

Manipulation of Consciousness: Psychological Mechanisms for Spreading Disinformation in Crisis Situations

By Mykola Babii¹, Olesia Posvistak², Yasmina Korokhod³, Yuriy Opoka⁴, Svitlana Kustova⁵

ABSTRACT:

During crises such as wars, pandemics, and economic or technological disasters, disinformation spreads faster and wider than accurate information, amplifying social anxiety and undermining institutional trust. Its effects are driven by cognitive biases (illusory truth effect, confirmation heuristics), group identity congruence, and emotional triggers. The aim of this study is to integrate these psychological mechanisms and assess the effectiveness of inoculation (prebunking) and cognitive (accuracy-prompt) interventions in countering crisis-related disinformation. A preregistered multi-wave online experiment with a 2×2×2 factorial design was conducted on a sample of about 2000 adults from Ukraine and Central/Eastern Europe. Results indicate that repeated exposure amplifies the illusory truth effect; congruence with social identity increases credibility; threatening emotional tone enhances virality. Stand-alone interventions (prebunking or accuracy prompts) reduced belief and sharing intention by about 10–15%, whereas their combination achieved more than a 20% reduction with lasting effects. Age, need for cognition, and conspiratorial beliefs moderated susceptibility to disinformation. These findings demonstrate that combined preventive strategies can substantially strengthen societal informational resilience during crises.

Keywords: Disinformation, infodemic, illusory truth effect, social identity, emotional triggers, prebunking, accuracy prompts, informational resilience, crisis communication

1. Introduction

The information space of the modern world is undergoing radical transformations, accompanied by an increase in the scale and speed of dissemination of false information. periods of social and political crises, such as armed conflicts, pandemics, economic or man-made disasters, are particularly vulnerable. Empirical studies

¹PhD in Psychology, Associate Professor, Department of Socio-humanitarian Technologies, Faculty of Digital, Educational and Social Technologies, Lutsk National Technical University, Lutsk, Ukraine.

²Doctor of Psychology Science, Professor, Department of International Communication and Political Science, Faculty of International Relations and Law, Khmelnytskyi National University, Khmelnytskyi, Ukraine.

³PhD in Political Science, Associate Professor, Expert-psychologist of the Center of Ukrainian Researchers in Austria, Vienna, Austria.

⁴Doctor of Social Science, Associate Professor of the Department of Social Disciplines, Lviv State University of Internal Affairs, Lviv, Ukraine.

⁵PhD in Public Administration, Associate Professor of the Department of European Integration Policy, Educational and Scientific Institute of Public Administration and Civil Service of Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.

show that false messages spread much faster, deeper, and wider on social media than true ones, and this effect is driven mainly by human behavior rather than automated bots (Vosoughi & Aral, 2018; Lazer *et al.*, 2018). At the same time, a review of definitions emphasizes that “fake news” is an umbrella term that covers different types of distorted content (Tandoc & Ling, 2017).

This imbalance of information flows creates infodemics that increase public anxiety, undermine trust in institutions, and complicate crisis response (World Health Organization, 2020; Nelson *et al.*, 2020). In the context of armed conflicts, disinformation causes specific types of harm to vulnerable groups, ranging from poor security decisions to escalation of violence (Ulbricht & Rizk, 2024). The Ukrainian regulatory context also reflects the need for systematic approaches: protocols for crisis communications during cyber incidents have been approved (Ministry of Health of Ukraine, 2023).

Citizens’ vulnerability to disinformation has a complex psychological nature, including cognitive, social, identification, and emotional mechanisms. Systematic reviews of the literature show that cognitive biases, emotional appeals, and motivated thinking related to group identity significantly determine the way information is processed and the behavior of online users (Munusamy *et al.*, 2024; Ecker *et al.*, 2022; Lewandowsky *et al.*, 2017). One of the key cognitive phenomena is the illusory truth effect: repeated exposure to the same message increases its perceived credibility and willingness to spread, even in the absence of factual confirmation (Udry & Barber, 2023; Vellani *et al.*, 2023). Experimental work shows that intensifying reliance on emotions (fear, anger, indignation) promotes belief in fake news, while stimulating reflective thinking reduces its impact (Martel *et al.*, 2020; Lutzke *et al.*, 2019). In addition, the alignment of information with group identity and prior beliefs leads to motivated acceptance; people tend to accept and share messages that confirm their worldview (Ecker *et al.*, 2022; Munusamy *et al.*, 2024; Ahmed & Rasul, 2022).

Importantly, while repetition, identity congruence, and emotional threat are widely documented mechanisms of misinformation susceptibility, their relative strength and interaction patterns may vary across crisis types and information environments. For example, war-related uncertainty and security threats may intensify identity-based motivated processing, whereas pandemic contexts can amplify reliance on heuristic cues under cognitive overload, and economic crises may heighten sensitivity to loss-framed narratives. In addition, media ecosystems and cultural settings (e.g., platform affordances, dominant communication channels, trust in institutions, and prevailing social norms) can modulate how these mechanisms translate into belief formation and sharing behavior. By explicitly acknowledging this likely contextual heterogeneity, the present study strengthens theoretical clarity and applied relevance while retaining the same experimental logic and analytic strategy.

