

Under the Hood of *The Debunking Handbook 2020*:  
A consensus-based handbook of recommendations  
for correcting or preventing misinformation

by

Stephan Lewandowsky, John Cook, and Ullrich K. H. Ecker

[stephan.lewandowsky@bristol.ac.uk](mailto:stephan.lewandowsky@bristol.ac.uk)

**Citation for the *Handbook*:**

Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E. Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., Vraga, E. K., Wood, T. J., Zaragoza, M. S. (2020). *The Debunking Handbook 2020*. Available at <https://sks.to/db2020>. DOI: 10.17910/b7.1182

**Purpose of this document:**

This document provides an overview of the process by which *The Debunking Handbook 2020* was created during May – October 2020. The new *Handbook* replaces the first *Debunking Handbook* (<http://sks.to/debunk>), published in 2011 by John Cook and Stephan Lewandowsky. The scientific knowledge about debunking has shifted considerably during the decade since publication of the first *Handbook*. (The evolution of the scientific evidence was already reflected in an update notice of the first *Handbook* published in June 2017 at <https://www.skepticalscience.com/Debunking-Handbook-now-freely-available-download.html#UpdateJune2017> ).

In addition to updating the science of debunking, *The Debunking Handbook 2020* also sought to represent the scientific consensus on debunking as of 2020. It was therefore developed using a preregistered approach that was, in part, inspired by research on the process of consensus formation in a medical/clinical context (e.g., Rosenfeld, Nnacheta, & Corrigan, 2015). Our approach was also informed by precedents in psychology, for example, the development and report of an expert consensus for the psychology of eyewitness testimony (Kassin, Tubb, Hosch, & Memon, 2001). Our approach departed from precedents by preregistering our methodology and intended completion schedule.

The remainder of this document summarizes the steps of document production, including the data that were generated along the way.

## The creation of the *Handbook*:

### 1. Selection of authors

Selection of authorship is crucial for creation of a consensus document. We adopted a two-tiered model from the Intergovernmental Panel on Climate Change (IPCC), which differentiates between coordinating lead authors (CLA) and contributing authors. The CLAs for *The Debunking Handbook 2020* were Stephan Lewandowsky, John Cook, and Ullrich K. H. Ecker, who originated the project and provided managerial oversight throughout.

The CLAs then assembled a team of contributing authors using the following three steps: (a) *Scopus* was searched for papers published from 2015 onwards (on 15 May 2020) using these keys: (misinformation OR disinformation OR "fake news" OR misconceptions OR myths OR "false information") AND (debunking OR corrections OR refutations OR retractions OR inoculation OR prebunking OR fact-checking OR "fact checking" OR backfire OR "belief updating" OR "knowledge revision" OR "belief echoes" OR backlash). All subject areas were excluded from the search except psychology, social sciences, arts and humanities, and "multidisciplinary." From the resulting list of published authors, researchers with only 1 publication and individuals who were known to be inactive were excluded. (b) The CLAs also independently generated other candidate names based on their own knowledge of the field. These steps yielded a preliminary list of 33 names. (c) This list was then discussed among CLAs and 5 more names were excluded because they had either published only theoretical papers or only preprints. The remaining 28 individuals were invited by personal email in late May 2020 to participate in the project, and those who accepted joined the team as contributing authors. The team was assembled in early June 2020. The initial team comprised 21 contributing authors, two of whom withdrew because of over-commitment early on. The remaining 19 contributing authors are listed as co-authors of *The Debunking Handbook 2020*.

### 2. Preregistration

All authors jointly developed the intended process for document production, which was preregistered on 26 June 2020. The preregistration is now public and available at <http://osf.io/dnu7z>.

All steps in the preregistration were followed except where indicated in the following. All milestones and deadlines in the preregistration document were met, with the exception of final release of the document which was targeted for late September 2020 but was postponed until early October 2020.

### 3. Scope of *The Debunking Handbook 2020*

During preparation of the preregistration document, the team defined the preliminary scope of the document as providing “a usable practice guide situated within the present technological and political/regulatory context. This goal necessarily limits the scope to cognitive and behavioral measures that are actionable now (rather than, for example, proposing redesigned social media platforms). In addition, our goal is to provide context by acknowledging how our guide sits within the ‘attention economy,’ and addressing the power of social media platforms (and how that power could be exploited by changing platform architecture and policy overnight with greater consequences than any debunking guide can ever achieve).”

### 4. Methodology and results

**4.1 Disclosure statements.** All authors completed disclosure statements, detailing funding and potential conflicts of interest (e.g., consultancies, sponsorships, grants). Authors also revealed their involvement in advice or testimony to governments or similar organizations (e.g., WHO, E.U.), and contributions to public media. The disclosures are available at [https://drive.google.com/file/d/18m\\_z4pQFtR7yPNb37tZTcuazNbHzbSco/view?usp=sharing](https://drive.google.com/file/d/18m_z4pQFtR7yPNb37tZTcuazNbHzbSco/view?usp=sharing).

**4.2 Nomination of phenomena and recommendations.** Together with the disclosures, all authors independently nominated up to 10 issues (i.e., findings and phenomena) for inclusion in *The Debunking Handbook 2020*, and up to 5 recommendations (for practitioners) that should be made in the document. Each nomination was made by brief descriptions (~50 words), optionally accompanied by ~3 key references. Authors also indicated whether they considered a nominated finding to have emerged from their own work. Authors were encouraged to highlight work and to provide citations to work done by others in addition to their own research. All responses were collected using a Google Form. The nominated phenomena and recommendations are available at [https://drive.google.com/file/d/18m\\_z4pQFtR7yPNb37tZTcuazNbHzbSco/view?usp=sharing](https://drive.google.com/file/d/18m_z4pQFtR7yPNb37tZTcuazNbHzbSco/view?usp=sharing).

**4.3 Team rating of nominated phenomena and recommendations.** The individual nominations were consolidated into a single list by merging apparent duplicates into a single description. The merged list of phenomena ( $N=57$ ) and recommendations ( $N=39$ ) was then turned into a questionnaire on the Qualtrics platform. As recommended by Rosenfeld et al. (2015), the authors then rated each item on two dimensions, namely confidence and importance. The response scales were based on criteria for evidence-based recommendations in education (<http://www2.ed.gov/policy/elsec/leg/essa/guidanceuseininvestment.pdf>) but were changed slightly during team discussion.

The response scale for *confidence* was:

- Replicated evidence—supported by multiple experiments that are known to be replicable. (1)
- Strong evidence—supported by one or more experiments. (2)
- Moderate evidence—supported by one or more quasi-experimental studies. (3)
- Promising evidence—supported by one or more correlational studies. (4)
- Demonstrates a rationale—conceptual grounding exists; awaiting empirical support (5)

The response scale for *importance* was:

- Pivotal—finding is crucial to debunking. (1)
- Important —finding is helpful to debunking. (2)
- Worth considering—debunking may benefit from this finding. (3)
- Not important—debunking is unaffected by this finding. (4)

Optionally, authors could provide brief justifications for their choices in addition to the ratings. A printout of the complete survey is available at

<https://drive.google.com/file/d/1KdwtoF90iPiED9DG504DXvHs6sIVilmh/view?usp=sharing>.

