668894>

¹⁶ Maldita.es, (March 23, 2023). What is pre-bunking and how to fight disinformation before a debunking? Via link: <https://maldita.es/nosotros/20230323/prebunking-que-es-antes-desmentido/>

It is worth noting that effectiveness depends on many factors, including the form of information delivery, the credibility of the source, visual and linguistic elements, and the audience's preconceptions. In the context of social networks, warning labels and cooperation with platforms become particularly important, as they reduce both belief in disinformation and its spread.

As mentioned earlier, although fact-checking plays a crucial role in detecting disinformation, it is not the only solution. The spread of disinformation is a constant process, and the fight against it requires continuous efforts and differentiated approaches. In this regard, it is essential that fact-checking organizations globally not only correct individual inaccuracies but also proactively combat fake news. The so-called prebunking aims to prepare the public in advance to recognize disinformation. This approach fosters critical thinking and creates a kind of "cognitive immunity."

Ultimately, fact-checking is an effective tool in the fight against disinformation, especially when integrated into a broader strategy that includes both technological and educational components.