Classical theories of social psychology provide a theoretical basis for understanding these effects. Allport and Postman, in their seminal work *The Psychology of Rumor*, emphasized the role of emotional tension and social context in the emergence and spread of rumors (Allport & Postman, 1947). The concept of cognitive dissonance explains the desire of individuals to avoid contradictions between beliefs and actions, which leads to selective perception and acceptance of congruent information (Festinger, 1957). Zaller’s model of the emergence of mass opinion postulates that people accept

messages that are consistent with their preconceptions and reject dissonant influences, forming a selective filter of information (Zaller, 1992).

At the same time, Elizabeth Noelle-Neumann's theory of the "spiral of silence" describes how social pressure and fear of isolation encourage people to hide unpopular views, and dominant narratives receive disproportionate visibility (Noelle-Neumann, 1974). Together, current reviews and experiments show that: first, cognitive biases (the familiarity effect, Dunning-Kruger) provide the ground for "easy" belief; second, social identity guides perceptual selectivity; third, negative affect and threat increase virality; fourth, platform algorithms amplify these processes through "information bubbles" and diffusion of sensational content (Dunning & Kruger, 1999; Lazer et al., 2018; Allcott & Gentzkow, 2017; Vosoughi et al., 2018).

Against this backdrop, preventive approaches – inoculation strategies and accuracy prompts – demonstrate the best reproducible effectiveness. In particular, prebunking in the form of games (Bad News), short videos, or flashcards shows consistent results across countries and languages (Roozenbeek & van der Linden, 2018, 2019; Roozenbeek et al., 2020; Bessarabova et al., 2024). Accuracy prompts reduce the spread of fake news without compromising the credibility of the news (Pennycook & Rand, 2022; Pennycook et al., 2021). Meta-analyses emphasize that effective corrections should be detailed, repeated, and offer alternative narratives (Prike & Ecker, 2023). At the same time, international organizations (World Health Organization, 2020) and national policies (Ministry of Health of Ukraine, 2023) call for combined approaches – both preventive and reactive – to combat infodemics.

In view of the above, this study aims to integrate the cognitive component (the effect of repetition and illusory truth), the social-identification component (congruence with group identity), the emotional component (threatening and emotionally charged frames), and algorithmic amplifiers into a single theoretical model of disinformation influence. In addition, it is planned to compare the effectiveness of inoculation interventions and accuracy prompts separately and in combination; to study their effect in the field in Central and Eastern Europe; to identify moderators of influence (age, need for cognition, conspiratorial beliefs); to formulate recommendations for strategic communications, media education and infodemic management during crises.

2. Theoretical Background

Modern disinformation research pays considerable attention to the cognitive mechanisms that determine the perception and spread of false messages. Empirical research shows that repeated exposure to the same claims increases their perceived credibility: the *illusory truth effect* is based on the fact that fluency is mistakenly perceived as a sign of credibility (Udry & Barber, 2023; Vellani et al., 2023). Even if the source is known to be dubious, repetition stimulates automatic responses and reduces attention to the content (Ecker et al., 2022). The key cognitive biases of include the confirmation bias and the availability heuristic, which lead to the selection of information that is consistent with prior beliefs, and the Dunning-Kruger effect, which explains that individuals with low levels of competence overestimate their own knowledge and unconsciously spread unverified claims (Dunning & Kruger, 1999). The reviews by Munusamy et al. (2024) and

Ecker et al. (2022) emphasize that unreliable memory and selective evidence seeking are basic determinants of fake news persistence. Lewandowsky et al. (2020) in *The Debunking Handbook 2020* recommend presenting truthful information at the beginning and end of a rebuttal (“truth sandwich”) and repeating corrective messages to neutralize the impact of disinformation.

An important factor is social identity: based on the theory of cognitive dissonance, people seek to avoid contradictions between their beliefs and behavior, which encourages them to select information that confirms their group membership (Festinger, 1957). The classic concepts of Zaller (1992) and Noelle-Neumann (1974) showed that the perception of political messages depends on the level of awareness, and the fear of social exclusion forms a “spiral of silence”. Modern reviews (Munusamy et al., 2024; Ecker et al., 2022) show that group identity enhances motivated thinking: people tend to trust messages that resonate with their social groups and reject information that causes dissonance. The influence of identity is modified by political ideology, education, and media literacy; together, these factors determine whether people are open to correction or remain in information bubbles.

Emotional triggers also have a significant impact on the virality and perception of disinformation. A study by Martel et al. (2020) found that subjective emotional orientation correlates with belief in fake news, while inducing analytical thinking reduces this effect. Negative emotions (fear, anger, indignation) stimulate virality because emotional messages attract attention and cause impulsive reactions. Back in, Allport and Postman (1947) proved that rumors spread in conditions of high social tension; this conclusion remains relevant for digital infodemics. Positive emotions (curiosity, excitement) can be used in curricula to increase interest in media education and critical thinking (Lutzke et al., 2019). The reviews by Munusamy et al. (2024) and Ecker et al. (2022) emphasize that emotional appeals and fear of threat unite audiences around congruent information. Positive emotions (curiosity, excitement) can be used in prevention programs to draw attention to educational materials and develop critical thinking.

Algorithmic and network factors are important in the structure of disinformation dissemination. Social media platforms use algorithms that rank content according to popularity and likelihood of interaction; these filters amplify sensationalized messages and facilitate the rapid spread of false stories, especially when they evoke strong emotions (Lazer et al., 2018). A study by Allcott and Gentzkow (2017) showed that during the 2016 US election, false news was more widely shared than true news. Vosoughi et al. (2018) used a large sample to show that fake news spreads faster and more widely than reliable information, especially when it evokes strong emotions. These mechanisms increase network polarization and contribute to the spread of disinformation in crisis situations. Modern graph analysis and machine learning methods can identify “super-spreaders,” but they are only beginning to integrate the psychological characteristics of users, leaving room for multidisciplinary research.

