Papers related to misinformation from **David Rand** and **Gordon Pennycook**'s research team

Key papers

- → The Psychology of Fake News <u>TiCS 2021 [X thread] [15 minute video summary</u>]
- → Durably reducing conspiracy beliefs through dialogues with AI <u>Science 2024 [NYTimes write up]</u>
 [X thread] [BlueSky thread] [9min presentation] [45min presentation] [Experimental materials]
 [Browse the conversations] [Try the bot yourself!]
- → Shifting attention to accuracy can reduce misinformation online Nature 2021 [X thread]
- → Understanding and Combatting Online Misinformation Across 16 Countries on Six Continents

 Nature Human Behaviour 2023 [X thread] [30 minute video presentation]
- → Quantifying the Impact of Misinformation and Vaccine-Skeptical Content on Facebook Science 2024 [X thread] [Scientific American op ed]
- → Differences in misinformation sharing can lead to politically asymmetric sanctions <u>Nature 2024</u> [<u>BlueSky thread</u>] [X thread]



Hyperlinked Table of Contents:

Key papers

Overview of our methodological approach Evaluating interventions to fight misinformation

Human-AI dialogues

Accuracy prompts

Crowdsourcing

Warnings/Corrections

Source Information

Role of reasoning in detecting versus falling for misinformation

Experimental papers

Correlational papers

Cross-cultural papers

Elite Party Cues/Messaging

Illusory truth and the effect of repetition

Politically motivated reasoning (or lack thereof)

Polarization, echo chambers, and moral outrage

Other papers related to misinformation

Overview of our methodological approach

- → A practical guide to doing behavioural research on fake news and misinformation Collabra:

 Psychology 2021 [X thread]
- → *REVIEW*: Field experiments on social media <u>Current Directions in Psychological Science 2021</u> [X thread]
- → High level of concordance across different news domain quality ratings PNAS Nexus 2023 [X thread]
- → How to think about whether misinformation interventions work Nature Human Behaviour 2023 [X thread]

Evaluating interventions to fight misinformation

Popular overview:

The Right Way to Fight Fake News NYTimes 2020 [X thread]

Human-AI dialogues

- → Durably reducing conspiracy beliefs through dialogues with AI <u>Science 2024 [NYTimes write up]</u>
 [X thread] [BlueSky thread] [9min presentation] [45min presentation] [Experimental materials]
 [Browse the conversations] [Try the bot yourself!]
- → Just the facts: How dialogues with AI reduce conspiracy beliefs <u>Working paper</u> [<u>BlueSky thread</u>] [X thread]
- → The Levers of Political Persuasion with Conversational AI <u>Working paper [BlueSky thread] [X thread]</u>
- → AI reduces conspiracy beliefs even when presented as a human expert <u>Working paper [BlueSky thread]</u>
- → Reducing belief in conspiracy theories as they unfold using large language models <u>Working paper</u> [<u>BlueSky thread</u>] [X thread]
- → Addressing climate change skepticism and inaction using human-AI dialogues <u>Working paper</u> [<u>BlueSky thread</u>] [<u>Try the bot yourself</u>] [<u>Browse the conversations</u>]
- → Personalized Dialogues with AI Effectively Address Parents' Concerns about HPV Vaccination Working paper [BlueSky thread] [Twitter thread] [Try the bot yourself]
- → Deep canvassing using AI Working paper [BlueSky thread] [Browse the conversations]
- → It's the Thought that Counts: Evaluating the Attempts of Frontier LLMs to Persuade on Harmful Topics Working paper
- → How Malicious AI Swarms Can Threaten Democracy Working paper

