1 Introduction

1.1 Rationale

This meta-analysis aims to quantify the comparative effectiveness and contextual parameters of different intervention strategies for reducing 1) susceptibility to misinformation and epistemically unwarranted beliefs, and 2) a willingness to share misinformation online. In recent years, a number of intervention strategies to tackle the spread of misinformation have received broad evidential support in different domains. For example, a number of games have been developed based on inoculation theory, which posits that encountering weakened versions of arguments as well as their relevant refutations provides psychological immunity against persuasive attacks on attitudes (see Banas & Rains, 2010; Compton, 2013; McGuire, 1964). These gamified intervention strategies have demonstrated that incentivised exposure to information detailing common approaches used by 'merchants of doubt' (e.g., Basol et al., 2020; van der Linden et al., 2017; see also Oreskes & Conway, 2010), radical extremist organisations (e.g., Saleh et al., 2021), and other bad faith actors pushing online disinformation (e.g., Basol et al., 2020; Maertens et al., 2021; Roozenbeek & van der Linden, 2019a, 2019b; Roozenbeek et al., 2020) subsequently reduces susceptibility to these attempts to persuade with the use of misinformation. Other approaches have focused on the use of 'accuracy nudges', aiming to improve truth discernment between real and fake news headlines and reduce willingness to share fake headlines through priming accuracy with cognitive reflection 'nudges' (e.g., Fazio, 2020; Pennycook et al., 2020, 2021). Other approaches include news media literacy training (e.g., Hameleers, 2020) and fact-checking (see Walter et al., 2019) among others.

While these various intervention strategies have largely been supported with individual research attempts, the existence of meta-analytic synthesis comparing the effectiveness of these different strategies is lacking (but see Chan et al., 2017 for a meta-analysis on debunking alone, and Walter et al., 2019 for a meta-analysis on fact-checking alone). This is important for a number of reasons. First, comparing the relative strength of each intervention strategy would provide useful insights for theoretical extensions in the literature alongside informing policymakers and researchers deciding which strategies to employ. Second, coding of moderating variables and analysis of their influence on the effectiveness of each intervention would serve to inform theoretical advancements on the nature of these approaches, alongside the planned implementation of certain interventions over others by policymakers based on certain contextual factors. Finally, meta-analytic synthesis can provide useful insights for the literature by analysing statistical biases that cannot be conducted through smaller research projects alone. These include the ability to determine the likelihood that findings are influenced by the *file drawer problem* (Rosenthal, 1979), varying study and measurement quality (see Peters et al., 2008), and robustly informed overviews for the likelihood of the directional hypothesis over the null with the use of Bayesian meta-analytic techniques. Therefore, we believe that the implementation of a meta-analytic approach comparing the effectiveness of different existing intervention strategies for reducing susceptibility and willingness to share misinformation online is timely and useful to effectively advance the literature and provide helpful information for policymakers hoping to tackle the *infodemic.*

1.2 Objectives

In this meta-analysis, we examine the relevant literature to answer the following research questions:

Q_0 *Are any of the existing intervention strategies significantly more or less effective at reducing susceptibility to misinformation and epistemically unwarranted beliefs than all other intervention strategies combined?* To test this research question, we will conduct a meta-analysis including the effects of all intervention strategies on reducing susceptibility to misinformation. Then, we will analyse whether the type of intervention moderates the strength of this effect.

Q_1 *Are any of the existing intervention strategies significantly more or less effective at reducing willingness to share misinformation than all other intervention strategies combined?* To test this research question, we will conduct a meta-analysis including the effects of all intervention strategies on reducing willingness to share misinformation. Then, we will analyse whether the type of intervention moderates the strength of this effect.

Q_2 *What are the contextual parameters (if any) to the findings from Q_{0-1}?* To test this research question, we will conduct moderation analyses on the two models from Q_{0-1} to determine whether the strength of the overall intervention effects vary based on factors associated with the type of misinformation susceptibility measured (e.g., truth discernment, awareness of manipulation attempts), psychometric qualities of the measures used (e.g., binary vs. continuous measures of truth discernment), context of the measures used (e.g., inoculation against conspiracy vs. polarising content), sample characteristics (e.g., Republican vs. Democrat partisan affiliation), and others.