Data collection was completed on 25 July 2020, with all authors contributing to the survey ( $N=22$ ). The data file is available at [https://docs.google.com/spreadsheets/d/1kbc-gppV\\_JXAWRFHDBwoh\\_HHbR9Z1BUdrInDSHAnFEQ/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1kbc-gppV_JXAWRFHDBwoh_HHbR9Z1BUdrInDSHAnFEQ/edit?usp=sharing).

Table 1 displays the list of items (phenomena and recommendations) together with their short names that are used in the figures to display the results.

Table 1. Short name (left column) used in graphs and full text (right column) of the phenomenon or recommendation. Short names prefixed by “Ph” refer to phenomena and those prefixed by “Rec” refer to recommendations.

Ph1	There is ideological symmetry in motivated reasoning processes.
Ph2	Motivated rejection of science may be overcome by increasing topic-specific knowledge.
Ph3	Repeated exposure to information, including false declarative statements and assertions, increases people's ratings of those ideas as true (“illusory truth” effect).
Ph4	People can acquire inaccurate understandings about the world even when that information is conveyed through fiction (and is known to be fictional).
Ph5	People may neglect their own general knowledge when assessing the veracity of a claim.
Ph6	Fake news is propagated more quickly and broadly than accurate news through social media networks.
Ph7	Partisan networks create echo chambers.

Ph8	Hyperpartisan news websites are an important source of misinformation.
Ph9	Ideologically congenial selective exposure is likely to reduce exposure to corrections after the misinformation is in place.
Ph10	The belief that one is capable of defending one's attitudes or beliefs may lead people to think of opposing messages as a challenge, and thus increase the likelihood of voluntary exposure to those messages.
Ph11	Information overload desensitizes people to mis/disinformation.
Ph12	People spend a short amount of time on news websites (and sometimes even share content prior to reading the associated article).
Ph13	Online source evaluations are challenging (but can be taught).
Ph14	Mis/disinformation is often more plausible or attention-grabbing than correct information—it has persuasive appeal.
Ph15	People do not routinely track and evaluate the credibility of sources—but when they do, impact of information from less credible sources can be reduced.
Ph16	Accuracy nudges and reminders increase the quality of people's sharing decisions on social media.
Ph17	Misinformation often continues to influence reasoning post-correction, and outdated information is not simply erased (the "continued influence effect").
Ph18	Effects of exposures to inaccurate information appear to last.
Ph19	Fact-checking and corrections work, at least in part and in many situations, but this does not mean fact-checking can eliminate all inaccurate beliefs.
Ph20	Updates to factual beliefs, even if successful, may not translate into attitude or behaviour change.
Ph21	Corrections are more effective if in addition to providing a simple retraction ("not true"), they propose a causal alternative, and generally if they provide substantive, relevant detail and establish coherence.
Ph22	Corrections are more effective if people are (made to be) suspicious of the source or intent of the misinformation.
Ph23	Corrections are more effective if they are repeated.
Ph24	Corrections are more effective if they involve multiple relevant counterarguments.
Ph25	Corrections are more effective if they come from trusted sources or highlight expert consensus.
Ph26	Expertise matters, especially for factual corrections.
Ph27	Source expertise does not matter much for corrections.
Ph28	Social media corrections which identify a distal source of the information are more effective at influencing beliefs than unsourced corrections.
Ph29	Detecting the inconsistency between the misinformation and the correction is a precondition for effective updating.

Ph30	Prompts to compare the correction with the initial belief or attitude can increase correction efficacy.
Ph31	A technique- or logic-based rebuttal has the advantage that it is based on the detection of generic violations of logic, and hence transfers outside a specific context.
Ph32	Well-designed graphs and videos can be helpful to convey corrections involving complex or “statistical” information clearly and concisely.
Ph33	Corrections with graphical elements tend to be less effective.
Ph34	Emotional content/language influences not only the spread of misinformation, but also corrective efforts.
Ph35	The language in some debunking texts (“vaccine denier”) runs the risk of stigmatizing specific groups and may thus enhance polarization.
Ph36	The efficacy of corrections depends in part on the recipient’s motivation to believe the statement.
Ph37	Warning people that they might be misinformed reduces later reliance on misinformation.
Ph38	“Prebunking” or inoculation is an alternative to debunking and the first line of defence against misinformation.
Ph39	Post-inoculation talk: when people spontaneously share debunking or inoculation messages with others after they receive them, they [UE1] [SL2] may spread “herd immunity” through a community.
Ph40	Failed attempts at correction are likely to strengthen belief in the initial misinformation.
Ph41	The familiarity backfire effect—corrections that repeat misinformation can ironically strengthen misconceptions—is not a robust phenomenon.
Ph42	The overkill backfire effect—“too many” counterarguments can ironically strengthen misconceptions—is not a robust empirical phenomenon.
Ph43	Individuals who are more motivated to resist disinformation tend to generate significantly more cognitive counterarguments.
Ph44	Corrections that challenge people's worldviews are typically less effective than worldview-consonant corrections.
Ph45	The worldview backfire effect — corrections of worldview-bolstering misinformation can ironically strengthen misconceptions — is not a robust empirical phenomenon.
Ph46	The worldview backfire effects can and do occur.
Ph47	Age and conservatism (in US) increases susceptibility to misinformation.
Ph48	Need for cognition and analytical thinking enable better differentiation between misinformation and accurate news.
Ph49	Encouraging readers to evaluate information as they read it can help reduce the likelihood of encoding inaccurate information.
Ph50	News literacy interventions can help people identify misinformation.

Ph51	News literacy interventions may lower credibility for accurate stories.
Ph52	Short debunkings save resources and may increase the willingness to engage with the debunking.
Ph53	Self-affirmation: the general notion of signalling that the recipient is valued and the correction is not a personal attack is important.
Ph54	It is important not to give exposure to fringe opinion and implicit conspiracy and false claims — “just asking questions” can be as corrosive as explicit messages.
Ph55	Fact-checks may spill over such that content that is not fact-checked may be assumed to be verified (the “implied truth” effect).
Ph57	In some circumstances (e.g., trolling) debunking may be the reason people pay attention to the misinformation.
Ph58	Ability of individuals to make a difference online. Corrections from users, experts, and algorithms can all be effective in reducing community misperceptions when responding to misinformation on social media across platforms.
Rec1	Get it right the first time: we have seen many times over how initial misreporting of certain information can have significant cascading effects that can be abused by malicious actors, even many years later (political context).
Rec2	Recognize that the structure of media systems may reduce the potential beneficial effects of debunking (political context).
Rec3	Tagging something as questionable or from an untrustworthy source is not enough in the face of repeated exposures (political context).
Rec4	We need to do more (for example, using algorithms or platform efforts) to ensure that people are seeing corrections to misinformation when it occurs and that corrections follow best practices (e.g., corrections in headlines, not questions) (political context).
Rec5	Any correction, even if effective, necessarily reinforces a rhetorical frame set by someone else. In that sense, any correction—even if successful—can create a “backfire” effect. One must therefore be mindful of not inadvertently reinforcing a misleading frame.
Rec6	If the misinformation is expected to reach a broad audience then invest in debunking. If the misinformation is expected to cause little attention, then do not waste your resources.
Rec7	Do not refrain from attempting to debunk or correct misinformation out of fear that doing so will backfire or increase beliefs in false information.
Rec8	Don’t worry too much about worldview backfire (despite issues around motivated reasoning and ideology-based selective sharing of fact-checks etc.).
Rec9	Avoid introducing corrections that have not proven to be effective.
Rec10	Make tentative judgments and stay open to change.
Rec11	Advise consumers to only share information you have verified to slow the spread of mis/disinformation (for citizens).