Accuracy prompts

- → REVIEW: Nudging Social Media Sharing Towards Accuracy AAAPSS 2022
- → Shifting attention to accuracy can reduce misinformation online Nature 2021 [X thread] [Scientific American op ed]
- → Reducing misinformation sharing at scale using digital accuracy prompt ads Working paper [X thread] [45 minute lecture video]
- → Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation Nature Comm 2022 [X thread]
- → Thinking more or thinking differently? Using drift-diffusion modeling to illuminate why accuracy prompts decrease misinformation sharing Cognition 2023 [X thread]
- → The social media context interferes with truth discernment Science Advances 2023 [X thread]
- → Inoculation and accuracy prompting increase accuracy discernment in combination but not alone Nature Human Behavior 2024 [X thread]
- → On the Efficacy of Accuracy Prompts Across Partisan Lines: An Adversarial Collaboration Psychological Science 2024 [X thread]
- → Examining partisan asymmetries in fake news sharing and the efficacy of accuracy prompt interventions Working paper [X thread]
- → Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy nudge intervention Psychological Science 2020 [X thread]
- → Developing an accuracy-prompt toolkit to reduce COVID-19 misinformation online
 <u>Misinformation Review 2021">HKS</u>

 Misinformation Review 2021">MISINFORMATION (NEW 2021)

 Misinformation Review 2021">MISINFORMATION (NEW 2021)

 Misinformation Review 2021">MISINFORMATION (NEW 2021)

 Misinformation Review 2021">MISINFORMATION (NEW 2021)

 MISINFORMATION (NEW 2021) MISINFORMATION (
- → Exploring Lightweight Interventions at Posting Time to Reduce the Sharing of Misinformation on Social Media <u>CSCW 2021</u>
- → Examining accuracy-prompt efficacy in combination with using colored borders to differentiate news and social content online HKS Misinfo Review 2023 [X thread]
- → Accuracy prompts protect professional content moderators from the illusory truth effect PNAS
 Nexus 2024 [BlueSky thread]

Crowdsourcing

- → *REVIEW*: Crowds Can Effectively Identify Misinformation at Scale <u>Perspectives on Psych Science 2023</u> [<u>Twitter thread</u>]
- → Scaling up fact-checking using the wisdom of crowds <u>Science Advances 2021</u> [X thread] [<u>Time op</u> ed]
- → Political motives help rather than hinder crowdsourced fact-checking Working paper
- → Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program CHI'22 [X thread] [Financial Times Op Ed]
- → Fighting misinformation on social media using crowdsourced judgments of news source quality PNAS 2019 [X thread] [The Hill op ed]

- → Will the crowd game the algorithm? Using layperson judgments to combat misinformation on social media by downranking distrusted sources. CHI '20 [X thread]
- → Republicans are flagged more often than Democrats for sharing misinformation on X's Community Notes PNAS 2025 [BlueSky thread] [LinkedIn thread]

Warnings/Corrections

- → *REVIEW*: Misinformation warning labels are widely effective: A review of warning effects and their moderating features <u>Current Opinion in Psychology 2023</u> [X thread]
- → The Implied Truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings Management Science 2020 [X thread]
- → Fact-checker warning labels are effective even for those who distrust fact-checkers Nature Human Behaviour 2024 [Twitter thread]
- → Labeling AI-Generated Media Online PNAS Nexus 2025
- → Labeling AI-Generated Content: Promises, Perils, and Future Directions MIT Policy Paper 2024
- → What label should be applied to content produced by generative AI? Working paper [X thread]
- → Perceived legitimacy of layperson and expert content moderators PNAS Nexus 2025
- → Partisan consensus and divisions on content moderation of misinformation Working paper
- → Perverse Downstream Consequences of Debunking: Being Corrected by Another User for Posting False Political News Increases Subsequent Sharing of Low Quality, Partisan, and Toxic Content in a Twitter Field Experiment CHI'21 [X thread]
- → Timing matters when correcting fake news PNAS 2021 [X thread]
- → Unbundling Digital Media Literacy Tips: Results from Two Experiments Working paper
- → Promoting engagement with social fact-checks online <u>PLoS ONE 2025</u>
- → Do explanations increase the effectiveness of AI-crowd generated fake news warnings? ICWSM
 ^{'22} [X thread]
- → You're definitely wrong, maybe: Correction style has minimal effect on corrections of misinformation online Media and Communication 2021 [X thread]
- → Debunking Handbook 2020 [PDF] [X thread]
- → How effective are TikTok misinformation debunking videos? <u>HKF Misinformation Review 2023</u>
 [X thread]