Q_3 *How effective are each of the existing intervention strategies for reducing susceptibility to misinformation and epistemically unwarranted beliefs?* To test this research question, we will conduct separate meta-analyses to obtain the effect of each intervention strategy on reducing susceptibility to misinformation and epistemically unwarranted beliefs.

Q_4 *How effective are each of the existing intervention strategies for reducing willingness to share misinformation?* To test this research question, we will conduct separate meta-analyses to obtain the effect of each intervention strategy on reducing willingness to share misinformation.

Q_5 *What are the contextual parameters (if any) to the findings from $Q_{3\text{-}4}$?* To test this research question, we will conduct moderation analyses on all models from *$Q_{3\text{-}4}$* to determine whether the strength of the overall intervention effects vary based on factors associated with the subtype of the intervention used (e.g., active vs. passive inoculation), type of misinformation susceptibility measured, psychometric qualities of the measures used, context of the measures used, sample characteristics, and others.

For $Q_{0\text{-}5}$ we aim to explore the following questions:

QA *What are the effect sizes of the different intervention strategies for reducing susceptibility to misinformation, epistemically unwarranted beliefs, and willingness to share misinformation?*

QB *How homogeneous and reliable are the effect sizes of the different intervention strategies for reducing susceptibility to misinformation, epistemically unwarranted beliefs, and willingness to share misinformation?*

QC *What are the moderating effects (if any) on the effectiveness of the different intervention strategies for reducing susceptibility to misinformation, epistemically unwarranted beliefs, and willingness to share misinformation?*

2 Methods

2.1 Eligibility criteria

In this meta-analysis, we are considering studies that include *i)* any measure of susceptibility to misinformation or epistemically unwarranted beliefs, *ii)* any measure of willingness to share misinformation, see *Types of outcome variables,* and *iii)* any manipulated intervention strategies to tackle these outcomes, see *Types of independent variables.* Below, we discuss these criteria in detail.

2.1.1 Types of studies

We will include experimental studies of all designs, as long as they fill the criteria outlined above. We will consider study design features as moderators of meta-analytic effect sizes.

2.1.2 Types of participants

We will include any participants from studies that fill the criteria outlined above. We will consider mean age, sample nationality, political ideology, partisan affiliation, and whether the sample consists exclusively of students as moderators.

2.1.3 Types of independent variables

We will include studies that manipulate **at least one**[1] of the independent variables that are associated with the following intervention strategies:

Fact-checking. We will consider any studies that manipulate the presentation of fact-checks to participants. These will include, but are not limited to, fact-checking from both crowdsourcing and experts, as well as producing fact-checks both before (pre-bunking) and after (de-bunking) the presentation of misinformation.

Source cues. We will consider any studies that manipulate the labelling of sources for the misinformation presented. These will include, but are not limited to, cues of non-credible versus credible sources, or hyperpartisan versus relatively impartial sources.

Accuracy nudges or prompts. We will consider any studies that manipulate reflective reasoning. These will include, but are not limited to, requesting accuracy ratings from participants, introducing "friction", as well as priming analytical or reflective versus intuitive or automatic thinking styles.

Inoculation. We will consider any studies that manipulate pre-emptive exposure to small doses of misinformation or disinformation strategies. These will include, but are not limited to, both passive (i.e., presenting prepared information) and active (i.e., asking participants to play a part in generating the relevant information) inoculation strategies, as well as the use of gamified procedures that inform participants of common disinformation strategies.

Personal incentives. We will consider any studies that manipulate personal incentives. These will include, but are not limited to, the indication of social or financial benefit as a result of accurate discernment.

[1] If we discover additional interventions we will also collect data for their effects.

Psychological skills training. We will consider any studies that manipulate cognitive and psychological skills that aid in the careful and informed appraisal of evidence. These will include, but are not limited to, critical thinking and media literacy training.

Protective motivations. We will consider any studies that manipulate the satisfaction of psychological needs to protect against motivated reasoning through biased justifications or psychological rigidity. These will include, but are not limited to, self-affirmation primes.

2.1.4 Types of outcome variables

We include studies that measure **at least one**[2] of the following outcome variables:

Misinformation susceptibility. We will consider studies that measure constructs such as sum, average, or difference ratings of veracity discernment, real news detection, fake news detection, perceived reliability of misinformation, perceived manipulativeness of misinformation, and perceived accuracy of misinformation.

Epistemically unwarranted beliefs. We will consider studies that measure constructs such as conspiracy beliefs, pseudoscientific beliefs, and bullshit receptivity.