Intervention strategies have been developed in response to the challenges. The classical basis is the theory of inoculation by McGuire (1961), which proved that prior familiarization with counterarguments builds resistance to manipulation. Based on this theory, Roozenbeek and van der Linden (2018, 2019) developed game and video prebunking formats. Roozenbeek et al. (2020) field experiments on YouTube in Central

and Eastern Europe showed that short videos significantly reduce the perceived credibility of disinformation. A meta-analysis by Bessarabova et al. (2024) confirmed the universality of inoculation approaches in the United States and Finland.

Another promising strategy is *accuracy prompts*, which remind people to evaluate the veracity before sharing. A meta-analysis by Pennycook and Rand (2022) and experiments by Pennycook et al. (2021) show that this strategy reduces the spread of fake news without compromising credible news. Prike and Ecker (2023) conclude that effective corrections should include explanations, alternative narratives, and repeatability. Lutzke et al. (2019) and van der Linden et al. (2017) showed that warning inoculation messages increase resistance to climate disinformation. However, combined strategies that combine prebunking and accuracy prompts have not been sufficiently studied, although preliminary results indicate that they are more effective.

Thus, the current literature confirms that cognitive biases, social identity, emotional triggers, and algorithmic amplifiers form a complex system that contributes to the spread of disinformation. Intervention approaches, such as inoculation and accuracy prompts, have been shown to be highly effective in different contexts, but there is a lack of integrated models that take all these factors into account simultaneously. Multidisciplinary research combining methods of social psychology, data analysis, and communication sciences is needed to develop comprehensive strategies for improving the information resilience of society.

3. Purpose of the Study, Research Questions and Hypotheses

Previous research suggests that cognitive biases, social identification motives, and emotional triggers contribute to the acceptance and spread of disinformation in different ways (Munusamy et al., 2024; Martel et al., 2020); however, there is a lack of integrative models that analyze these mechanisms simultaneously (Ecker et al., 2022) and evaluate the effectiveness of combined preventive measures.

The purpose of our empirical study is to develop and test an integrated model that combines cognitive (repetition/fluency), social-identity (congruence with group identity), emotional (threatening tone), and algorithmic factors of disinformation exposure and evaluates the effectiveness of inoculation interventions and accuracy prompts, separately and in combination. An additional goal is to identify moderators (age, need for cognition, conspiratorial beliefs) and formulate recommendations for strategic communications, media literacy, and infodemic management in times of pandemics, hybrid wars, or other social upheavals.

Taking into account theoretical and empirical developments, the authors formulate the following research questions. *RI1*: How do the factors of repetition, congruence with group identity, and emotional intensity interact in predicting disinformation belief and intention to spread during crises? *RI2*: what is the relative effectiveness of inoculation videos, precise cues, and their combination in reducing belief and intention to spread crisis disinformation? *RI3*: Do individual characteristics (need for knowledge, level of conspiracy beliefs, political identity) and demographic variables (age, gender) modify the impact of the above factors on the perception of disinformation?

In response to these questions, the following *hypotheses* are put forward. *H1 (illusory truth effect)*. Repeated exposure to a disinformation message increases its perceived credibility and intention to spread (Udry & Barber, 2023; Vellani *et al.*, 2023). *H2 (social congruence)*. Disinformation that is consistent with the respondent's group identity (political or national) elicits greater belief and intention to spread than incongruent information (Ecker *et al.*, 2022; Munusamy *et al.*, 2024). *H3 (emotional triggers)*. Messages with high emotional intensity (fear, anger) stimulate higher belief and virality than emotionally neutral messages (Lutzke *et al.*, 2019; Martel *et al.*, 2020). *H4 (interventions)*. A combined intervention that combines inoculation messages (prebunking) and accurate cues, is more effective in reducing belief and intent to spread disinformation than either approach alone (McGuire, 1961; Pennycook & Rand, 2022; Roozenbeek *et al.*, 2020). *H5 (moderators)*. The effects of repetition, congruence, and emotional intensity are stronger for respondents with low cognition need and high levels of conspiracy beliefs, and vary by age and education (Dunning & Kruger, 1999; Munusamy *et al.*, 2024).

4. Methods

The study was conducted as part of a pre-registration multiwave online experiment with a 2×2×2 factorial design, which allows simultaneous analysis of the influence of several psychological factors and their interactions. The sample consisted of about 2,400 adult respondents from Ukraine and Central and Eastern Europe; after filtering by quality criteria (completion of control questions, minimum time to complete, uniqueness of IP address, absence of straight-line), approximately 2,000 participants were included in the analysis. Stratified selection ensured balance in terms of age, gender, and region. Recruitment took place through online panels and social media, with the purpose of the study, guarantees of confidentiality, and the ability to withdraw at any time. A detailed description of the design, procedure, and statistical approaches is provided in Appendix A.

