Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

Review

Interventions to counter misinformation: Lessons from the Global North and applications to the Global South

Robert A. Blair[1], Jessica Gottlieb[2], Brendan Nyhan[3], Laura Paler[4], Pablo Argote[5] and Charlene J. Stainfield[6]

Abstract

We synthesize evidence from 176 experimental estimates of 11 interventions intended to combat misinformation in the Global North and Global South, which we classify as informational, educational, sociopsychological, or institutional. Among these, we find the most consistent positive evidence for two informational interventions in both Global North and Global South contexts: inoculation/prebunking and debunking. In a complementary survey of 138 misinformation scholars and practitioners, we find that experts tend to be most optimistic about interventions that have been least widely studied or that have been shown to be mostly ineffective. We provide a searchable database of misinformation randomized controlled trials and suggest avenues for future research to close the gap between expert opinion and academic research.

Addresses

[1] Department of Political Science and Watson Institute for International and Public Affairs, Brown University, United States
[2] Hobby School of Public Affairs, University of Houston, United States
[3] Department of Government, Dartmouth College, United States
[4] Department of Government, School of Public Affairs, American University, United States
[5] Department of Political Science and International Relations, University of Southern California, United States
[6] Department of Political Science, Ohio State University, United States

Corresponding author: Nyhan, Brendan (Brendan.J.Nyhan@dartmouth.edu)

Current Opinion in Psychology 2024, **55**:101732

This review comes from a themed issue on **The Psychology of Misinformation (2024)**

Edited by **Gordon Pennycook** and **Lisa K. Fazio**

For complete overview about the section, refer The Psychology of Misinformation (2024)

Available online 9 November 2023

https://doi.org/10.1016/j.copsyc.2023.101732

Keywords:
Misinformation, Disinformation, Randomized controlled trials, Experiments, Expert survey, Global South, Global North.

Introduction

Research on misinformation and interventions designed to combat it has proliferated in recent years.[1] Numerous literature reviews have already been published, but these focus almost exclusively on research from the Global North [2–8]. By contrast, the literature in the Global South remains limited, despite the pervasiveness and severity of misinformation there.[2] This evidence gap is especially important because interventions that effectively reduce misinformation in the Global North may perform differently in the Global South due to differences in factors such as news consumption, media literacy, state capacity, etc.

In this article, we synthesize evidence from 176 experimental estimates of the effects of interventions intended to combat misinformation that were conducted in both the Global North and Global South. These estimates are drawn from 155 unique randomized controlled trials (RCTs) identified in our literature review.

Specifically, we evaluate 11 prominent interventions, which we classify into four categories. Interventions in the first three categories aim to reduce consumers' propensity to believe or share misinformation, including the following: *informational interventions*, which provide corrective information to counter specific pieces of misinformation; *educational interventions*, which seek to impart skills that will make consumers less susceptible to misinformation more generally; and *sociopsychological interventions*, which use priming or appeals to social identity and social costs to discourage the spread of misinformation. Interventions in the fourth category, *institutional interventions*, instead seek to change the behavior of the producers and distributors of misinformation, including platforms, politicians, and journalists.[3]

Because of the lack of evidence from the Global South on the effectiveness of many of these interventions, we

[1] Misinformation is defined as purportedly factual claims that are either demonstrably false or widely contradicted by existing evidence and expert opinion, though definitional challenges remain [1].

[2] We use the term "Global South" to refer to developing countries in East Asia, Latin America, the Middle East and North Africa, South Asia, and Sub-Saharan Africa.

[3] See the Online Appendix for detailed definitions of each intervention.

complement our literature review with an original survey of 138 experts, including scholars, policymakers, and practitioners with expertise on misinformation and/or governance in the Global North, Global South, or both. This survey allows us to elicit expert opinion on the interventions that are most and least likely to be effective in the Global South.

We make several contributions. First, we identify two informational interventions—inoculation/prebunking and debunking—that effectively reduce belief in misinformation in both the Global North and Global South. Experts, however, are most enthusiastic about platform alterations, media literacy, and journalist training—institutional interventions that have been the least widely studied or for which the (limited) evidence base is decidedly mixed (e.g., media literacy). We suggest several avenues for future research to reconcile the mismatch between expert opinion and academic research. Finally, we provide readers with a separate, more detailed literature review [9] and a searchable database of the studies we considered.[4] The database can be filtered using characteristics of each study and the context in which it was conducted.

Scope and methodology

Our synthesis of the evidence began with a search for literature reviews and meta-analyses, which we used to formulate key search terms, identify specific interventions, and sort those interventions into broader categories. Using the key search terms, we then compiled a comprehensive list of relevant studies from Google Scholar and Elicit, restricting the scope of the analysis to RCTs. To account for recent literature reviews and meta-analyses of the most heavily studied interventions, we imposed year restrictions for select searches. (See the Online Appendix for more details on our methodology.)

Table 1 summarizes the evidence base for 11 of the most prominent interventions grouped by intervention type and region. Notably, our assessments of these studies are not based on a formal (quantitative) meta-analysis of the literature but instead are qualitative judgments made by the authors.[5]

Findings

Literature review

As shown in Table 1, available evidence on combating misinformation is heavily tilted toward the Global North. Of the 155 studies included in this review, over 80% were conducted in one or more Global North countries. This imbalance highlights the challenges of drawing conclusions about effective strategies for countering misinformation in the Global South from existing research. Likewise, there is substantial variation in the strength of the evidence base across leading interventions. Some interventions have been relatively well studied—for example, 56 of the 155 unique studies we reviewed test the effectiveness of debunking. But other potentially promising interventions such as sociopsychological "frictions" are relatively understudied. Institutional interventions, which arguably hold the greatest promise because they aim to reduce the supply of misinformation, have been studied the least, with no studies on politician messaging or journalist training in Global South countries to date. We summarize the findings for each intervention in the following and in Table 2. (See our report for a more detailed review of the findings for each intervention [9].