Sharing of fake news. We will consider studies that measure any behavioural indications or intentions to share misinformation, fake news, or real news online.

2.1.5 Other criteria

We will not use any other eligibility criteria for the meta-analysis.

2.2 Information sources

We will search the *PsycINFO, Web of Science, ProQuest Conference Papers & Proceedings and Dissertations & Theses,* and *Europe PMC Pre-Prints* databases. This process will be carried out until 1st June 2022. We aim to call for unpublished data around 1st March 2022 by sending a message to the mailing lists of all relevant societies, they will be given 31 days to respond. We will also email researchers that are prolific in the relevant research at this time to call for any other unpublished data. Any uncertainties or missing data will be noted, and

[2] If we discover additional outcome measures we will also collect their data.

the authors will be emailed to request this information. During the data extraction process we will follow relevant citations to find any potential papers that we may have missed previously.

2.3 Search strategy

When searching the databases, we will use the following search terms:

(Misinfo* OR disinfo* OR conspir* OR fake AND news OR pseudosci* OR epistem* AND unwarrant* AND belief* OR fals* AND news OR bullshit AND receptiv*) AND interven*

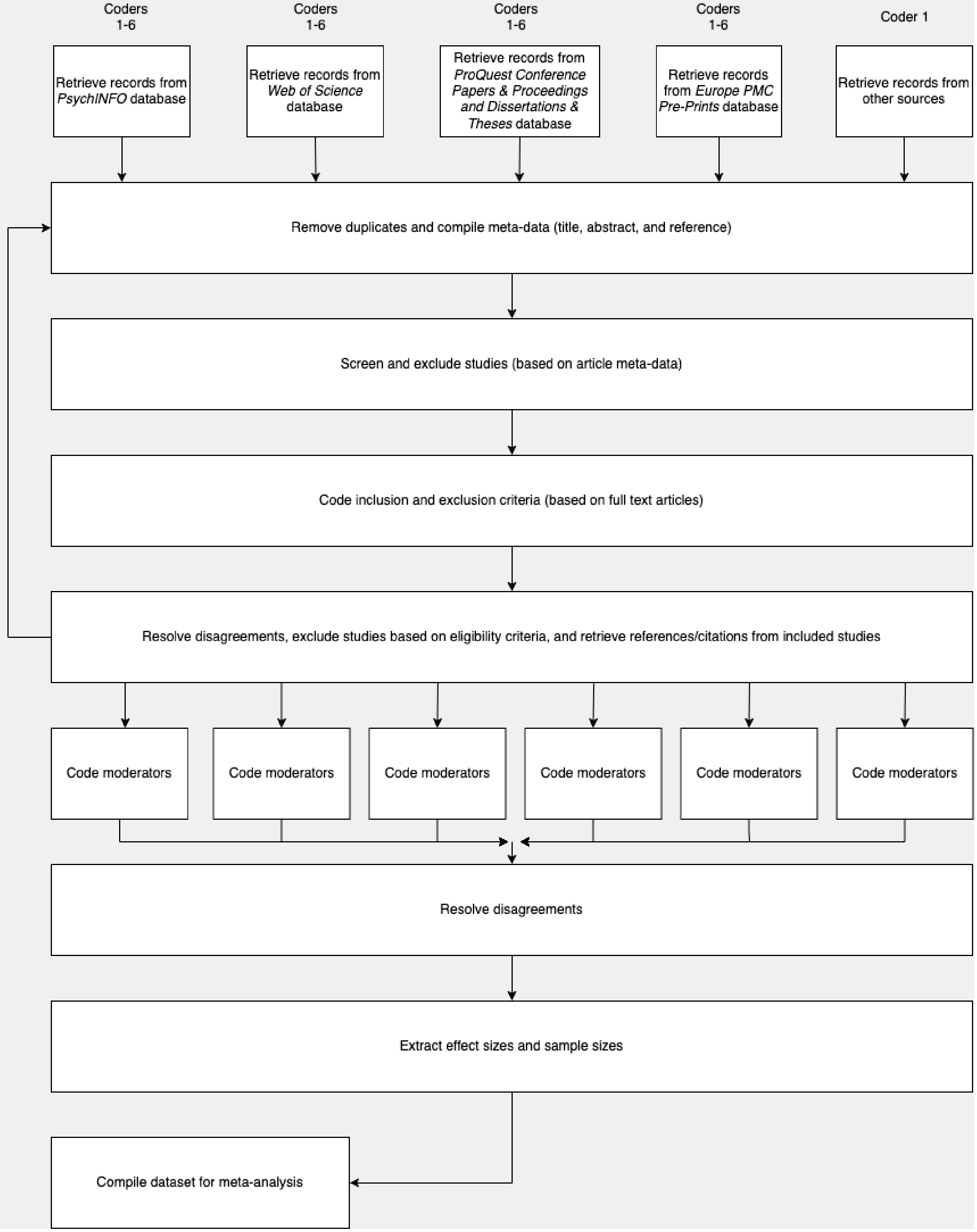
2.4 Data Management

Once the relevant papers are selected, each paper will be given a specific identification number and added to shared *Zotero* folders. Once duplicates have been removed, we will create a spreadsheet to compile the meta-data. Once the data has been extracted, new spreadsheets will be created to be saved as CSV files for the respective meta-analyses using R.

2.5 Study selection

We will select relevant studies in three steps. First, we will screen records based on their title, abstract, and keywords. We will then divide records between six coders who each screen a sixth of the records. For each record, coders will decide whether the record meets the eligibility criteria (yes, maybe, no). We will exclude all records for which coders answered “no”, but keep all records for which coders answered “yes” or “maybe”. Second, the six coders will each review each of the full-text manuscripts and code whether any sample in the manuscript fulfils the eligibility criteria (see Section 2.1). Third, we will resolve any disagreements between the six coders (by consensus) and exclude ineligible studies. At this stage, we will also search for relevant records that cite, or are cited by, any of the included studies or relevant reviews. Figure 1 illustrates the study selection process.

Figure 1. Plan for the search strategy, study selection and data extraction process.


Coders 1-6
Coders 1-6
Coders 1-6
Coders 1-6
Coder 1
Retrieve records from *PsychINFO* database
Retrieve records from *Web of Science* database
Retrieve records from *ProQuest Conference Papers & Proceedings and Dissertations & Theses* database