The experiment manipulated three main factors: *repetition*, *congruence with group identity*, and *emotional tone* of the messages. Repetition involved one or three exposures to the same message to simulate the illusory truth effect reported in recent studies (Vellani *et al.*, 2023). Congruence was based on previously measured self-identifications (political and national): some messages contained frames aligned with the respondent's group, while others remained neutral, allowing for the assessment of motivated thinking and the confirmation effect (Ecker *et al.*, 2022). The emotional tone varied between threatening/emotionally charged and neutral wording, as negative emotions such as fear or anger have been shown to be a key factor in virality (Martel *et al.*, 2020; Lutzke *et al.*, 2019).

All experimental messages were related to crisis topics and covered three domains: security/war, health, and economy. Eight truthful and disinformation variants were developed for each domain, with the latter containing typical manipulative techniques (emotional language, pseudo-expertise, conspiracy theories, false dilemmas). The fact-checkers independently verified the truthful messages. The rehearsal (three exposures) and control of the time between waves (48–72 hours) allowed us to assess both immediate and delayed effects of repetition.

The interventions were based on prebunking theory and accuracy prompts. Intervention 1 (prebunking) consisted of a 90-second video exposing the mechanisms of common manipulations and included short recognition exercises; the concept was borrowed from the *Bad News* inoculation game, which has been shown to be effective across cultures (Roozenbeek & van der Linden, 2018, 2019). Intervention 2 (accuracy prompt) was a concise text message reminding users to assess the accuracy of news before sharing it, which has previously been shown to reduce the sharing of fake headlines (Pennycook & Rand, 2022). Intervention 3 combined both approaches. The control group did not receive any prompts or videos, which allowed us to measure the natural level of receptivity. An example of the instructions for participants is provided in Appendix B.

The dependent variables included: (1) perceived truthfulness (scale 1–7); (2) intention to share (scale 1–7); (3) factual memory (number of statements correctly recalled); (4) trust in the source; and (5) emotional tone assessment. Moderators: the need for cognition (short version of the Need for Cognition Scale), conspiratorial beliefs (Generic Conspiracy Beliefs Scale), political and national identity, self-assessment of media literacy, and demographic variables (age, gender, education). Two mindfulness checklists were used. Standardized scales were used to measure the moderators (Appendix C1). The stimulus material consisted of true and false messages in the areas of health, economy, and security (Appendix C2).

The procedure consisted of three waves. In wave 1, participants gave informed consent, answered moderated questionnaires, and were then randomly assigned to groups. Depending on the assigned condition, they either watched or did not watch the inoculation video; they were shown a precise prompt before reading the messages, if necessary. Each message was shown on the screen, after which respondents assessed its veracity, reported their intention to share, determined the emotional tone, and answered control questions. After 48–72 hours (Wave 2), for the repeated groups, half of the disinformation messages were presented again with the same ratings. Wave 3, 14 days later, assessed delayed effects; upon completion, all participants received a detailed debriefing with reliable facts and recommendations for media literacy (Lewandowsky et al., 2020). Visualizations of key findings are presented in Appendix D.

Ethical considerations were consistent with the principles of the Declaration of Helsinki; the study was approved by the ethics committee of the respective institution. Informed consent was mandatory; data were stored anonymously; disinformation materials were created specifically for the experiment and corrected after the debriefing.

Statistical analysis was performed in R (lme4/brms packages). Mixed linear and logistic models (cross-classified) with random intercepts for participants and messages were used; fixed effects included repetition, congruence, emotionality, interventions, and all their interactions. Moderators and demographic variables were entered as covariates. Power was calculated to detect small effects ($f^2 \approx 0.01$) with a power ≥ 0.90 . The Benjamini-Hochberg method was used to correct for multiple comparisons, and Bayesian intervals were also estimated. Scripts and data will be made available on OSF after publication.

5. Results

The first stage of the analysis confirmed that the experimental manipulations worked properly and reproduce psychological patterns typical of the information society. *Repeated exposure* to disinformation headlines significantly increased both the perceived truthfulness and the intention to share them, a phenomenon known as the illusory truth effect (Udry & Barber, 2023; Vellani et al., 2023). Its relevance is amplified in the world of continuous news and recommendation algorithms: the daily avalanche of information in times of pandemic or war encourages people to accept familiar messages as true (Nelson et al., 2020). *The congruence of information with group identity* also increased truthfulness ratings, confirming motivated thinking and selective processing (Ecker et al., 2022; Munusamy et al., 2024). *Emotional headlines* with fear or anger increased perception and virality, which is consistent with research on the role of affect in the spread of fakes (Lutzke et al., 2019; Martel et al., 2020).

The mixed linear models showed significant main effects of repetition, congruence, and emotionality, as well as significant interactions: repetition was enhanced in the identical group and threatening frames conditions, while the interaction between congruence and emotionality was not significant. This suggests that even emotionally neutral messages can gain credibility through repeated repetition and identity congruence, a characteristic feature of modern platforms.

Next, the authors analyze *the effectiveness of the interventions*. Prebunking (expository videos) and precise prompts separately reduced perceived truthfulness and intention to share by about 10-15%, replicating the results of previous meta-analyses (Pennycook & Rand, 2022). *The combined strategy* (prebunking+accuracy prompts) showed the greatest effect: a decrease of more than 20% compared to the control. This confirms that the combination of information inoculation, which raises awareness of manipulation, and cognitive priming, which shifts attention to fact-checking, creates synergy (Pennycook & Rand, 2022; Roozenbeek et al., 2020).