Among informational interventions, there is strong evidence that inoculation/prebunking and debunking are generally effective at countering misinformation in both Global North and Global South countries [10–20]. Together, these two informational interventions are the most widely studied among all treatments included in this review. In direct comparisons, debunking is more effective than inoculation/prebunking, and inoculation/prebunking appears to be more effective than credibility labels, although relatively few direct comparisons have been conducted to date [21–25]. These effects are strongest in the short term, however, and are typically measured only among people who are exposed to the debunking or inoculation/prebunking treatment (thus abstracting away from dissemination challenges in the real world).

The sole educational intervention we consider is media literacy, one of the most popular strategies for combating misinformation.[6] The evidence on its effectiveness is mixed. Though there are relatively few studies, many are of especially high quality, including eight studies conducted in five Global South countries. The majority of studies on media literacy in the Global North find no effect on misinformation beliefs [26–28]. Findings in the Global South are mixed, with some evidence that media literacy works but only among populations with high baseline literacy and education levels [29–31].

Sociopsychological interventions that alter social norms around misinformation have potential, although more evidence is needed. Both accuracy prompts and friction interventions tend to increase discernment between true and false information [32–39]. There is compelling evidence from the Global North that social norms interventions, which typically use messaging from in-group members to change perceived norms around sharing or believing misinformation, are generally

[4] Database: https://www.democratic-erosion.com/briefs/misinformation-intervention-database/.

[5] As noted earlier, data on the studies we identified are available online in a searchable format that allows filtering by study- or context-specific indicators: https://www.democratic-erosion.com/briefs/misinformation-intervention-database/.

[6] As we discuss in the Online Appendix, technique inoculation can also be viewed as an educational intervention in many cases.

Table 1

Summary of intervention types.

Category	Intervention	Studies cited	Global North	Global South
Informational	Inoculation/prebunking	25	18	7
Informational	Debunking	56	49	9
Informational	Credibility labels/tags	24	23	1
Informational	Contextual labels/tags	8	7	1
Educational	Media literacy	16	9	8
Sociopsychological	Accuracy prompts	13	11	3
Sociopsychological	Frictions	3	3	0
Sociopsychological	Social norms	14	11	3
Institutional	Platform alterations	10	8	2
Institutional	Politician messaging	4	4	0
Institutional	Journalist training	3	3	0

Note: The "Global North" and "Global South" columns do not always sum to the "Studies cited" column because some studies contain evidence from both regions.

effective at reducing both misinformed beliefs and sharing intentions [40–43]. While the Global South literature on social norms interventions is still small and has found mixed results [44], these interventions represent a promising avenue for future research because of the relative strength of community ties in many Global South contexts.

Institutional interventions, which instead target the supply of misinformation, have been the least studied, with few Global North and no Global South studies to date on politician messaging or journalist training. Interventions that aim to alter the supply of misinformation by platforms, politicians, or journalists can be harder to implement and evaluate than those discussed earlier because they often require behavioral change by high-level actors who might lack the will or ability to reduce misinformation supply. The small number of existing Global North studies nevertheless suggests that interventions involving deplatforming [45], deactivating from platforms [46], warning politicians their messages will be fact-checked [47–49], and training or incentivizing journalists to fact-check [50] are promising (though evidence is mixed). There remains an urgent need to improve the evidence base on these interventions for the Global South.

While all intervention types aim to reduce belief in or sharing of misinformation, outcomes are not always measured in the same way, limiting the extent to which we can make direct comparisons. Table 3 presents the frequency of outcomes measured across intervention types, with at least ten country studies each.[7] Accuracy prompt studies typically measure the sharing of misinformation, whereas debunking and media literacy usually measure beliefs.

But even among studies that measure beliefs, there is another important distinction: some measure belief in true or false content, while others measure ability to discern between true and false content. We see some patterns across outcome measures: debunking and social norms studies more often measure belief in false content, whereas the majority of accuracy prompt, credibility tag, and inoculation/prebunking studies measure discernment. Given this correlation between interventions and outcome measures, we caution readers when making direct comparisons of effectiveness across intervention types. More helpfully, Table 3 highlights how different interventions may be better suited to address different parts of the misinformation problem. For instance, we may not expect that a simple accuracy prompt would change long-lasting beliefs, but it should reduce the impulse to share dubious content.

This review was motivated in part by a concern that interventions designed to combat misinformation are likely to work differently across contexts. For instance, populations in the Global South have, on average, lower levels of digital and media literacy, which could make them more vulnerable to misinformation but which could also potentially increase the impact of interventions to promote digital and media literacy. The information environment is also different: Global South consumers are more likely to use WhatsApp than Twitter (X) or Facebook. The capacity of media outlets and journalists is also lower, with important implications for the feasibility of interventions. To allow readers to make their own judgments about these issues, we provide an interactive database of the studies considered in this review that can be filtered by study- and country-level characteristics: https://www.democratic-erosion.com/briefs/misinformation-intervention-database/

In addition to assessing the efficacy of interventions in combating information, policymakers and practitioners

[7] This cross-tabulation was constructed using this data set: https://www.democratic-erosion.com/briefs/misinformation-intervention-database/.