Retrieve records from *Europe PMC Pre-Prints* database
Retrieve records from other sources
Remove duplicates and compile meta-data (title, abstract, and reference)
Screen and exclude studies (based on article meta-data)
Code inclusion and exclusion criteria (based on full text articles)
Resolve disagreements, exclude studies based on eligibility criteria, and retrieve references/citations from included studies
Code moderators
Code moderators
Code moderators
Code moderators
Code moderators
Code moderators
Resolve disagreements
Extract effect sizes and sample sizes
Compile dataset for meta-analysis

2.6 Data collection

We will collect the relevant data in two steps. First, we will code moderating variables to be used in (exploratory) meta-regression analyses (see Section 2.9.3). We will decide exactly which moderators to code based on a first review of the included studies. So far we plan to code the categorical moderators of publication status (published, unpublished), study design (e.g., mixed, between-subjects), type of intervention strategy (e.g., accuracy nudge), intervention sub-type (e.g., active vs. passive inoculation), intervention context (e.g., conspiracy, polarisation), susceptibility measure (e.g., real vs. fake news discernment), type of sharing measure, score type (e.g., average, sum, or difference), sample nationality, partisan affiliation, and whether the sample comprises exclusively students. Further, we plan to code the continuous moderators of mean age of the sample and political ideology. Six coders will each review the included full-text manuscripts.

Firstly, the six coders will each randomly select 10 papers to extract the data from. Each paper will be double-checked by another coder to avoid mistakes and inconsistencies. Once the piloting process is completed, all six coders will extract the data and code the moderating variables from their independent *Zotero* folders. 10% of this data will then be double-checked and re-calculated in different ways by one other coder for reliability. If there are any concerns about reliability, all data from the relevant coder will be checked.

2.7 Outcome selection

Due to the rapid growth of misinformation intervention research, we aim to collect as many effect sizes to be analysed as possible. Therefore, Pearson's *r*, Cohen's *d*, Hedges' *g,* and Fisher's *z* will be collected for each data point. We do not need to prioritise selection of multiple variables on one sample due to our planned use of Robust Variance Estimation (RVE).

2.8 Individual study bias

For each data point, the publication status, mean age, sample nationality, whether the sample was exclusively students, and variable types will all be included in the moderation analyses to investigate individual study bias. As RVE is the planned analysis, this will also allow us to control for sample-level bias (i.e., multiple analyses on the same sample) and conduct sensitivity and homogeneity analyses.