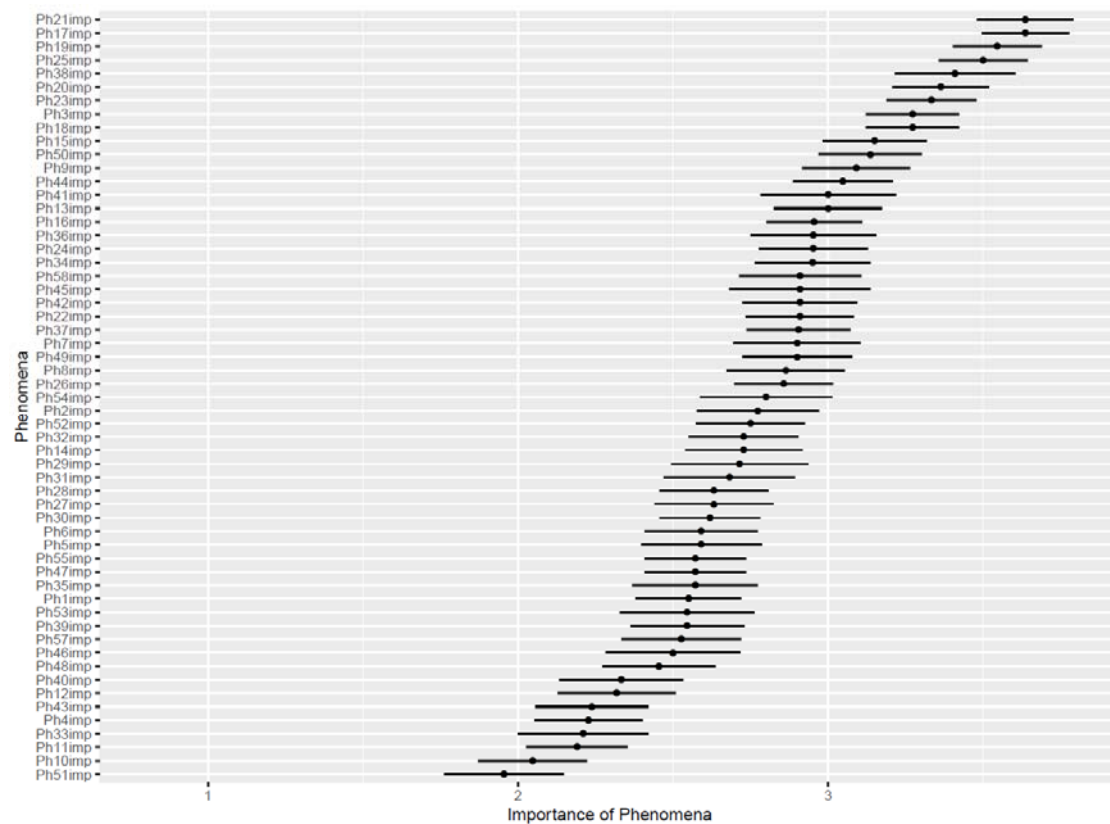
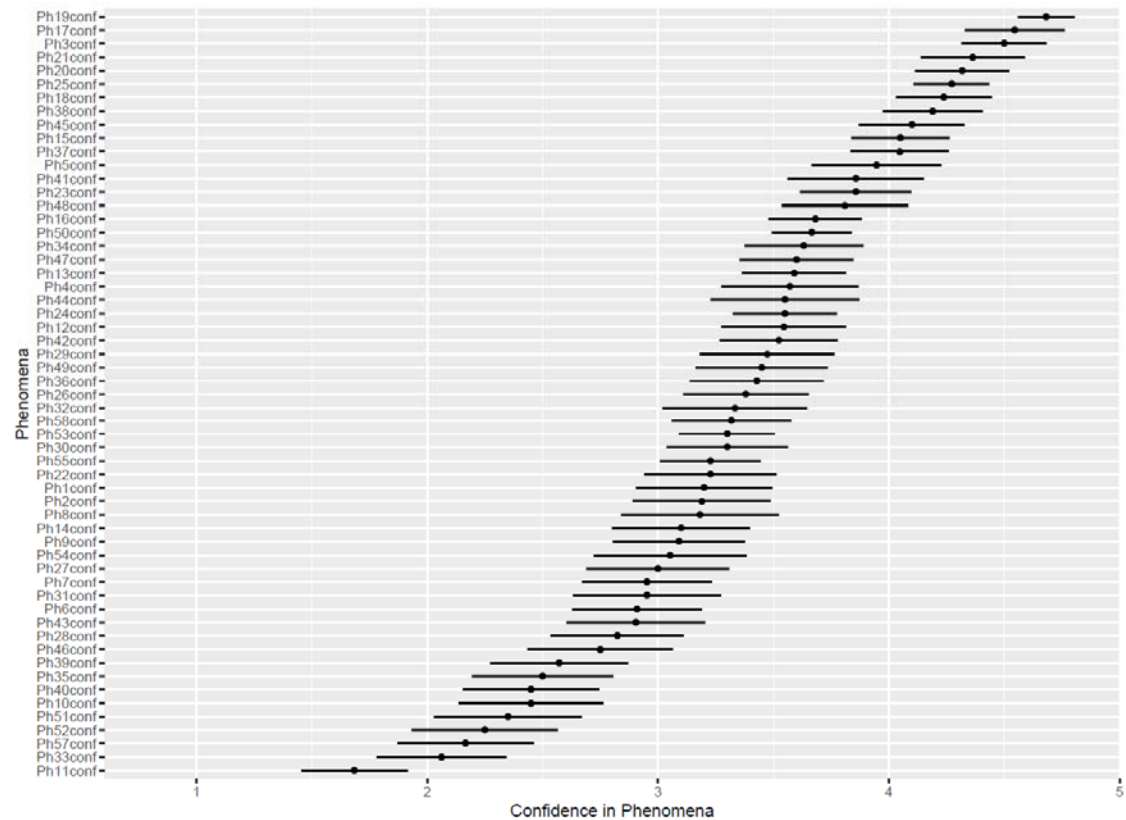
Rec12	Mobilizing users is a promising solution to responding to misinformation (engage citizens).
Rec13	Monitor and introduce agents in public social media networks that are actively disseminating information (engage citizens).
Rec14	Encourage people to slow down: Shifting from system 1 to system 2 thinking and engage analytical thinking (for citizens).
Rec15	Promote critical thinking and defensive confidence in the public (for citizens).
Rec16	Focus on interpersonal effects in online communication—"see something, say something" (for citizens).
Rec17	Advise consumers to follow sourcing best practices and become your own fact checker (for citizens).
Rec18	Make sure your corrective claim is plausible.
Rec19	Be transparent in describing the process and source materials in coming to the correct information.
Rec20	Do not polarize/stigmatize unnecessarily – Do use inclusive language.
Rec21	Avoid scientific jargon/ technical language but do use numbers.
Rec22	Repeat the misinformation, only once, directly prior to the correction. While multiple repetitions of the misinformation prior to the correction should be avoided, one repetition is beneficial to belief updating.
Rec23	Evaluate connections between evidence sources and alternative claims.
Rec24	Simple negations are not effective - it is important to provide people with a clear explanation of (1) why the mistaken information was thought to be correct in the first place and (2) why it is now clear it is wrong and (3) why the alternative is correct.
Rec25	Factual (causal) alternatives: The alternative should not be more complex and have the same explanatory relevance as the original misinformation.
Rec26	Encouraging and supporting people's detection of discrepancies can help reduce the effects of inaccurate presentations and existing inaccurate understandings.
Rec27	Effort should be invested in translating complicated ideas so they are readily accessible to the target audience, thus facilitating acceptance of and memory for a given claim.
Rec28	Detailed corrections promote sustained belief change over time and protect against belief regression.
Rec29	Ensure the misinformation is clearly and saliently paired with the correction. It should be virtually impossible for the individual to ignore, overlook, or not notice the corrective element, even when skimming.
Rec30	Juxtapose the correction with the mistaken information. It is important for people to see the inconsistency in order to resolve it.
Rec31	Finding new ways to engage in de- or pre-bunking, such as through social games (like John Cook's "Cranky Uncle" or the "BadNews" game) or humour is a promising way forward.