Source Information

- → Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media HKS Misinformation Review 2020 [X thread]
- → Displaying News Source Trustworthiness Ratings Reduces Sharing Intentions for False News Posts <u>J Trust and Safety 2023 [X thread]</u>
- → Partisan Source Cues and Trust in Global News Working paper
- → Depoliticizing Public Media: Polarization and the Dynamics of Media Trust in Poland Working

Role of reasoning in detecting versus falling for misinformation

Popular overview: "Why do people fall for fake news?" NYTimes op ed 2019

Experimental papers

- → Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines JEP:General 2020 [X thread]
- → Reliance on emotion promotes belief in fake news <u>Cognitive Research</u>: <u>Principles and Implications</u> 2020 [X thread] [Blog post]
- → Reasoning about climate change PNAS Nexus 2023 [X thread]
- → Emotion may predict susceptibility to fake news but emotion regulation does not seem to help Cognition and Emotion 2022 [X thread]
- → Does deliberation decrease belief in conspiracies? <u>Journal of Experimental Social Psychology 2022</u>
- → Uncommon Errors: Adaptive Intuitions in High-Quality Media Environments Increase Susceptibility to Misinformation Working paper [X thread]
- → Emotional language reduces belief in false claims <u>JDM 2025</u> [X thread]

Correlational papers

- → Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning Cognition 2019
- → Cognitive reflection correlates with behavior on Twitter Nature Communications 2021 [X thread]
- → Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking <u>Journal of Personality 2019</u>
- → Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking <u>Journal of Applied Memory & Cognition 2019</u>
- → Beyond "fake news": Analytic thinking and the detection of false and hyperpartisan news headlines <u>Judgment and Decision Making 2021</u>
- → What Makes News Shareable on Social Media? <u>Journal of Quantitative Description: Digital Media</u> 2023 [X thread]
- → Digital literacy is associated with more discerning accuracy judgments but not sharing intentions <u>HKS Misinfo Review 2021</u> [X thread] [Scientific American op ed]
- → Overconfidently conspiratorial: Conspiracy believers are dispositionally overconfident and massively overestimate how much others agree with them <u>Personality and Social Psychology</u> <u>Bulletin 2025</u> [<u>BlueSky thread</u>]

Cross-cultural papers

→ Beliefs about COVID-19 in Canada, the U.K., and the U.S.A.: A novel test of political polarization

- and motivated reasoning Personality and Social Psychology Bulletin 2022 [X thread]
- → Happiness and surprise are associated with worse truth discernment of COVID-19 headlines among social media users in Nigeria HKS Misinformation Review, 2021 [X thread]
- → Does Analytic Thinking Insulate Against Pro-Kremlin Disinformation? Evidence from Ukraine.

 Political Psychology 2022 [X thread] [SPSP Blog post]

Elite Party Cues/Messaging

- → Measuring exposure to misinformation from political elites on Twitter <u>Nature Communications</u> 2022 [X thread] [Web app to show any Twitter user's exposure to elite misinfo]
- → Partisans' receptivity to persuasive messaging is undiminished by countervailing party leader cues Nature Human Behaviour 2023 [X thread] [The Hill op ed]
- → Correcting Misperceptions of Out-partisans Decreases American Legislators' Support for Undemocratic Practices PNAS 2023 [X thread]
- → Elite Party Cues Increase Vaccination Intentions among Republicans PNAS 2021 [X thread] [WashPo Op Ed]
- → Pro-Climate Statements from Elon Musk can Persuade Republicans on Climate Change Working

 Paper
- → Examining false beliefs about voter fraud in the wake of the 2020 Presidential Election <u>HKS</u> <u>Misinformation Review 2021 [X thread] [The Hill op ed]</u>

Illusory truth and the effect of repetition on belief

- → Prior exposure increases perceived accuracy of fake news JEP:General 2018
- → Blatantly false news increases belief in news that is merely implausible Working paper
- → Repetition increases perceived truth equally for plausible and implausible statements. <u>Psychonomic Bulletin & Review 2019</u>
- → Replicability and generalizability of the repeated exposure effect on moral condemnation of fake news Nature Communications 2025
- → Investigating the robustness of the illusory truth effect across individual differences in cognitive ability, need for cognitive closure, and cognitive style. Personality and Social Psychology Bulletin 2019
- → Illusory Truth Occurs Even with Incentives for Accuracy Working paper [X thread]
- → The Not So Illusory Truth Effect: A Rational Foundation for Repetition Effects Working paper