2.9 Data synthesis (meta-analysis)

A portion of studies with an experimental design may not report Cohen's *d*, only the means and standard deviations for each group. Therefore, the mean, standard deviation and sample size of both study conditions will be inputted into the *Campbell Collaboration Effect Size Calculator* (see Wilson, 2001) to obtain a between-subjects Cohen's *d.* To obtain Hedges' *g*, we will input the respective group sample sizes, standard deviations, and means into the *Statology Hedges' g* calculator (Zach, 2021). The sample size for each group may not be reported; in this case, the authors will be contacted in the hopes of obtaining this information. If this fails, the total sample size will be divided equally to obtain an estimate of group sample sizes.

If an analysis employs a within-subjects design, Cohen's d_z will be calculated with the following formula:

$$d_z = \frac{m_1 - m_2}{SD_{pooled}}$$

This will then be converted into Cohen's *d* with the following formula:

$$d = d_z\sqrt{2}$$

Cohen's *d* will be converted into its Pearson's *r* equivalent effect size with the following formula:

$$r = \frac{d}{\sqrt{d^2 + a}}$$

In that formula, *a* represents 4 if both groups had equal sample sizes. However, if both sample sizes are not equal, *a* is calculated using the following formula:

$$a = \frac{(n_1 + n_2)^2}{n_1 n_2}$$

The effect size reported may be an Odds Ratio (OR). In these cases, OR will be converted into Cohen's *d* with the following formula (and subsequently into Pearson's *r* with the two previous formulas):

$$d = LogOddsRatio \times \frac{\sqrt{3}}{\pi}$$

Once all data points have a Pearson's *r* effect size equivalent, the Pearson's *r* and sample size will be inputted into the online *Campbell Collaboration Effect size calculator* to obtain Fisher's *z.*

The variance of Fisher's *z* will be calculated with the following formula:

$$V_z = \frac{1}{n-3}$$

The standard error of Fisher's *z* will be calculated with the following formula:

$$se_z = \frac{1}{\sqrt{n-3}}$$

Finally, the lower and upper bounds of the Fisher's *z* confidence interval will be calculated with the following formula:

$$lci_z \; or \; uci_z = z \pm 1.96 \times se_z$$

2.9.1 Confirmatory analyses.

We plan to conduct confirmatory analyses for all research questions (Section 1.2). Specifically, we will run separate RVE models (without moderators) to analyse the effectiveness of each of the intervention strategies for reducing both susceptibility and sharing. If significant moderation effects based on categorical variables are obtained, we will run RVE models that separate out these variables (e.g., separate RVE models for Republican and Democrat partisan affiliations).

2.9.2 Sensitivity analyses.

For RVE, sensitivity analysis is important because if between-study covariances are considerably smaller than within-study covariances, the estimated effect size may not be appropriately weighted for the meta-regression model (see Fisher & Tipton, 2015, p. 9–10). Therefore, sensitivity in this case will be used to analyse the average effect size and its associated standard error for different values of our conservative *rho* coefficient range (0, 1). If there is considerable variation in effect size estimation when sensitivity analysis is run on a model, a more conservative coefficient interval will be selected, and the model will be re-run.

If *df* < 4 in any RVE model, this may indicate unbalanced covariates due to too few samples included in the analysis (see Fisher & Tipton, 2015). In this case, the significance level will be adjusted to the more conservative $p < .01$. We will run analyses of influential cases on all models to determine whether any of the effect sizes are outliers, and thus potentially biasing our results. If any cases are considered outliers, we will consider *winsorising* them.

We will run exploratory models using alternative heterogeneity estimators as additional robustness checks to determine the relative Probability of Benefit (POB) of applying certain methods over others (see van der Linden & Goldberg, 2020). These will include Residual Maximum Likelihood (REML), the DerSimonian and Laird approach (DL), and the Hartung, Knapp, Sidik and Jonkman (HKSJ) method. If results of a given model are sensitive to the estimation method applied, we will consider which method to apply based on the consistency and nature of the data analysed.

2.9.3 Exploratory analyses.

Firstly, if $df < 4$ in any of the RVE models, we will especially consider the range of our credibility intervals in comparison to the effect size distribution (see Wiernik et al., 2017). Specifically, smaller effect sizes require tighter credibility intervals than larger effect sizes. Therefore, we will compare the credibility intervals of our smaller study size analyses to discuss and infer their robustness.