Rec32	Because people may superficially engage with online content (skimming headlines, or sharing before reading), when addressing large-scale misinformation, efforts at raising the profile of facts may be more effective than detailed corrections.
Rec33	Long debunkings may not backfire but may be a waste of resources.
Rec34	Connecting corrections and claims towards an individual's identity can have big effects on the likelihood of their taking up and/or endorsing a particular view.
Rec35	Use photos or other semantic aids that facilitate conceptual processing and highlight important elements in factual claims.
Rec36	Exploit benefits of technique rebuttal/logic-based debunking.
Rec37	"Prebunking" or inoculating can offer a potentially more powerful alternative to debunking.
Rec38	Focus on manipulation techniques rather than debunking individual examples of misinformation.
Rec39	Find individuals who people trust to provide the correction - expertise can overwhelm people and does not win them over.

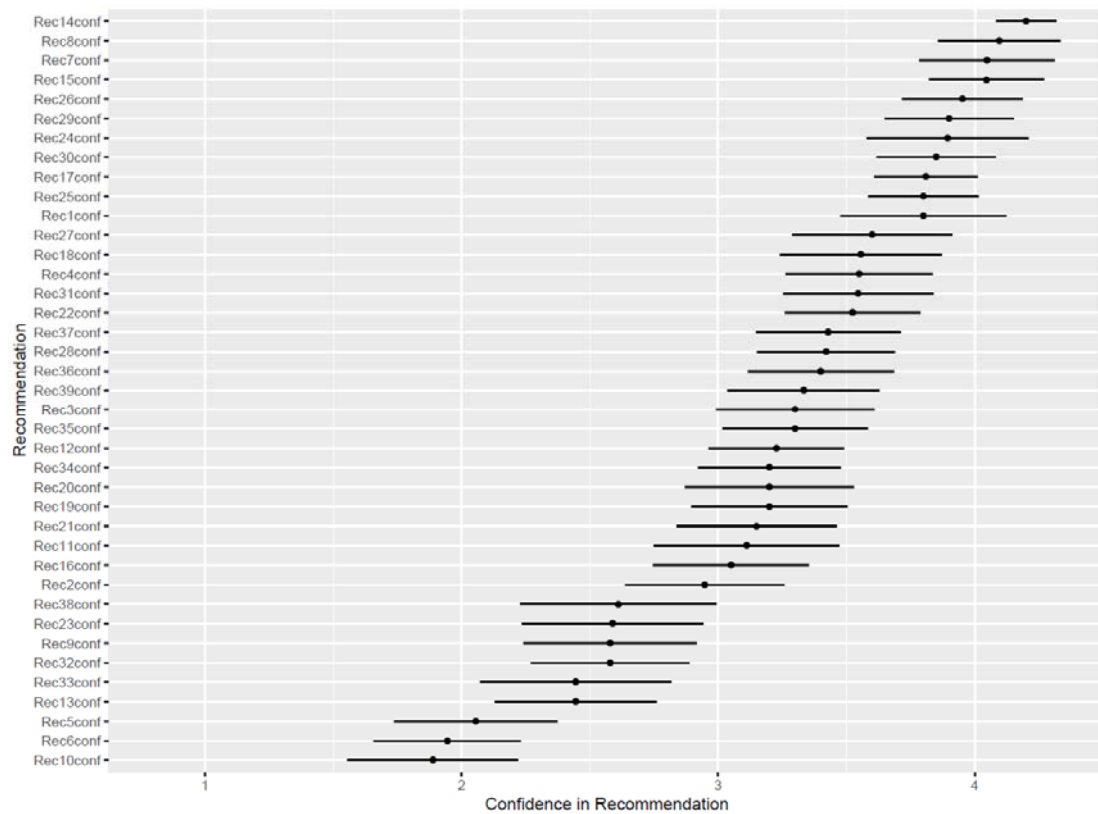
The next figure shows the mean judged confidence (top panel) and importance (bottom panel) for *phenomena*, rank ordered from highest confidence ("conf" suffix) and importance ("imp" suffix), respectively, to lowest. The error bars represent standard errors. Labels on the ordinate use the coding introduced in Table 1.

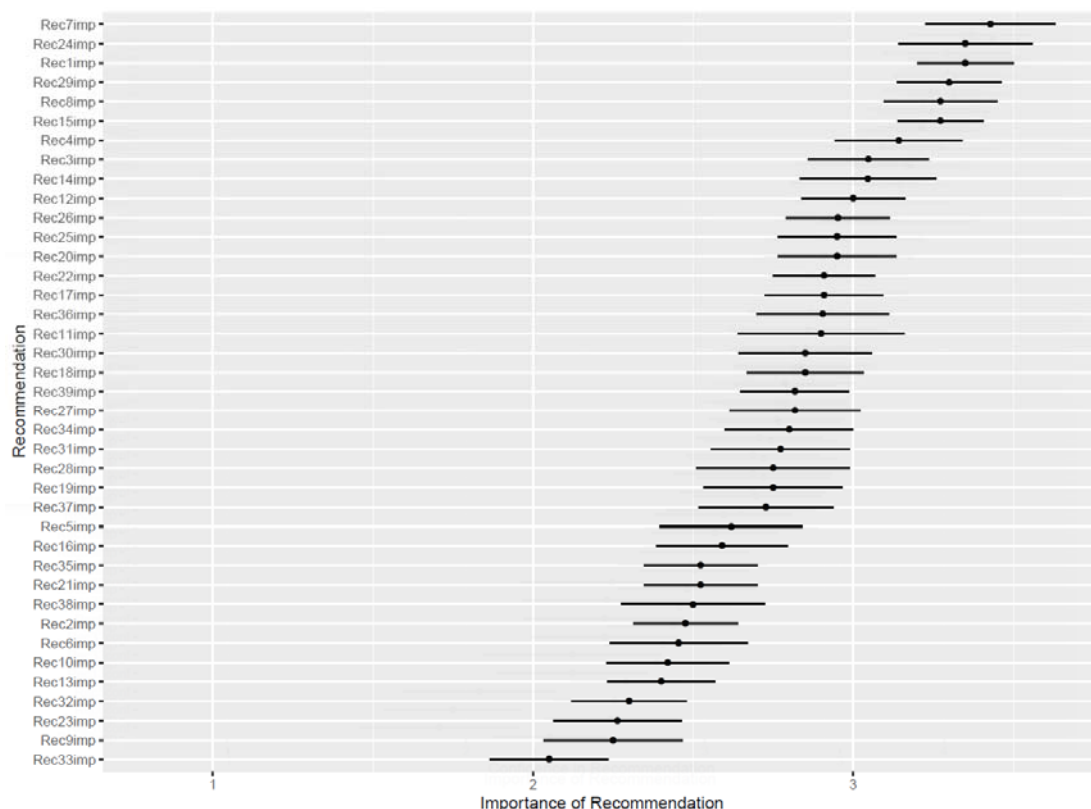
The correlation between importance and confidence ratings for phenomena was  $r(55) = .78$ ,  $p < .0001$ . (Source code for the analysis weaved together with the output can be found at [https://drive.google.com/file/d/1sm\\_yotJm54o3gGck7TDFFjy9a951IYU0/view?usp=sharing](https://drive.google.com/file/d/1sm_yotJm54o3gGck7TDFFjy9a951IYU0/view?usp=sharing)).