The reduction in the combined condition persisted after two weeks, while the changes in the control condition were minimal, indicating that the “psychological immunity” lasted. Extended results on intention to spread are presented in Appendix C4.

Table 1: Mean scores of perceived truthfulness and intention to share in different intervention conditions

Intervention condition	Perceived truthfulness (M±SD)	Intention to share (M±SD)
Control (no intervention)	4.2±1.1	3.8±1.2
Prebunking	3.8±1.0	3.2±1.1
Accuracy prompts	3.9±1.0	3.4±1.0
Combined intervention (prebunking + accuracy prompts)	3.5±0.9	2.9±0.9

Source: Own calculations based on empirical analysis of the experimental data (results of mixed linear models ($N \approx 2000$, $p < 0.01$))

Note: 1. See Appendix C3 for extended data; 2. Percentage reductions reported in the text are calculated relative to the control condition mean score; 3. The initial sample included approximately 2,400 respondents; after data quality filtering, the final analytic sample consisted of 1,988 participants.

The results of the mixed linear model are summarized in Table 2 (β -coefficients, standard errors, t/F -statistics, and p -values). It demonstrates significant main effects:

repetition ($\beta \approx 0.32$), congruence ($\beta \approx 0.23$), and emotional tone ($\beta \approx 0.27$), as well as important interactions, including repetition \times congruence ($\beta \approx 0.18$) and repetition \times emotionality ($\beta \approx 0.15$).

Contrasts show that the combined intervention ($\beta \approx 0.70$) outperforms both prebunking and accuracy prompts separately.

Table 2: Mixed linear model results for main effects, interactions, and interventions

Effect	β (SE)	t/F (df)	p-value	Interpretation
Repeat (H1)	0.32 (0.05)	t = 6.40 (1988)	< 0.001	Repeated exposure increases believability and intention to share
Identity congruence (H2)	0.23 (0.04)	t = 5.75 (1988)	< 0.001	Consistency with group identity increases belief
Emotional tone (H3)	0.27 (0.04)	t = 6.75 (1988)	< 0.001	Threatening/emotional frames increase virulence
Repetition \times Congruence	0.18 (0.06)	t = 3.02 (1988)	0.003	Synergy of repetition and identity
Repetition \times Emotionality	0.15 (0.06)	t = 2.50 (1988)	0.012	Synergy of repetition and emotionality
Congruence \times Emotionality	0.03 (0.06)	t = 0.50 (1988)	0.62	No significant interaction
Prebunking vs. Control	-0.40 (0.07)	t = -5.71 (1988)	< 0.001	Prebunking intervention reduces belief and spread
Accuracy of prompts vs Control	-0.30 (0.07)	t = -4.29 (1988)	< 0.001	Accuracy prompts reduce belief and spread
Combined vs Control	-0.70 (0.08)	t = -8.75 (1988)	< 0.001	Combination of two interventions is most effective

Source: created by the author based on the results of mixed linear models ($N \approx 2000$, $p < 0.01$)

Note: Reported coefficients are unstandardized fixed effects from mixed linear models; dependent variables were measured on 7-point Likert scales.

To conceptualize the interplay of cognitive (repetition and fluency effects), social-identity, emotional, and algorithmic factors, Figure 1 presents an integrated model of disinformation exposure. It demonstrates how these components together shape beliefs and behaviors regarding the spread of disinformation in the context of crisis information flows.

Figure 1. Integrated model of disinformation impact

Source: created by the author

Conceptual scheme that integrates the cognitive component (repetition and fluency effect), social identification component (congruence with group identity), emotional component (threatening and emotional frames), and algorithmic amplifiers (Figure 1). Arrows indicate the direction of influence; smaller arrows between modules indicate synergies and the potential impact of algorithmic reinforcers on each mechanism.

The moderators found important individual differences. Respondents with a high need for cognition showed lower receptivity and stronger reactions to the interventions; in contrast, participants with high conspiracy beliefs were more likely to believe fakes to be true and less likely to respond to prebunking and accuracy prompts, consistent with research on the effects of conspiracy beliefs (Dunning & Kruger, 1999; Munusamy *et al.*, 2024). Age had a U-shaped effect: middle-aged people (30-50 years old) were the most vulnerable, while younger and older people were more skeptical; this may be due to a combination of high social media activity and lack of media literacy. No gender differences were found.

Delayed effects showed that single interventions only provide a short-term reduction in beliefs; without interventions, beliefs remained almost unchanged after two weeks. Only the combined strategy maintained a significant effect (a decrease of 0.5-0.6 points), which emphasizes the need for ongoing and comprehensive measures against disinformation.

The results clearly confirm the hypotheses: repetition and fluency increase belief in false messages; congruence with group identity and emotional intensity enhance this effect; the combined intervention is the most effective, and individual factors (need for cognition, conspiratorial beliefs, age) modify the results. In today's realities of pandemics, wars, and economic instability, these findings emphasize that disinformation is not only a consequence of technology, but also a reflection of psychological and social processes. Pre-banking and accurate prompts, implemented systematically and over the long term, can significantly strengthen the information resilience of society.