Politically motivated reasoning (or lack thereof)

- → Thinking clearly about causal inferences of politically motivated reasoning: Why paradigmatic study designs often prevent causal inference <u>Current Opinion in Behavioral Science_2020</u> [X thread]
- → Rethinking the link between cognitive sophistication and identity-protective bias in political belief formation JEP:General 2021 [X thread]

- → No association between numerical ability and politically motivated reasoning in a large US probability sample PNAS 2023
- → Bayesian or biased? Analytic thinking and political belief updating Cognition 2020 [X thread]
- → Increasing accuracy motivations using moral reframing does not reduce Republicans' belief in false news HKS Misinfo Review 2023
- → Science beliefs, political ideology, and cognitive sophistication <u>JEP:General 2022</u> [X thread]
- → On the belief that beliefs should change according to evidence: Implications for conspiratorial, moral, paranormal, political, religious, and science beliefs <u>Judgment and Decision Making 2020</u> [X thread]
- → Sincere or Motivated? Partisan Bias in Non-political Information Processing <u>Judgment and</u>
 <u>Decision Making 2023</u>
- → Self-persuasion does not imply self-deception Cognition 2025

Polarization, echo chambers, and moral outrage

- → Megastudy testing 25 treatments to reduce antidemocratic attitudes and partisan animosity <u>Science</u> 2024 [X thread]
- → Information Gerrymandering and Undemocratic Decisions Nature 2019 [X thread]
- → Political Sectarianism in America Science 2020 [X thread]
- → Divergent patterns of engagement with partisan and low-quality news across seven social media platforms PNAS 2025 [BlueSky thread] [X thread]
- → Shared Partisanship Dramatically Increases Social Tie Formation in a Twitter Field Experiment PNAS 2021 [X thread]
- → Partisans neither expect nor receive reputational rewards for sharing falsehoods over truth online PNAS Nexus 2024
- → Psychological underpinnings of partisan bias in tie formation on social media <u>JEP:General 2024</u> [X thread]
- → Blocking of counter-partisan accounts drives political assortment on Twitter PNAS Nexus 2024 [X thread]
- → Misinformation and harmful language are interconnected, rather than distinct, challenges <u>PNAS</u>

 Nexus 2024 [X thread] [BlueSky thread]
- → Interventions Reducing Affective Polarization Do Not Necessarily Improve Anti-Democratic Attitudes Nature Human Behaviour 2022 [X thread]
- → From Foe to Friend and Back Again: The Temporal Dynamics of Intra-Party Bias in the 2016 U.S. Presidential Election <u>Judgment and Decision Making 2019</u>
- → Signaling when no one is watching: A reputation heuristics account of outrage and punishment in one-shot anonymous interactions <u>JPSP 2020 [X thread] [NYtimes Op Ed]</u>
- → Why Do We Hate Hypocrites? Evidence for a Theory of False Signaling Psych Science 2017

 [NYTimes Op Ed]

- → Third-party punishment as a costly signal of trustworthiness Nature 2016 [NYTimes Op Ed]
- → Dynamic Remodeling of In-Group Bias During the 2008 Presidential Election PNAS 2009

Other papers related to misinformation

- → The (Minimal) Persuasive Advantage of Political Video over Text PNAS 2021 [X thread]
- → The distorting effects of producer strategies: Why engagement does not reveal consumer preferences for misinformation PNAS 2024 [X thread]
- → Quantifying the Potential Persuasive Returns to Political Microtargeting PNAS 2023 [X thread]
- → Addressing misperceptions takes more than combating fake news <u>TiCS 2025</u>
- → Character Deprecation in Fake News: Is it in Supply or Demand? <u>Group processes and intergroup</u> relations 2021
- → Cognitive reflection and the 2016 U.S. Presidential Election <u>Personality and Social Psychology</u> Bulletin 2019
- → Human Cooperation and the Crises of Climate Change, COVID-19, and Misinformation <u>Annual Review of Psychology 2022</u>