The next figure shows the mean judged confidence (top panel) and importance (bottom panel) for *recommendations*, rank ordered from highest confidence (importance) to lowest. The error bars represent standard errors. Labels on the ordinate use the coding introduced in Table 1.

The correlation between importance and confidence ratings for recommendations was  $r(55) = .81, p < .0001$ .





*4.4 Selection of phenomena and recommendations for inclusion.* Several criteria for inclusion of phenomena and recommendations were discussed by the team of authors. The discussion was conducted on Google doc by authors commenting on a proposed list of phenomena and recommendations for inclusion produced by the CLAs. The discussion is archived here: [https://drive.google.com/file/d/1tUZWWAk\\_gweU34yMWw-f7HuyLxJcZcMk/view](https://drive.google.com/file/d/1tUZWWAk_gweU34yMWw-f7HuyLxJcZcMk/view).

The discussion culminated in a Zoom meeting with most authors on 17 August 2020. The meeting minutes can be found at <https://drive.google.com/file/d/1xVMU3gmVBe5QoUyF1PZrOlXuznSO4eR0/view?usp=sharing>. The meeting decided to use a “majority cutoff” for the phenomena and recommendations; that is, all those items were considered for inclusion in *The Debunking Handbook 2020* that received one of the top two ratings for both confidence and importance by a majority of authors.

The meeting also resulted in the creation of two “task forces” that resolved two areas of uncertainty and debate among team members. One of those related to the role of source expertise during debunking, and the other related to the role and existence of backfire effects. The expertise task force was led by Ullrich Ecker and the backfire task force by Stephan Lewandowsky (complete membership listed in minutes of Zoom meeting at the above link). Both task forces resolved the issues by 25 August 2020.

The expertise task force produced a consensus statement available at

[https://drive.google.com/file/d/1-68x9\\_fNc7V-8NiTSs1AzNLmC8k5EZPE/view?usp=sharing](https://drive.google.com/file/d/1-68x9_fNc7V-8NiTSs1AzNLmC8k5EZPE/view?usp=sharing).

The backfire task force collected 5 (anonymous) preferred narratives (using Google Forms) from its members and discussed those in another Zoom meeting, with the narratives available at

[https://drive.google.com/file/d/1TyBExPvPbD-71zGx3Xg\\_8YTK56QFNEy/view?usp=sharing](https://drive.google.com/file/d/1TyBExPvPbD-71zGx3Xg_8YTK56QFNEy/view?usp=sharing).

The final selection of items for inclusion in *The Debunking Handbook 2020* was as follows (ordered by topic area; numbers in brackets are confidence and importance ranks; item numbers are prefixed by “P” and “R”, respectively, to indicate phenomena and recommendations):

### **What is the problem and where does misinformation originate?**

- P1. Repeated exposure to information, including false declarative statements and assertions, increases people’s belief in its truth (“illusory truth effect”) [9/4].
- P2. Misinformation continues to influence people’s thinking even after correction [8/7] (the “continued influence effect”) [2/2].

### **What can be done generally / preemptively?**

- P3. Encouraging readers to evaluate information as they read it can help to reduce the likelihood of encoding inaccurate information [25/28].
- P4. Accuracy nudges and reminders increase the quality of people's sharing decisions on social media [17/20].
- P5. People do not routinely track and evaluate the credibility of sources. However, when they do, impact of misinformation from less-credible sources can be reduced [10/8]. Online source evaluations are challenging (but can be taught) [13/16].
- P6. News literacy interventions can help people to identify misinformation [12/21].
- P7. “Prebunking” or inoculation is an alternative to debunking and the first line of defence against misinformation [4/9].
- P8. Warning people that they might be misinformed reduces later reliance on misinformation [24/11].

### **What can be done to deal with specific pieces of misinformation?**

- P9. Fact-checking and corrections work, at least in part and in many situations, but this does not mean fact-checking can eliminate all inaccurate beliefs [3/1].
- P10. Corrections are more effective if:

- a. in addition to providing a simple retraction (“not true”), they propose a causal alternative, and generally if they provide substantive, relevant information [1/3].
  - b. they are repeated [7/13].
  - c. they involve multiple relevant counterarguments/explanations [18/18].
  - d. people are (made to be) suspicious of the source or intent of the misinformation [21/35].
  - e. they come from trusted sources or highlight expert consensus [5/6]. (There is some disagreement regarding expertise: Expertise matters, especially for factual corrections [28/29] vs. Source expertise does not matter much for corrections [34/40].)
- P11. The overkill backfire effect—“too many” counterarguments can ironically strengthen misconceptions—is *not* a robust empirical phenomenon [16/22].
- P12. The familiarity backfire effect—corrections that repeat misinformation can ironically strengthen misconceptions—is not a robust phenomenon [15/14].
- P13. The efficacy of corrections depends in part on the recipient’s motivation to believe the statement [19/26]. Corrections that challenge people’s worldviews are typically less effective than worldview-consonant corrections [11/19]. The worldview backfire effect—corrections of worldview-bolstering misinformation can ironically strengthen misconceptions—is *not* a robust empirical phenomenon [22/10].
- P14. Emotional content/language influences not only the spread of misinformation, but also corrective efforts [20/23].
- P15. Updates to factual beliefs, even if successful, may not translate into attitude or behaviour change [6/5].

**Additional candidates** (included by majority but missed out on other cutoff):

- P16. Well-designed graphs and videos can be helpful to convey corrections involving complex or statistical information clearly and concisely [33/30].
- P17. Prompts to compare the correction with the initial belief or attitude can increase correction efficacy [38/33].