6. Discussion

The experimental results confirm that disinformation spreads through the interaction of cognitive, social-identification, and emotional mechanisms, further amplified by algorithmic systems. Repeated exposure, alignment with group identity, and alarmist emotional tone form a particularly hazardous combination that increases the perceived credibility of false content and stimulates its virality (Vellani *et al.*, 2023). While prebunking messages and accuracy prompts independently reduce susceptibility to disinformation, the strongest and most durable effect emerges when these interventions are combined, underscoring the importance of integrated preventive strategies. These findings are consistent with inoculation theory (McGuire, 1961; Roozenbeek & van der Linden, 2018, 2019) and recent meta-analyses demonstrating the effectiveness of accuracy-based interventions (Pennycook & Rand, 2022).

From an applied perspective, the results indicate that crisis communication strategies should extend beyond reactive refutation of false claims and incorporate proactive measures that enhance citizens' understanding of manipulative techniques and strengthen evaluative skills. The findings further highlight the importance of accounting for individual differences: individuals with stronger conspiratorial worldviews or lower cognitive engagement may require more intensive or adaptive interventions. Accordingly, international guidelines and national crisis communication frameworks (World Health Organization, 2020; Ministry of Health of Ukraine, 2023) would benefit from systematically integrating preventive and corrective psychological tools.

Theoretically, these results support and extend classical concepts in social psychology. The observed synergy between identity congruence and repetition aligns with Festinger's theory of cognitive dissonance, which posits that individuals strive to minimize informational inconsistencies and that repeated, identity-congruent messages facilitate dissonance reduction (Festinger, 1957). The strong role of emotional framing likewise echoes Allport and Postman's classic observation that rumors flourish under conditions of heightened emotional tension and uncertainty (Allport & Postman, 1947).

Importantly, the present findings not only replicate well-established effects in the misinformation literature—such as the illusory truth effect, motivated reasoning, and affective virality—but also extend them by demonstrating their combined and mutually reinforcing operation within crisis contexts. Whereas much prior research has examined these mechanisms in isolation, this study advances an integrated model in which cognitive repetition, social identity alignment, and emotional threat framing jointly shape belief formation and sharing intentions. This holistic perspective is consistent with recent meta-analyses and experimental syntheses in psychology and communication research that emphasize interactions among cognitive, social, and affective processes rather than single-factor explanations. By empirically testing these mechanisms together, the study offers a more comprehensive account of misinformation dynamics under conditions of crisis.

The results further suggest that identity-based and emotional processes influence not only individual belief formation but also online behavior, reinforcing mechanisms such as the spiral of silence (Noelle-Neumann, 1974) and selective acceptance of political messages (Zaller, 1992). At the same time, the effectiveness of the intervention approach builds on robust empirical evidence supporting prebunking and accuracy prompts, highlighting the importance of integrating classical theory with contemporary psychological and technological practices.

Several limitations of the present study should be acknowledged. First, the empirical sample is geographically and culturally situated, focusing on Ukraine and countries of Central and Eastern Europe. Although this context is theoretically and empirically relevant given the prevalence of large-scale crises in the region, it may limit the generalizability of the findings to other cultural settings, media systems, or political environments. Replication in different regions would help assess the cross-cultural robustness of the observed effects. Second, the study relies on self-reported intentions to share information rather than directly observed sharing behavior. While prior research suggests that sharing intentions are meaningful predictors of actual behavior, this methodological choice may constrain external validity, particularly in real-world digital environments shaped by algorithmic amplification, social feedback, and platform-specific affordances. Accordingly, the results should be interpreted with appropriate caution when extending them to actual online behavior across platforms.

Building on these limitations, several promising directions for future research can be identified. These include: (a) testing combined interventions across different platforms and crisis contexts (e.g., pandemics, natural disasters, political elections); (b) examining interactions between psychological interventions and algorithmic changes, such as modifications to news feeds or recommendation systems; (c) developing personalized interventions that account for individual profiles, including critical thinking skills, worldview orientations, and levels of institutional trust; (d) investigating long-term

resistance effects as well as potential unintended consequences; and (e) analyzing the social dynamics of disinformation communities, including mechanisms of social influence and the spiral of silence. Advancing interdisciplinary research that integrates social psychology, information science, network analysis, and infodemic management will be essential for refining evidence-based strategies to counter disinformation in future crises.

6. Implications and further research

Summarizing the experimental results, the study demonstrates that repeated exposure to messages, their alignment with group identity, and negative emotional tone—both independently and in combination—increase belief in disinformation and intentions to share it, forming a “trivium” of cognitive, social, and emotional mechanisms (Pennycook & Rand, 2022). These effects are further intensified by algorithmic systems that preferentially deliver sensationalized and identity-congruent content, thereby fostering information bubbles and polarization. The combination of inoculative messages with accuracy prompts emerged as the most effective countermeasure, significantly reducing both belief in and virality of disinformation while producing sustained effects over time (Pennycook & Rand, 2022; Roozenbeek *et al.*, 2020). In the long term, these findings suggest that comprehensive preventive interventions that simultaneously teach recognition of manipulation and activate critical thinking represent the most promising direction for mitigating disinformation.

Taken together, the findings highlight the value of a combined intervention model that integrates cognitive, social, emotional, and algorithmic dimensions of misinformation dynamics. Rather than treating repetition effects, identity-based reasoning, and emotional framing as isolated drivers, the study demonstrates that their interaction yields more robust and durable improvements in belief accuracy and sharing behavior. The superior effectiveness of the combined prebunking and accuracy-prompt intervention underscores the practical importance of preventive and continuous strategies, which outperform reactive fact-checking approaches applied only after misinformation has already spread. This integrated framework provides a coherent basis for strengthening crisis communication and enhancing societal informational resilience under conditions of sustained uncertainty.