Table 2

Summary of intervention effects for Global North and Global South.

Category	Intervention	Global North findings	Global South findings
Informational	Inoculation/prebunking	Generally effective in the short term, although evidence is mixed regarding technique inoculation. Inoculation games may be less effective than first thought.	Tends to be most successful when implemented over long stretches of time in partnership with local news entities.
Informational	Debunking	Generally effective at reducing misinformed beliefs but effects may vary depending on credibility of or trust in the correction source.	More effective when delivered or endorsed by a trusted or in-group source; mixed evidence regarding effect on behavioral intentions.
Informational	Credibility labels/tags	Most effective when labels/tags provide justification for their presence, provide clear true or false ratings, and reference expert fact-checkers.	One study shows the potential for labels to decrease individual sharing of misinformation; more evidence needed.
Informational	Contextual labels/tags	Mixed effects on belief and sharing intentions; more evidence needed.	One study shows reduced perceived credibility of misinformation; more evidence needed.
Educational	Media literacy	Limited evidence of effectiveness and substantial variation in durability.	Mixed evidence of effectiveness despite intensive nature of some interventions.
Sociopsychological	Accuracy prompts	Generally effective at increasing discernment, though effect sizes vary.	Sometimes effective in increasing discernment and reducing sharing but effects are small.
Sociopsychological	Friction	May increase truth discernment and reduce sharing intentions for false news; more evidence needed.	No evidence; studies needed.
Sociopsychological	Social norms	Generally effective, although effects may vary by ideology.	Broadly effective; more evidence needed to determine contextual variation.
Institutional	Platform alterations	Generally positive, though most evidence gathered from simulating platforms as opposed to actual platform change.	Experimental evidence needed.
Institutional	Politician messaging	Shows initial promise at reducing misinformation supply; more evidence needed.	No evidence; studies needed.
Institutional	Journalist training	Shows initial promise for combating misinformation; more evidence needed.	No evidence; studies needed.

may wish to consider other dimensions when selecting and designing interventions, including *feasibility* (how easily an intervention can be implemented), *scalability* (how easily an intervention can be expanded to reach a large number of people), and *durability* (how long effects persist beyond the immediate postintervention period). Few interventions score well on all four dimensions, so deciding which intervention type to implement will likely involve trade-offs along these dimensions. For instance, both debunking and accuracy prompts have shown evidence of impact in the Global North and the Global South, but the effects of debunking tend to be larger and more durable. Debunking messages, however, only tends to reduce beliefs in the false claim that they target and are more feasible in contexts with relatively high levels of media capacity and legitimacy to identify and refute misinformation. By contrast, accuracy prompts are more easily scaled because they can affect beliefs on a wide range of topics and are more feasible in places with weaker media infrastructure.

Expert survey

Given the paucity of evidence from Global South countries, we complemented our literature review with an expert survey conducted from March 21–April 26, 2023, of 138 academics, policymakers, and practitioners with expertise on governance and/or misinformation in the Global North, Global South, or both. The goal of the expert survey was to measure beliefs about which interventions would be most effective in the Global South.

We presented respondents with the following hypothetical allocation exercise: "imagine USAID has funds

Table 3

Outcome measures by intervention type.

	Accuracy prompt	Credibility tags	Debunking	Inoculation/prebunking	Media literacy	Social norms
What is measured?						
Beliefs	2	11	62	20	16	8
Sharing	27	4	1	1	1	4
Both	0	3	7	9	1	5
Other	0	2	1	1	4	0
Total	29	24	73	26	17	17
How measured?						
False content	1	3	54	8	5	11
True content	0	3	1	3	2	0
Discernment	28	18	18	15	10	6
Total	29	24	73	26	17	17

to allocate to programs aimed at combating misinformation in developing countries. Given this goal, how would you advise USAID to divide up [100 units of funding] across the 12 intervention types described below?"[8]

Figure 1 displays the average amounts that respondents chose to allocate to each of the intervention types disaggregated by role and region of expertise.

Notably, respondents of all types and areas of regional expertise systematically preferred institutional and educational interventions over informational or socio-psychological ones. This finding is especially striking, given that the evidence base is much stronger for the latter than for the former in both the Global North and Global South. In addition, where there is a reasonably sized evidence base for these interventions in the Global South, as in the case of media literacy, the evidence of their effectiveness is quite limited. Experts also tended to be most optimistic about interventions that are most difficult to study experimentally (e.g., platform alterations and journalist training).

Some differences across experts emerged, however. As Figure 1a shows, researchers were most optimistic about platform alterations, while practitioners were most optimistic about journalist training and media interventions. As per Figure 1b, respondents with expertise on the Global North (mainly researchers) were also most optimistic about platform alterations, whereas respondents with expertise on the Global South were most optimistic about media literacy, followed closely by journalist training and platform alterations.

Finally, as we show in Figure 2, the three most popular interventions among experts—media literacy, journalist training, and platform alterations—are also among the least studied (29 studies in total).

This total is roughly equivalent to the number of unique studies on inoculation/prebunking (25) or credibility labels (24) alone and is less than half the number of unique studies on debunking (56).