## Selection of Recommendations

(numbers in brackets are importance ranks)

- R1. Get it right the first time: we have seen many times over how initial misreporting of certain information can have significant cascading effects that can be abused by malicious actors [Importance: #3]
- R2. Encouraging and supporting people's detection of discrepancies can help reduce the effects of inaccurate presentations and existing inaccurate understandings [8].
  - a. Promote critical thinking and defensive confidence in the public [5].
  - b. Encourage people to slow down: Shifting from system 1 to system 2 thinking and engage analytical thinking [10].
  - c. Advise consumers to follow sourcing best practices and become your own fact checker [16].
- R3. We need to do more (for example, using algorithms or platform efforts) to ensure that people are seeing corrections to misinformation when it occurs and that corrections follow best practices (e.g., corrections in headlines, not questions) [7].
- R4. Do not refrain from attempting to debunk or correct misinformation out of fear that doing so will backfire or increase beliefs in false information [1]. Don't worry too much about worldview backfire (despite issues around motivated reasoning and ideology-based selective sharing of fact-checks etc.) [6].
- R5. Simple negations are not effective—it is important to provide people with a clear explanation of (1) why the mistaken information was thought to be correct in the first place, (2) why it is now clear it is wrong, and (3) why the alternative is correct [2].
- R6. Make sure your corrective claim is plausible [21]. When using factual (causal) alternatives, the alternative should not be more complex and have the same explanatory relevance as the original misinformation [14].
- R7. Ensure the misinformation is clearly and saliently paired with the correction. It should be virtually impossible for the individual to ignore, overlook, or not notice the corrective... [4]. Juxtapose the correction with the mistaken information; it is important for people to see the inconsistency in order to resolve it [15]. Repeat the misinformation, only once, directly prior to the correction. While multiple repetitions of the misinformation prior to the correction should be avoided, one repetition is beneficial to belief updating [17].
- R8. Do not polarize/stigmatize unnecessarily—do use inclusive language [13].

R9. Effort should be invested in translating complicated ideas so they are readily accessible to the target audience, thus facilitating acceptance of and memory for a given claim [18].

R10. Exploit benefits of technique rebuttal/logic-based debunking [19]. "Prebunking" or inoculating can offer a potentially more powerful alternative to debunking [26]. Find new ways to engage in de- or pre-bunking, such as through social games ("Cranky Uncle" and "Bad News" games) and humour [20].

*4.5 Production of first draft.* The CLAs produced a first draft of the document based on the above list of phenomena and recommendations chosen for inclusion, and also based on the outcome of the two task forces. The draft was discussed by the team of authors, again using Google doc. The discussion is archived here:

<https://drive.google.com/file/d/1MSrrDDpSXLfhB6Kwhit9gb-vOpn291bM/view?usp=sharing/>.

*4.6 Review.* Once the draft had been refined through team discussion, it was sent for review to four additional experts. This step was not preregistered and was mainly intended to guard against errors and problematic claims, rather than to further improve the draft through detailed editorial comments.

*4.7 Final product.* The typeset document was error-checked twice by all team members. An annotated version is enclosed below that cross-references the content to the phenomena and recommendations nominated for inclusion (see above).



## Quick guide to responding to misinformation



## Misinformation can do damage

Misinformation is false information that is spread either by mistake or with intent to mislead. When there is intent to mislead, it is called disinformation. Misinformation has the potential to cause substantial harm to individuals and society. It is therefore important to protect people against being misinformed, either by making them resilient against misinformation before it is encountered or by debunking it after people have been exposed.



## Misinformation can be sticky!

Fact-checking can reduce people's beliefs in false information. However, misinformation often continues to influence people's thinking even after they receive and accept a correction—this is known as the “continued influence effect”. Even if a factual correction seems effective—because people acknowledge it and it is clear that they have updated their beliefs—people frequently rely on the misinformation in other contexts, for example when answering questions only indirectly related to the misinformation. It is therefore important to use the most effective debunking approaches to achieve maximal impact.



## Prevent misinformation from sticking if you can

Because misinformation is sticky, it's best pre-empted. This can be achieved by explaining misleading or manipulative argumentation strategies to people—a technique known as “inoculation” that makes people resilient to subsequent manipulation attempts. A potential drawback of inoculation is that it requires advance knowledge of misinformation techniques and is best administered before people are exposed to the misinformation.



## Debunk often and properly

If you cannot preempt, you must debunk. For debunking to be effective, it is important to provide detailed refutations<sup>1,2</sup>. Provide a clear explanation of (1) why it is now clear that the information is false, and (2) what is true instead. When those detailed refutations are provided, misinformation can be “unstuck”. Without detailed refutations, the misinformation may continue to stick around despite correction attempts.

## Misinformation can do damage

Misinformation damages society in a number of ways<sup>3,4</sup>. If parents withhold vaccinations from their children based on mistaken beliefs, public health suffers<sup>5</sup>. If people fall for conspiracy theories surrounding COVID-19, they are less likely to comply with government guidelines to manage the pandemic<sup>6</sup>, thereby imperiling all of us.

It's easy to be misled. Our feelings of familiarity with truth are often linked. We are more likely to believe things that we have heard many times than new information.

*“Objective truth is less important than familiarity: we tend to believe falsehoods when they are repeated sufficiently often.”*

**P1** This phenomenon is called the “illusory truth effect”<sup>7,8</sup>. Thus, the more people encounter a piece of information they do not challenge, the more the misinformation seems true, and the more it sticks. Even if a source is identified as unreliable or is blatantly false and inconsistent with people's ideology, repeated exposure to information still tips people towards believing its claims<sup>9,10,11</sup>.

**P14** Misinformation is also often steeped in emotional language and designed to be attention-grabbing and have a persuasive appeal. This facilitates its spread and can boost its impact<sup>12</sup>, especially in the current online economy in which user attention has become a commodity<sup>13</sup>.

Misinformation can also be intentionally suggested by “just asking questions”, a technique that allows provocateurs to hint at falsehoods or conspiracies while maintaining a facade of respectability<sup>14</sup>. For example, in one study, merely presenting questions that hinted at a conspiracy relating to the Zika virus induced significant belief in the conspiracy<sup>15</sup>. Likewise, if you do not read past a headline such as “Are aliens amongst us?” you might walk away with the wrong idea.

## Where does misinformation come from?

Misinformation ranges from outdated news initially thought to be true and disseminated in good faith, to technically true but misleading half-truths, to entirely fabricated disinformation spread intentionally to mislead or confuse the public. People can even acquire misconceptions from obviously fictional materials<sup>16,17</sup>. Hyper-partisan news sources frequently produce misinformation<sup>18</sup>, which is then circulated by partisan networks. Misinformation has been shown to set the political agenda<sup>19</sup>.

## Misinformation can be sticky!

*“Misinformation is sticky—even when it seems to have been corrected.”*

**P2** A fundamental conundrum with misinformation is that even though corrections may seem to reduce people's beliefs in false information, the misinformation often continues to influence people's thinking—this is known as the “continued influence effect”. The effect has been replicated many times. For example, someone might hear a relative has fallen ill from food poisoning. Even if they later learn that the information was incorrect—and even if the person accepts and remembers this correction—they might still show a lingering reliance on the initial misinformation in different contexts (e.g., they might avoid the restaurant allegedly involved).