The results also have important practical and theoretical implications. They suggest that strategic communication and infodemic management should shift from reliance on isolated fact-checking toward systemic preventive approaches that combine inoculation materials with cognitive prompts and are implemented on an ongoing basis. The identified moderators further emphasize the need for personalized interventions: audiences characterized by lower cognitive engagement or stronger conspiratorial beliefs may require more intensive and adaptive strategies, while particular attention should be given to demographic groups identified as more vulnerable to disinformation. At a theoretical level, the findings update classical theories of rumor, cognitive dissonance, mass opinion, and the spiral of silence by demonstrating their continued relevance in the digital information environment and by integrating them with contemporary concepts of fluency, motivated reasoning, and algorithmic amplification.

Finally, the study raises important policy and educational questions concerning institutional readiness to implement sustained, combined interventions at scale. While experimental evidence supports the effectiveness of integrated prebunking and accuracy-based strategies, their real-world impact depends on whether public institutions, educational systems, and digital platforms are prepared to move beyond ad hoc fact-checking toward continuous, preventive communication models. Translating these insights into practice requires embedding psychological resilience tools into media literacy curricula, incorporating accuracy prompts and inoculation mechanisms into platform design, and developing regulatory frameworks that promote transparency and responsibility in algorithmic content distribution. Particular attention should be given to vulnerable demographic groups to ensure that interventions are inclusive, adaptive, and context-sensitive. Addressing these challenges is essential for transforming experimental findings into sustainable strategies that strengthen societal informational resilience.

Acknowledgment: Not applicable.

References

- Ahmed, S., & Rasul, M. E. (2022). Social media news use and COVID19 misinformation engagement: Survey study. *Journal of Medical Internet Research* 24(9), e38944. DOI: 10.2196/38944
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31(2), 211-236. DOI: 10.1257/jep.31.2.211
- Allport, G. W., & Postman, L. (1947). *The psychology of rumor*. H. Holt & Company. <https://archive.org/details/in.ernet.dli.2015.221801>
- Bessarabova, E., Banas, J. A., Reinikainen, H., Talbert, N., Luomaaho, V., & Tsetsura, K. (2024). Assessing inoculation's effectiveness in motivating resistance to conspiracy propaganda in Finnish and United States samples. *Frontiers in Psychology* 15, 1416722. DOI: 10.3389/fpsyg.2024.1416722
- Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring belief in conspiracy theories: The Generic Conspiracist Beliefs Scale. *Frontiers in Psychology*, 4, 279. DOI: 10.3389/fpsyg.2013.00279
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology* 42(1), 116-131. DOI: 10.1037/0022-3514.42.1.116
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment* 48(3), 306-307. DOI: 10.1207/s15327752jpa4803_13
- Dunning, D., & Kruger, J. (1999). Unskilled and unaware of it: How difficulties in recognizing one's incompetence lead to inflated selfassessments. *Journal of Personality and Social Psychology* 77(6), 1121-1134. DOI: 10.1037/00223514.77.6.1121
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., & Fazio, L. K. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology* 1, 13-29. DOI: 10.1038/s44159-021-00006-y
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press. <https://www.sup.org/books/sociology/theory-cognitive-dissonance>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science* 359(6380), 1094-1096. DOI: 10.1126/science.aa02998
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "posttruth" era. *Journal of Applied Research in Memory and Cognition* 6(4), 353-369. DOI: 10.1016/j.jarmac.2017.07.008
- Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., ... Zaragoza, M. S. (2020). *The Debunking Handbook 2020*. Databrary. DOI: 10.17910/b7.1182