Promising areas for future research

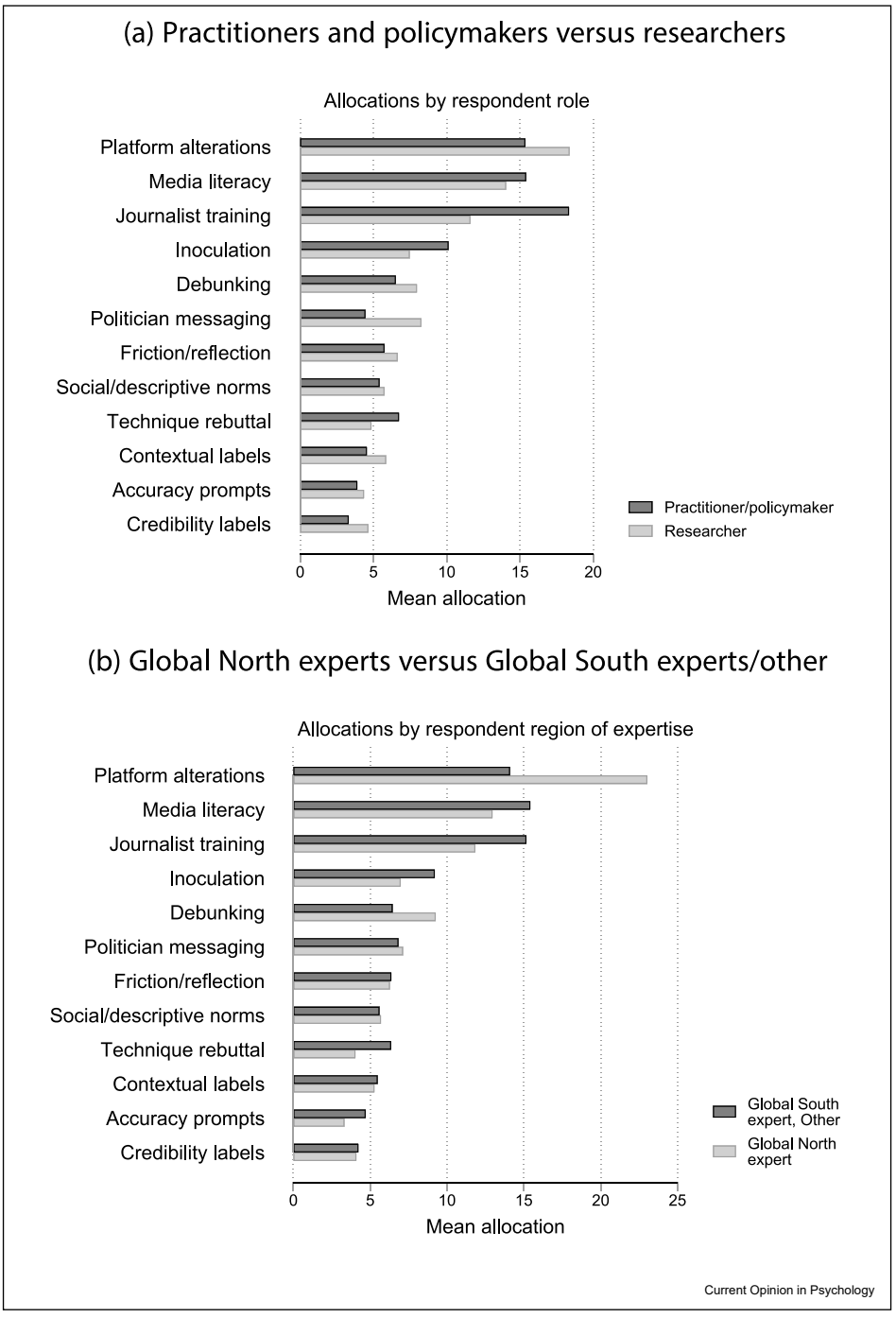
We conclude by suggesting some promising directions for future research. Given the current evidence gap, one strategy is to assess whether interventions are effective in both the Global North and Global South. The most rigorous (but also challenging and costly) way to do this is to test the same intervention in multiple countries simultaneously. Some researchers have already begun to take this approach, demonstrating its promise and feasibility [29,15,20,37].

Given that the characteristics of misinformation can differ by context, it is important to determine whether some types of misinformation are more or less susceptible to correction than others. However, most studies focus on a particular topic or type of misinformation. More researchers should consider incorporating multiple types of misinformation into their designs [51].

Third, scholars should consider how social norms and identities could contribute to curbing misinformation in their context of interest. For example, state weakness in many countries in the Global South leads to greater reliance on social institutions, such as ethnic or religious groups, whose leaders function as intermediaries between the state and its citizens. These institutions and intermediaries could be valuable partners in interventions to combat misinformation; for example, consumers may be more likely to believe information they receive from sources with whom they share political, racial, ethnic, or religious ties [52,53].

[8] In the expert survey, we listed "technique rebuttal" as a separate intervention type, but we subsequently decided to classify it as a form of either debunking or inoculation/prebunking.