**P9** Fact-checking and corrections appear to “work” when you ask people directly about their beliefs. For example, people may report the correction accurately and state that they no longer believe the original misinformation. But it doesn't guarantee that the misinformation will not pop up elsewhere, for example when answering questions making indirectly related decisions.

**P15** Even though misinformation is sticky, we have opportunities to respond. We can prevent misinformation from taking root in the first place. Or we can apply best practices to debunk misinformation successfully.

*“Once experienced, even corrected misinformation can linger in memory but we can often undo its influence if we follow best practices.”*

## Sticky myths leave other marks

There is much evidence that updates to factual beliefs, even if successful, may not translate into attitude or behaviour change. For example, in polarized societies (e.g., the US) people indicate that they will continue to vote for their favored politician even if they discover that the majority of the politician's statements are false<sup>20,21</sup>. Fortunately, it does not have to be that way. In less polarized societies (e.g., Australia), people's voting intentions are sensitive to politicians' truthfulness<sup>22</sup>.

Nevertheless, do not refrain from debunking because you are worried it will not change behaviour. Successful debunking can affect behaviour—for example, it can reduce people's willingness to spend money on questionable health products or their sharing of misleading content online<sup>23,24</sup>.

## Prevent misinformation from sticking if you can

**R1** Misinformation is hard to dislodge, preventing it from taking root in the first place is one fruitful strategy. Several prevention strategies are known to be effective.

**P8** Preemptively warning people that they might be misinformed can reduce later reliance on misinformation<sup>25,26</sup>. Even general warnings (“the media sometimes does not check facts before publishing information that turns out to be inaccurate”) can make people more receptive to later corrections. Specific warnings that content may be false have been shown to reduce the likelihood that people will share the information online<sup>27</sup>.

**P7** The process of inoculation or “prebunking” includes a forewarning as well as a preemptive refutation and draws the biomedical analogy<sup>28</sup>. By exposing people to a severely weakened dose of the techniques used in misinformation (and by preemptively refuting them), “cognitive antibodies” can be cultivated. For example, by explaining to people how the tobacco industry rolled out “fake experts” in the 1960s to create a chimerical scientific “debate” about the harms from smoking, people become more resistant to subsequent persuasion attempts using the same misleading argumentation in the context of climate change<sup>29</sup>.

**R10** Effectiveness of inoculation has been shown repeatedly and across many different topics<sup>30,31,32,33,34</sup>. Recently, it has been shown that inoculation can be scaled up through engaging multimedia applications, such as cartoons<sup>35</sup> and games<sup>36,37</sup>.

## Simple steps to greater media literacy

**P3** Simply encouraging people to critically evaluate information as they read it can reduce the likelihood of taking in inaccurate information<sup>38</sup> or help people become more discerning in their sharing behavior<sup>39</sup>.

**P6** Educating readers about specific strategies to aid in this critical evaluation can help people develop important habits. Such strategies include: Taking a “buyer beware” stance towards all information on social media; slowing down and thinking about the information provided, evaluating its plausibility in light of alternatives<sup>40,41</sup>; always considering information sources, including their track record, their expertise, and their motives<sup>42</sup>; and verifying claims (e.g., through “lateral reading”<sup>43</sup>) before sharing them<sup>44</sup>. Lateral reading means to check other sources to evaluate the credibility of a website rather than trying to analyse the site itself. Many tools and suggestions for enhancing digital literacy exist<sup>45</sup>.

**P5** You cannot assume that people spontaneously engage in such behaviours<sup>46</sup>. People do not routinely track, evaluate, or use the credibility of sources in their judgments<sup>47</sup>. However, when they do, the impact of misinformation from less-credible sources can be reduced (see next textbox).

### The strategic landscape of debunking

If you are unable to prevent misinformation from sticking, then you have another arrow in your quiver: Debunking! However, you should first think about a few things before you start debunking.

Everyone has limited time and resources, so you need to pick your battles. If a myth is not spreading widely, or does not have the potential to cause harm now or in the future, there may be no point in debunking it. Your efforts may be better invested elsewhere, and the less said about an unknown myth the better.

Corrections have to point to the misinformation so they necessarily raise its familiarity. However, hearing about misinformation in a correction does little damage, even if the correction introduces a myth that people have never heard of before<sup>10</sup>. Nonetheless, one should be mindful not to give undue exposure to fringe opinion and conspiracy claims through a correction. If no one has heard of the myth that caravans can dissolve concrete, why correct it in public?

Debunkers should also be mindful that any correction necessarily reinforces a rhetorical frame (i.e., a set of "talking points") created by someone else. You cannot correct someone else's myth without talking about it. In that sense, any correction—even if successful—can have unintended consequences, and choosing one's own frame may be more beneficial. For example, highlighting the enormous success and safety of a vaccine might create a more positive set of talking points than debunking a vaccine-related myth<sup>11</sup>. And they are your talking points, not someone else's.



The Debunking Handbook 2020

8

### Who should debunk?

Successful communication rests on the communicator's credibility.

Information from sources that are perceived to be credible typically creates stronger beliefs<sup>12</sup> and is more persuasive<sup>13,14</sup>. By and large, this also holds for misinformation<sup>15,16</sup>. However, credibility may have limited effects when people pay little attention to the source<sup>16,17</sup>, or when the sources are news outlets rather than people<sup>18,19</sup>.

**P10** Source credibility also matters for corrections of misinformation, although perhaps to a lesser extent<sup>20,21</sup>. When breaking down credibility into trustworthiness and expertise, perceived trustworthiness of a debunking source may matter more than its perceived expertise<sup>16,17</sup>. Sources with high credibility on both dimensions (e.g., health professionals or trusted health organizations) may be ideal choices<sup>22-24</sup>.

It is worth keeping in mind that the credibility of a source will matter more to some groups than others, depending on content and context<sup>25-27</sup>. For example, people with negative attitudes toward vaccines distrust verbal sources of vaccine-related information (including generally-trusted health organizations)<sup>24</sup>.

**P10** Deliver the message to the audience and use a messenger trusted by the target group<sup>28</sup>. Discredit misinformation sources that have vested interests<sup>29</sup>.

### The elusive backfire effects

**P12** years ago, scholars and practitioners were concerned that corrections may "backfire", that is, ironically strengthen misconceptions rather than reduce them. Recent research has allayed these concerns: backfire effects occur only occasionally and the risk of occurrence is lower in most situations than once thought.

**R4** not refrain from attempting to debunk or correct misinformation out of fear that doing so will backfire or increase beliefs in false information<sup>30,31,32</sup>.

**Definition**  
**Backfire Effect:** A backfire effect is where a correction backfires: it increases belief in, or reliance on, misinformation relative to a pre-correction or no-correction baseline.

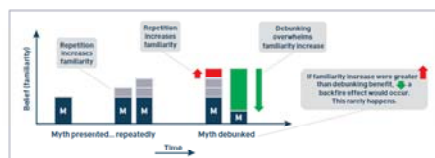
**P12** Backfire effects are not as common as we used to think. We cannot reliably predict the circumstances under which they occur.<sup>33</sup>

The Debunking Handbook 2020

9

### Familiarity backfire effect

**P11** Repetition makes information more familiar, and familiar information is generally perceived to be more truthful than novel information (the aforementioned illusory-truth effect). Because a myth is necessarily repeated when debunked, the risk arises that debunking may backfire by making a myth more familiar (see figure below). Early evidence was supportive of this idea, but more recently, exhaustive experimental attempts to induce a backfire effect through familiarity alone have come up empty<sup>34-36</sup>. Thus, while repeating misinformation generally increases familiarity and truth ratings, repeating a myth while refuting it has been found to be safe in many circumstances, and can even make the correction more salient and effective<sup>37</sup>.