- Lutzke, L., Drummond, C., Slovic, P., & Árvai, J. (2019). Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change* 58, 101964. DOI: 10.1016/j.gloenvcha.2019.101964
- Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications* 5(1), 47. DOI: 10.1186/s41235-020-00252-3
- McGuire, W. J. (1961). The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion. *Sociometry* 24(2), 184–197. DOI: 10.2307/2786067
- Ministry of Health of Ukraine. (2023). Order of the Ministry of Health of Ukraine dated 06.12.2023 No. 2076 “On Approval of the Protocol for Crisis Communications during Response to Cyberattacks and Cyber Incidents”. <https://moz.gov.ua/uk/decrees/nakaz-moz-ukraini-vid-06122023--2076-prozatverdzhennja-protokolu-krizovih-komunikacij-pid-chas-reaguvannja-na-kiberatiki-ta-kiberincidenti>
- Munusamy, S., Syasyila, K., Abu Hassan Shaari, A., Pitchan, M. A., Kamaluddin, M. R., & Jatnika, R. (2024). Psychological factors contributing to the creation and dissemination of fake news among social media users: A systematic review. *BMC Psychology* 12, 673. DOI: 10.1186/s40359-024-02129-2
- Nelson, T., Kagan, N., Critchlow, C., Hillard, A., & Hsu, A. (2020). The danger of misinformation in the COVID19 crisis. *Missouri Medicine* 117(6), 510–512. <https://pmc.ncbi.nlm.nih.gov/articles/PMC7721433/>
- Noelle-Neumann, E. (1974). The spiral of silence: A theory of public opinion. *Journal of Communication* 24(2), 43–51. DOI: 10.1111/j.1460-2466.1974.tb00367.x
- Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications* 13, 2049. DOI: 10.1038/s41467-022-30073-5
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature* 592(7855), 590–595. DOI: 10.1038/s41586-021-03344-2
- Prike, T., & Ecker, U. K. H. (2023). Effective correction of misinformation. *Current Opinion in Psychology* 54, 101712. DOI: 10.1016/j.copsyc.2023.101712
- Roozenbeek, J., & van der Linden, S. (2018). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research* 22(5), 570–580. DOI: 10.1080/13669877.2018.1443491
- Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications* 5, 65. DOI: 10.1057/s41599-019-0279-9
- Roozenbeek, J., van der Linden, S., & Nygren, T. (2020). Prebunking interventions based on inoculation theory can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review* 1(1). DOI: 10.37016/mr-2020-008
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2017). Defining “fake news”: A typology of scholarly definitions. *Digital Journalism* 6(2), 137–153. DOI: 10.1080/21670811.2017.1360143
- Ulbricht, B., & Rizk, J. (2024). How harmful information on social media impacts people affected by armed conflict: A typology of harms. *International Review of the Red Cross* 106, 823–862. DOI: 10.1017/S1816383124000572
- Udry, J., & Barber, S. J. (2023). The illusory truth effect requires semantic coherence across repetitions. *Cognition* 241, 105607. DOI: 10.1016/j.cognition.2023.105607
- Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges* 1(2), 1600008. DOI: 10.1002/gch2.201600008
- Vellani, V., Zheng, S., Ercelik, D., & Sharot, T. (2023). The illusory truth effect leads to the spread of misinformation. *Cognition* 236, 105421. DOI: 10.1016/j.cognition.2023.105421
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science* 359(6380), 1146–1151. DOI: 10.1126/science.aap9559
- World Health Organization. (2020). Managing the COVID19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation (Policy brief, 23 Sept 2020). <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation>
- Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge University Press. DOI: 10.1017/CBO9780511818691

Appendix A

Extended description of the research methodology (see Appendix B in the text of the article)

- **Design:** 2×2×2 factorial (replication × congruence × intervention).
- **Sample:** 2034 participants, after selection – 1987. Mean age = 34.7 years; 51.8% women.
- **Interventions:** prebunking (video), precision cueing (text), combined intervention.
- **Procedure:** three waves with intervals of 48-72 hours and 14 days.
- **Variables:** truthfulness, intention to share, need for cognition, conspiratorial beliefs.
- **Statistical analysis:** multilevel linear models (R, lme4). Fixed effects: all factors and their interactions. Random effects: message and respondent levels.

Source: author’s own development of the methodology.

Appendix B

Instructions for participants

Example of an introductory text: “In this study, you will be asked to read several short messages from different areas (health, economy, security). After each message, please rate how true it is (1–7) and how likely you are to share it on social media (1–7). It is important to answer honestly – there are no right or wrong answers.”

Appendix C

Questionnaire and baseline data (empirical base)

Table C1: Example of questionnaire questions to measure moderators (moderators and main variables) (see Appendix A1 in the text of the article)

Question	Scale / answer options
Your age	Please indicate the number of years
Gender	<input type="checkbox"/> Male <input type="checkbox"/> Female <input type="checkbox"/> Other
Level of education	<input type="checkbox"/> Secondary <input type="checkbox"/> Higher <input type="checkbox"/> Other
Need for Cognition Short Scale (Cacioppo & Petty, 1982, adapted)	“I like tasks that require a lot of thinking” (1 = strongly disagree ... 7 = strongly agree) “I enjoy solving problems that make me think.” (1 = strongly disagree ... 7 = strongly agree)
Conspiratorial beliefs (Generic Conspiratorial Beliefs Scale; Brotherton et al., 2013, adapted)	“Many important events in the world are controlled by secret groups” (1–7) “The government hides the real causes of important events from citizens.” (1–7) “Many world decisions are made by secret groups. (1–7)

Source: adapted from Cacioppo and Petty (1982; 1984), Brotherton et al. (2013)

Table C2: Example of information messages (stimulus material) (see Appendix A2 in the text of the article)

Type of message	Example text (in Ukrainian)	Category
Fake news	“New flu strain created in the laboratory to reduce birth rate”	Health
Fake news	“The government is hiding from the population plans to introduce mandatory confiscation of savings”	Economy
True	“The UN approved a resolution to continue humanitarian aid to Ukraine”	Security
True	“The Ministry of Health has introduced a national vaccination program”	Health
Fake	“The government plans to confiscate private savings in the coming years”	Economy

Source: adapted by the author from open public sources

Table C3: Average values of truthfulness assessment by groups (see Appendix A3 in the text of the article)

Condition	N	M (truthfulness)	SD
Control	500	3.45	0.88
Pre-banking	500	3.00	0.85
Accurate tip	500	3.05	0.82
Combined intervention	500	2.75	0.80

Source: Own empirical data (N≈2000)

Table C4: Intention to disseminate messages (mean values, scale 1–7) (see Appendix A4 in the text of the article)

Condition	One-time exposure	Multiple exposure
Control	3.20	3.60
Prebunking	2.80	3.10
Accurate tip	2.85	3.05
Combined intervention	2.55	2.85

Source: Own empirical data (N≈2000)

Appendix D

Visualizations of the results

Figure D1. Mean values of truthfulness scores by condition (bar chart)

Figure D2. Intention to share by condition (boxplot)
Source: author's own data and visualizations