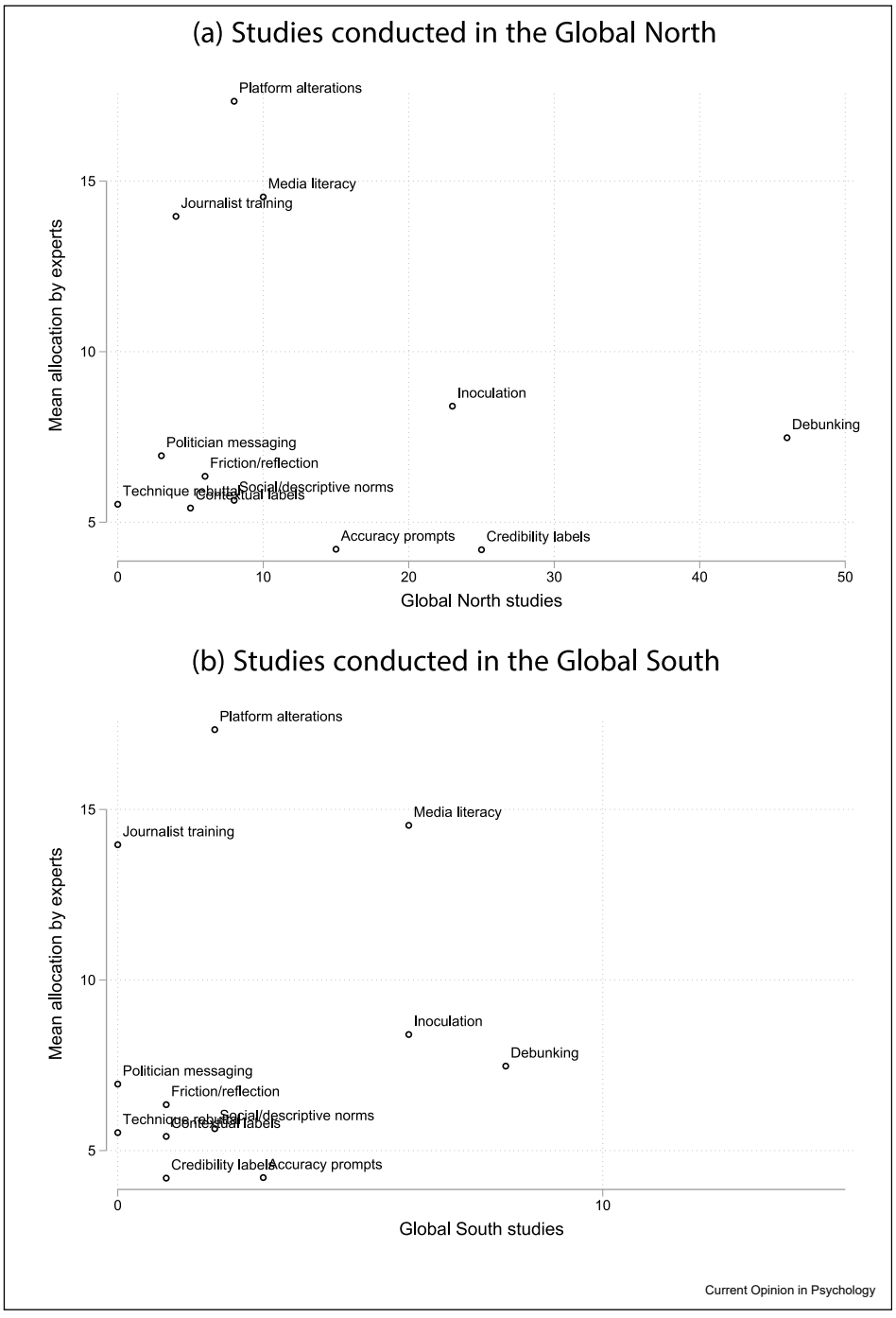
Figure 1



Mean allocations to each intervention type by respondent characteristics. Expert survey of 138 practitioners, policymakers, and academic researchers.

Finally, some interventions have been much more extensively studied than others. Not only is the balance of the evidence skewed more toward the Global North, but also little evidence exists about some important intervention types. Most notably, the three most popular interventions among experts—media literacy, journalist training, and platform alterations—are also among the least studied.

Figure 2



Comparing expert evaluations with quantity of evidence. Expert survey of 138 practitioners, policymakers, and academic researchers.

This finding highlights the fact that interventions that are the most feasible to rigorously study may not be the most promising in terms of scalability or durability—two features of interventions that policymakers and practitioners value. Focusing on interventions for which the evidence base is weak but

expected effectiveness is high in terms of impact, scale, feasibility, and durability strikes us as an especially important direction for future research.

Author contributions

PA and CJS conducted the literature review. RAB, JG, BN, LP, PA, and CJS wrote and edited the article.

Funding

This work was supported by the United States Agency for International Development (USAID) [GS-10F-0033M/7200AA18M00016, Tasking N067].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data from our literature review are available at https://www.democratic-erosion.com/briefs/misinformation-intervention-database/. Data and code to replicate our results from the expert survey can be found at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ZMGNG4.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.copsyc.2023.101732.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Vraga EK, Bode L: **Defining misinformation and understanding its bounded nature: using expertise and evidence for describing misinformation**. *Polit Commun* 2020, **37**, https://doi.org/10.1080/10584609.2020.1716500.

2. Nyhan B: **Facts and myths about misperceptions**. *J Econ Perspect* 2020, **34**:220–236, https://doi.org/10.1257/jep.34.3.220.

3. Nyhan B: **Why the backfire effect does not explain the durability of political misperceptions**. *Proc Natl Acad Sci USA* 2021, **118**, e1912440117, https://doi.org/10.1073/pnas.1912440117.

4. Pennycook G, Rand DG: **The psychology of fake news**. *Trends Cognit Sci* 2021, **25**:388–402. https://psycnet.apa.org/doi/10.1016/j.tics.2021.02.007.

5. Ecker UKH, Lewandowsky S, Cook J, Schmid P, Fazio L, Brashier N, Kendeou P, Vraga EK, Amazeen MA: **The psychological drivers of misinformation belief and its resistance to correction**. *Nat Rev Psychol.* 2022, **1**, https://doi.org/10.1038/s44159-021-00006-y.