After all confirmatory analyses are completed, the moderation models will be run. Specifically, Wiernik et al. (2017) recommends running moderation models on effects that show wide credibility intervals in comparison with effect size distribution. All effects with comparatively wide credibility intervals will be tested for moderation effects.

2.9.4 Meta-biases

To test for publication bias, we will produce contour-enhanced funnel plots with ‘trim-and-filled’ corrections, alongside Kendall’s Rank Tests and a regression coefficient between the meta-analytic effect size and τ^2 to assess funnel plot asymmetry. We will also conduct *P-Uniform* analyses to detect potential publication bias and obtain corrected effect sizes. To test the evidential value of our data, we will produce *P-curves* for the respective analyses. We will also conduct Bayesian analyses on the main effects to determine the likelihood of the directional hypotheses over the null.

References

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs*, *77*(3), 281-311. https://doi.org/10.1080/03637751003758193

Basol, M., Roozenbeek, J., & Van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, *3*(1), 2. https://doi.org/10.5334/joc.91

Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, *28*(11), 1531-1546. https://doi.org/10.1177/0956797617714579

Compton, J. (2013). Inoculation theory. *The Sage handbook of persuasion: Developments in theory and practice*, *2*, 220-237.

Fazio, L. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-009

Fisher, Z., & Tipton, E. (2015). *robumeta: An R-package for robust variance estimation in meta-analysis.* PsyarXiv. https://doi.org/arXiv: 1503.02220

Hameleers, M. (2020). Separating truth from lies: Comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands. *Information, Communication & Society*, 1-17. https://doi.org/10.1080/1369118x.2020.1764603

Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, *27*(1), 1-16. https://doi.org/10.1037/xap0000315

McGuire, W. J. (1964). Inducing resistance to persuasion. Some contemporary approaches. *CC Haaland and WO Kaelber (Eds.), Self and Society. An Anthology of Readings, Lexington, Mass.(Ginn Custom Publishing) 1981, pp. 192-230.*

Oreskes, N., & Conway, E. M. (2010). Defeating the merchants of doubt. *Nature*, *465*(7299), 686-687. https://doi.org/10.1038/465686a

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. https://doi.org/10.31234/osf.io/3n9u8

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy nudge intervention. https://doi.org/10.31234/osf.io/uhbk9

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991-996. https://doi.org/10.1016/j.jclinepi.2007.11.010

Roozenbeek, J., & van der Linden, S. (2019a). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1). https://doi.org/10.1057/s41599-019-0279-9

Roozenbeek, J., & van der Linden, S. (2019b). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research*, *22*(5), 570-580. https://doi.org/10.1080/13669877.2018.1443491

Roozenbeek, J., van der Linden, S., & Nygren, T. (2020). Prebunking interventions based on the psychological theory of "inoculation" can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016//mr-2020-008

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638-641. https://doi.org/10.1037/0033-2909.86.3.638

Saleh, N. F., Roozenbeek, J., Makki, F. A., McClanahan, W. P., & van der Linden, S. (2021). Active inoculation boosts attitudinal resistance against extremist persuasion techniques: A novel approach towards the prevention of violent extremism. *Behavioural Public Policy*, 1-24. https://doi.org/10.1017/bpp.2020.60

van der Linden, S., & Goldberg, M. H. (2020). Alternative meta-analysis of behavioral interventions to promote action on climate change yields different conclusions. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-17613-7

van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, *1*(2), 1600008. https://doi.org/10.1002/gch2.201600008

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

Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2019). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, *37*(3), 350-375. https://doi.org/10.1080/10584609.2019.1668894

Wiernik, B. M., Kostal, J. W., Wilmot, M. P., Dilchert, S., & Ones, D. S. (2017). Empirical Benchmarks for Interpreting Effect Size Variability in Meta-Analysis. *Industrial and Organizational Psychology*. https://doi.org/10.1017/iop.2017.44

Wilson, D. B. (2001). *Practical Meta-Analysis Effect Size Calculator* [Online Calculator]. https://campbellcollaboration.org/research-resources/effect-size-calculator.html

Zach (2021). *Hedges' g calculator* [Online Calculator]. Statology. https://www.statology.org/hedges-g-calculator/