"Debunking a myth makes it more familiar but the debunking usually overpowers the increase in familiarity."

### P11 Arkill backfire effect

This effect refers to the idea that providing "too many" counterarguments against a false claim might produce unintended effects or even backfire. The only study to directly examine this notion, however, found no evidence of this effect and instead concluded that a greater number of relevant counterarguments generally leads to greater reduction of misconceptions<sup>38</sup>.

### Worldview backfire effect

**P13** The worldview backfire effect is presumed to occur when a correction that challenges people's worldview causes belief in the misinformation. While there was initially some evidence for the worldview backfire effect<sup>39</sup>, most research indicates that it is not a pervasive and robust empirical phenomenon.

### Personal experience vs. evidence

**R4** Although communicators may observe backfire effects in their everyday lives, many experiments have shown that, in fact, such behavior is unusual. Social scientists are still figuring out why some people "backfire" but not others, and why those effects occur on some occasions but not others. However, the accumulated evidence to date is clear that the worldview backfire effect is not a sufficient reason to avoid debunking and fact-checking.

**P12** Several studies have failed to obtain a backfire effect even in theoretically favourable circumstances<sup>40,41,42,43,44</sup>, as, while there are reports of worldview backfire effects emerging under specific conditions (e.g., when publicans are presented with information concerning climate mitigation measures<sup>45</sup>) concern about worldview backfire has been disproportionate.

### Role of worldview in belief confirmation

Even if worldview backfire effects are infrequent, there are other ways that worldview can affect debunking.

Worldview can affect what content people choose to consume<sup>46,47,48</sup>. This process of selective exposure may mean that people are more likely to be exposed to worldview-consistent false or misleading claims in the first place, and by implication, less likely to be exposed to corrective information about such claims after exposure. To illustrate, one analysis showed that 62% of visits to fake news websites came from the 20% of Americans with the most conservative information diet<sup>49</sup>.

**P13** The efficacy of corrections depends in part on the recipient's willingness to believe the statement. Activating group identities likely induces constraints in how people think about an issue—depending on the identity and the issue, this may ameliorate or exacerbate misperceptions, and it may affect whom a person will believe. This highlights the importance of using inclusive language and avoiding the stigmatization of groups for holding inaccurate beliefs. Doing so is likely to polarize more than generate desired updating.

**R8** Recent research suggests that although (mis-)information diets may differ across the political spectrum, some of the motivated reasoning processes just described may be symmetric for liberals and conservatives<sup>50</sup>.

**R4** On balance, recent evidence provides no reason to avoid debunking fear of a backfire effect. Debunking is likely to be at least partially effective, except for some limited circumstances when people's worldviews are being challenged.<sup>51</sup>

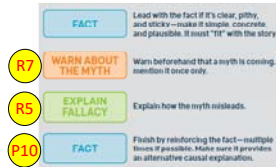
The Debunking Handbook 2020

11

### Debunk often and do it properly

**R5** Corrections on their own are unlikely to fully unstick misinformation. Tagging something as questionable from an untrustworthy source is not enough in the face of repeated exposures.

Debunking is more likely to be successful if you apply the following 3 or 4 components:



#### FACT: State the truth first

If it's easy to do in a few clear words, state what is true first. This allows you to frame the message—you lead with your talking points, not someone else's.

- R3** The best corrections are as prominent (in the headlines, not buried in questions) as the misinformation.
- P5** Do not rely on a simple retraction ("this claim is not true").
- P10** Providing a factual alternative, that is an alternative that fills a causal "gap" in explaining what happened if misinformation is corrected, is an effective method of debunking. Having a causal alternative facilitates "switching out" the inaccurate information in an individual's initial understanding and replaces it with a new version of what happened.
- R6** The alternative should not be more complex and should have the same explanatory relevance as the original misinformation.<sup>1, 36, 37</sup>
- P5** There may, however, be circumstances in which the facts are so nuanced that they escape pithy summary. In these cases, it may be better to lead with an explanation of why the myth is false before explaining the facts.

### Debunk often and do it properly

**R5** Corrections on their own are unlikely to fully unstick misinformation. Tagging something as questionable from an untrustworthy source is not enough in the face of repeated exposures.

Debunking is more likely to be successful if you apply the following 3 or 4 components:



#### FACT: State the truth first

If it's easy to do in a few clear words, state what is true first. This allows you to frame the message—you lead with your talking points, not someone else's.

- R3** The best corrections are as prominent (in the headlines, not buried in questions) as the misinformation.
- P5** Do not rely on a simple retraction ("this claim is not true").
- P10** Providing a factual alternative, that is an alternative that fills a causal "gap" in explaining what happened if misinformation is corrected, is an effective method of debunking. Having a causal alternative facilitates "switching out" the inaccurate information in an individual's initial understanding and replaces it with a new version of what happened.
- R6** The alternative should not be more complex and should have the same explanatory relevance as the original misinformation.<sup>1, 36, 37</sup>
- P5** There may, however, be circumstances in which the facts are so nuanced that they escape pithy summary. In these cases, it may be better to lead with an explanation of why the myth is false before explaining the facts.

### General guidelines:

Avoid scientific jargon or complex, technical language<sup>38</sup>.

**P16** Well-designed graphs, videos, photos, and other semantic aids can be helpful to convey corrections involving complex or statistical information clearly and concisely<sup>35, 36, 37</sup>.

**R9** The truth is often more complicated than some viral false claim. You must invest effort in translating complicated issues so they are readily accessible to the target audience...so they can be easily read, easily imagined, and easily acted<sup>38, 39, 40</sup>.

### Collective action: Debunking on social media

**P4** Accuracy nudges (e.g., "most people want to receive accurate information") and reminders increase the quality of what's shared on social media<sup>41</sup>.

Mobilize social media users to respond quickly to misinformation by sharing facts. A platform's efforts may not be sufficient in scope or scalable to misinformation; user correction can work if people feel emboldened to engage in it<sup>40, 42</sup>.

*"Focus on interpersonal effects in online communication: 'see something, say something'."*<sup>43</sup>

**P10** Individuals have the ability to make a difference online. Corrections from users, experts, and algorithms (e.g., recommending related articles that contain a correction) can all be effective in reducing community misperceptions when responding to misinformation<sup>44, 45, 46</sup>.

Seeing someone else on social media being corrected (known as observational corrections) can lead to more accurate attitudes on various topics<sup>45</sup>.

Conversely, not speaking out can lead to a "spiral of silence," both for the person being corrected and for the observer, where a mute majority codes a narrative to a vocal but misinformed minority<sup>46, 47, 48</sup>.

### Example of a Refutation



### References

- Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: A new survey of the experts. *American Psychologist*, 56, 405-416.
- Rosenfeld, R. M., Nnacheta, L. C., & Corrigan, M. D. (2015). Clinical consensus statement development manual. *Otolaryngology—Head and Neck Surgery*, 153, S1-S14.