6. Johansson P, Enock F, Hale S, Vidgen B, Bereskin C, Margetts H, Bright J: **How can we combat online misinformation? A systematic overview of current interventions and their efficacy.** *arxiv:2212.11864 [cs.CY]* 2022. https://doi.org/10.48550/arXiv.2212.11864; 2022.

7. Kozyreva A, Lorenz-Spreen P, Herzog S, Ecker UKH, Lewandowsky S, Hertwig R, Basol M, Berinsky AJ, Betsch C, Cook J, *et al.*: **Toolbox of interventions against online misinformation and manipulation**. *PsyArXiv* 2022. http://doi.org/10.31234/osf.io/x8ejt; 2022.

8. van der Linden S: **Misinformation: susceptibility, spread, and interventions to immunize the public**. *Nat Med* 2022, **28**:460–467, https://doi.org/10.1038/s41591-022-01713-6.

9. Blair RA, Gottlieb J, Nyhan B, Paler L, Argote P, Stainfield CJ: **Interventions to counter misinformation: lessons from the Global North and applications to the Global South**. *USAID Development Experience Clearinghouse* 2023. https://pdf.usaid.gov/pdf_docs/PA0215JW.pdf; 2023.

10. Toni GLA, Meer VD, Jin Y: **Seeking formula for misinformation treatment in public health crises: the effects of corrective information type and source**. *Health Commun* 2020, **35**:560–575, https://doi.org/10.1080/10410236.2019.1573295.

11. Nyhan B, Porter E, Reifler J, Wood TJ: **Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability**. *Polit Behav* 2020, **42**:939–960, https://doi.org/10.1007/s11109-019-09528-x.

12. Bowles J, Larreguy H, Liu S: **Countering misinformation via WhatsApp: preliminary evidence from the COVID-19 pandemic in Zimbabwe**. *PLoS One* 2020, **15**:1–11, https://doi.org/10.1371/journal.pone.0240005.

13. Wang Y: **Debunking misinformation about genetically modified food safety on social media: can heuristic cues mitigate biased assimilation?** *Sci Commun* 2021, **43**:460–485, https://doi.org/10.1177/10755470211022024.

14. Clayton K, Finley C, Flynn DJ, Graves M, Nyhan B: **Evaluating the effects of vaccine messaging on immunization intentions and behavior: evidence from two randomized controlled trials in Vermont**. *Vaccine* 2021, **39**:5909–5917, https://doi.org/10.1016/j.vaccine.2021.08.047.

15. ** Porter E, Wood TJ: **The global effectiveness of fact-checking: evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom**. *Proc Natl Acad Sci USA* 2021, **118**:1–7, https://doi.org/10.1073/pnas.2104235118.
The authors describe fact-checking experiments conducted simultaneously in Argentina, Nigeria, South Africa, and the United Kingdom, in which, they studied whether fact-checking can durably reduce belief in misinformation. In total, 22 fact-checks were evaluated, including two that were tested in all four countries. Fact-checking reduced belief in misinformation, with most effects still apparent more than 2 weeks later. A meta-analytic procedure indicates that fact-checks reduced belief in misinformation by at least 0.59 points on a 5-point scale. Exposure to misinformation, however, only increased false beliefs by less than 0.07 points on the same scale.

16. Vivion M, Sidi EAL, Betsch C, Dionne M, Dubé E, Driedger SM, Gagnon D, Graham J, Greyson D, Hamel D, *et al.*: **Prebunking messaging to inoculate against COVID-19 vaccine misinformation: an effective strategy for public health**. *J Commun Healthc* 2022, **15**:232–242, https://doi.org/10.1080/17538068.2022.2044606.

17. Buczel M, Szyszka PD, Siwiak A, Szpitalak M, Polczyk R: **Vaccination against misinformation: the inoculation technique reduces the continued influence effect**. *PLoS One* 2022, **17**, e0267463, https://doi.org/10.1371/journal.pone.0267463.

18. Garg N, Yadav M: *Learning to resist misinformation: a field experiment in India*. 2022. https://assets.namangarg.net/papers/jmp_latest.pdf.

19. Bertolotti M, Catellani P: **Counterfactual thinking as a prebunking strategy to contrast misinformation on COVID-19**. *J Exp Soc Psychol* 2023, **104**, https://doi.org/10.1016/j.jesp.2022.104404.

20. Porter E, Velez YR, Wood TJ: **Correcting COVID-19 vaccine misinformation in 10 countries**. *R Soc Open Sci* 2023, **10**:1–12, https://doi.org/10.1098/rsos.221097.

21. Brashier NM, Pennycook G, Berinsky AJ, Rand DG, Levi M: **Timing matters when correcting fake news**. *Proc Natl Acad Sci USA* 2021, **118**:1–3, https://doi.org/10.1073/pnas.2020043118.

22. Tay LQ, Hurlstone MJ, Kurz T, Ecker UKH: *A comparison of prebunking and debunking interventions for implied versus explicit*. 2022, https://doi.org/10.1111/bjop.12551.

23. Amazeen MA, Krishna A, Eschmann R: **Cutting the bunk: comparing the solo and aggregate effects of prebunking and debunking covid-19 vaccine misinformation**. *Sci Commun* 2022, **44**:387–417, https://doi.org/10.1177/10755470221111558.

24. McPhedran R, Ratajczak M, King E: **Inoculation against misinformation in Australia: a replication study**. *PsyArXiv* 2022. https://doi.org/10.31234/osf.io/87nza; 2022.

25. McPhedran R, Ratajczak M, Mawby M, King E, Yang Y, Gold N: **Psychological inoculation protects against the social media infodemic**. *Sci Rep* 2023, **13**:5780, https://doi.org/10.1038/s41598-023-32962-1.

26. Panizza F, Ronzani P, Martini C, Mattavelli S, Morisseau T, Motterlini M: **Lateral reading and monetary incentives to spot disinformation about science**. *Sci Rep* 2022, **12**:5678, https://doi.org/10.1038/s41598-022-09168-y.

27. Vraga E, Tully M, Bode L. *Assessing the relative merits of news literacy and corrections in responding to misinformation on Twitter*, **24**. New Media Soc; 2022:2354–2371, https://doi.org/10.1177/1461444821998691.

28. Qian S, Shen C, Zhang J: **Fighting cheapfakes: using a digital media literacy intervention to motivate reverse search of out-of-context visual misinformation**. *J COMPUT-MEDIAT COMM* 2023, **28**:1–12, https://doi.org/10.1093/jcmc/zmac024.

29. Guess AM, Lerner M, Lyons B, Montgomery JM, Nyhan B, Reifler J, Sircar N: **A digital media literacy intervention increases discernment between mainstream and false news in the United States and India**. *Proc Natl Acad Sci USA* 2020, **117**:15536–15545, https://doi.org/10.1073/pnas.1920498117.

30. Oberiri Destiny Apuke, Omar Bahiyah, Elif Asude Tunca: **Literacy concepts as an intervention strategy for improving fake news knowledge, detection skills, and curtailing the tendency to share fake news in Nigeria**. *Child Youth Serv* 2023, **44**:88–103, https://doi.org/10.1080/0145935X.2021.2024758.

31. * Ali A, Qazi IA: **Countering misinformation on social media through educational interventions: evidence from a randomized experiment in Pakistan**. *J Dev Econ* 2023, **163**:103108, https://doi.org/10.1016/j.jdeveco.2023.103108.
This study evaluates the effectiveness of two educational interventions to counter misinformation among low digital literacy populations in urban Pakistan using a randomized experiment. Authors find no evidence that video-based general educational messages about misinformation have any statistically significant impact on the ability to correctly identify news items as true or fake. However, when such messages are augmented with personalized feedback based on individuals' past engagement with fake news, the authors find an improvement of 3.3 percentage points or 4.5 percent in correctly identifying news (true or fake) relative to the control group. Results suggest that educational interventions can enable information discernment but that their effectiveness depends on how well their features and delivery are customized for the population of interest.

32. Pennycook G, Epstein Z, Mosleh M, Arechar AA, Eckles D, Rand DG: **Shifting attention to accuracy can reduce misinformation online**. *Nature* 2021, **592**:590–595, https://doi.org/10.1038/s41586-021-03344-2.

33. Jahanbakhsh F, Zhang AX, Ajl Berinsky, Pennycook G, Rand DG, Karger DR: **Exploring lightweight interventions at posting time to reduce the sharing of misinformation on social media**. *Proc ACM Hum Comput Interact* 2021, **5**(1–42), https://doi.org/10.1145/3449092.

34. Capraro V, Celadin T: **"I think this news is accurate": endorsing accuracy decreases the sharing of fake news and increases the sharing of real news**. *Pers Soc Psychol Bull* 2022:1–11, https://doi.org/10.1177/01461672221117691.

35. Sharevski F, Alsaadi R, Jachim P, Pieroni E: **Misinformation warnings: Twitter's soft moderation effects on COVID-19 vaccine belief echoes**. *Comput Secur* 2022, **114**:102577, https://doi.org/10.1016/j.cose.2021.102577.

36. Epstein Z, Sirlin N, Arechar A, Pennycook G, Rand D: **The social media context interferes with truth discernment**. *Sci Adv* 2023, **9**:1–8, https://doi.org/10.1126/sciadv.abo6169.

37. * Arechar AA, Allen J, Berinsky AJ, Cole R, Garimella K, Gully A, Lu JG, Ross RM, Zhang Y, Pennycook G, *et al.*: **Understanding and combating misinformation across 16 countries on six continents**. *Nat Human Behav* 2023, **7**, https://doi.org/10.1038/s41562-023-01641-6.
Authors conducted an experiment in 16 countries across 6 continents (N = 33,480) to investigate predictors of susceptibility to misinformation and interventions to combat misinformation. In every country, participants with a more analytic cognitive style and stronger accuracy-related motivations were better at discerning truth from falsehood; valuing democracy was also associated with greater truth discernment, whereas political conservatism was negatively associated with truth discernment in most countries. Subtly prompting people to think about accuracy was broadly effective at improving the veracity of news that people were willing to share, as were minimal digital literacy tips. Finally, crowd-sourced accuracy evaluation was able to differentiate true from false headlines with high accuracy in all countries.

38. ** Offer-Westort M, Rosenzweig LR, Athey S: *Battling the coronavirus infodemic among social media users in Africa*. 2023, https://doi.org/10.48550/arXiv.2212.13638 [Unpublished results].
An experiment on Facebook Messenger with users from Kenya and Nigeria tested interventions designed to decrease intentions to share COVID-19 misinformation without decreasing intentions to share factual posts. The initial stage of the study incorporated (i) a factorial design with 40 intervention combinations and (ii) a contextual adaptive design, increasing the probability of assignment to treatments that worked better for previous subjects with similar characteristics. The second stage evaluated the best-performing treatments and a targeted treatment assignment policy estimated from the data. Tips to identify misinformation and nudges to consider information's accuracy reduced misinformation sharing by 4.2% and 4.9%, respectively.

39. ** Athey S, Cersosimo M, Koutout K, Li Z: *Emotion- versus reasoning-based drivers of misinformation sharing: a field experiment using text message courses in Kenya*. 2022, https://doi.org/10.2139/ssrn.4489759 [Unpublished results].
To learn whether interventions to counter reasoning- or emotion-based techniques are more effective, or whether the approaches are complementary, authors evaluate three distinct versions of a five-day, low-cost, and scalable text-message educational course in a field experiment with approximately 9000 participants in Kenya. They assess the impact of the courses through a pre–post survey design eliciting intentions to share, finding that all treatment courses work, decreasing misinformation sharing 28% on average relative to the no-text-message course. Treating the emotional drivers of misinformation sharing in the "Emotions" course is more effective than teaching about reasoning-based techniques either alone in the "Reasoning" course or in combination with emotion-based techniques in the "Combo" course. Moreover, the "Emotions" course performs best on misinformation posts that use emotion-based techniques, and it does no worse than the Reasoning or Combo courses on misinformation posts that use reasoning-based techniques. In a follow-up experiment, approximately two months later, 88% of the treatment effect of the three courses on misinformation sharing persists.

40. Gimpel H, Heger S, Olenberger C, Utz L: **The effectiveness of social norms in fighting fake news on social media**. *J Manag Inf Syst* 2021, **38**:196–221, https://doi.org/10.1080/07421222.2021.1870389.

41. Ecker UKH, Sanderson JA, McIlhiney P, Rowsell JJ, Quekett HL, Brown GDA, Lewandowsky S: **Combining refutations and social norms increases belief change**. *Q J Exp Psychol* 2022, **76**:1275–1297, https://doi.org/10.1177/17470218221111750.

42. Orosz G, Paskuj B, Farago L, Kreko P: **A prosocial fake news intervention with durable effects**. *Sci Rep* 2023, **13**:3958, https://doi.org/10.1038/s41598-023-30867-7.

43. Badrinathan S, Chauchard S: *"I don't think that's true, bro!" Social corrections of misinformation in India*. Int. J. Press Polit.; 2023, https://doi.org/10.1177/19401612231158770.

44. * Gottlieb J, Adida C, Moussa R: *Reducing misinformation in a polarized context: experimental evidence from Côte d'Ivoire*. 2022, https://doi.org/10.31219/osf.io/6x4wy [Unpublished results].
This study proposes three interventions that leverage insights on how social identity shapes behavior and evaluates each with an information experiment in Côte d'Ivoire, a polarized country. The authors find that a standard digital literacy intervention fails to curb the belief in and spread of misinformation, while their social-identity-based interventions limits both. The findings confirm that misinformation spreads at least in part because individuals are motivated to consume biased media, and they caution against strategies that ignore these directional motives.

45. Thomas DR, Wahedi LA: **Disrupting hate: the effect of deplatforming hate organizations on their online audience**. *Proc Natl Acad Sci USA* 2023, **120**, e2214080120, https://doi.org/10.1073/pnas.2214080120.

46. Ventura T, Majumdar R, Nagler J, Tucker JA: **WhatsApp increases exposure to false rumors but has limited effects on beliefs and polarization: evidence from a multimedia-constrained deactivation**. *SSRN* 2023. http://dx.doi.org/10.2139/ssrn.4457400; 2023.

47. Nyhan B, Reifler J: **The effect of fact-checking on elites: a field experiment on U.S. State legislators**. *Am J Polym Sci* 2015, **59**: 628–640, https://doi.org/10.1111/ajps.12162.

48. Diermeier M: **Tailoring the truth – evidence on parliamentarians' responsiveness and misinformation toleration from a field experiment**. *Eur Political Sci Rev* 2023, **15**:1–21, https://doi.org/10.1017/s1755773922000613.

49. ** Ma S, Bergan D, Ahn S, Carnahan D, Gimby N, McGraw J, Virtue I: **Fact-checking as a deterrent? A conceptual replication of the influence of fact-checking on the sharing of misinformation by political elites**. *Hum Commun Res* 2023, **49**: 321–338, https://doi.org/10.1093/hcr/hqac031.
The current study presents a conceptual replication and extension of Nyhan and Reifler (2015) by utilizing a similar treatment that leverages a recent partnership between local media outlets and fact-checking organizations, assessing the effects of the treatment on the accuracy of legislators' statements on Twitter around the first impeachment trial of Donald Trump. Results provide limited evidence of the effects of treatment on the accuracy of legislators' posts, even among legislators within media markets directly affected by this partnership.

50. Graves L, Nyhan B, Reifler J: **Understanding innovations in journalistic practice: a field experiment examining motivations for fact-checking**. *J Commun* 2016, **66**:102–138, https://doi.org/10.1111/JCOM.12198.

51. Bowles J, Croke K, Larreguy H, Liu S, Marshall J: **Sustained exposure to fact-checks can inoculate citizens against misinformation**. *SSRN* 2023. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4582703; 2023.

52. Armand A, Kameshwara KK, Bancalari Q, Augsburg B: **Countering misinformation with targeted messages: experimental evidence using mobile phones**. *EconStor* 2021. http://hdl.handle.net/10419/267949; 2021.

53. Pretus C, Javeed A, Hughes DR, Hackenburg K, Tsakiris M, Vilarroya O, Van Bavel JJ: **The Misleading count: an identity-based intervention to counter partisan misinformation sharing**. *PsyArXiv* 2022. https://doi.org/10.31234/osf.io/7j26y; 2022.