

RESEARCH ARTICLE | SPECIAL ISSUE: FROM VULNERABILITY TO VIGILANCE

True discernment or blind scepticism? Comparing the effectiveness of four conspiracy belief interventions

Cian O'Mahony^{1*}, Gillian Murphy¹ & Conor Linehan¹

Received: May 3, 2024 | **Accepted:** November 6, 2024 | **Published:** November 15, 2024 | **Edited by:** Jonas R. Kunst

¹School of Applied Psychology, University College Cork. *Please address correspondence to Cian O'Mahony, cianomahony@ucc.ie, School of Applied Psychology, University College Cork, College Road, Cork T12 K8AF, Ireland. This article is published under the Creative Commons BY 4.0 license. Users are allowed to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

In this research, we compared the effectiveness of four promising interventions (Priming, Inoculation, Active Inoculation, and Discernment), on a range of conspiracy thinking measures. Across two studies (total $N = 1,766$), we found that the inoculation-based interventions (but not Priming and Discernment conditions) were effective at reducing susceptibility to novel implausible conspiracy theories but did not improve critical appraisal of novel plausible conspiracy theories. We found that only the Discernment condition, which discouraged blind scepticism of conspiracy theories, significantly improved critical appraisal of both plausible and implausible conspiracy theories. The inoculation-based interventions (but not Priming and Discernment conditions) were moderately successful at reducing epistemically unwarranted beliefs. However, no intervention statistically significantly reduced general conspiracy ideation or significantly affected likelihood judgements for hypothetical conspiracy theories. The overall intervention effects ranged from small to moderate ($ds = 0.14 - 0.72$). These findings highlight both the importance of teaching discernment and measuring discernment as an outcome, as many well-established interventions designed to reduce belief in conspiracy theories may have either no effect or negatively impact participants' ability to critically reason about plausible conspiracy theories.

Keywords: conspiracy theories, critical appraisal, discernment, inoculation, intervention, misinformation

1. INTRODUCTION

Conspiracy theories are arguments put forward by individuals or groups, suggesting that powerful agents act covertly in the pursuit of often malicious goals (Wood et al., 2012). At their core, they are a means for people to make sense of the world in times and situations of uncertainty (Douglas et al., 2017). Previous research has shown that conspiracy beliefs are associated with several negative consequences on both a personal and societal level, such as anti-social behaviour, dangerous health practices, political extremism, and science denial (Douglas et al., 2017; Erisen, 2022; Grawitch & Lavigne, 2021; Jolley et al., 2022; Sallam et al., 2021; van Prooijen et al., 2015). As a result, researchers have become increasingly interested in designing interventions to challenge belief in conspiracy theories in order to mitigate the harms they may facilitate (O'Mahony et al., 2023).

It is important to note at the outset that conspiracy theories are not inherently untrue. Indeed, many real conspiracies have been exposed throughout history by journalists and activists. In this paper, we distinguish between simply believing a conspiracy to be afoot (which is not in itself inaccurate or harmful) and believing in a conspiracy theory that is very unlikely to be true, for example, due to a lack of evidence, or due to the absurdly complex explanation put forward. This might include a conspiracy theory that hinges on the covert co-operation of multiple state bodies, the media, and the judicial system, such as the claim that Joe Biden stole the 2016 US Presidential election (Stall & Petrocelli, 2023).

In a recent systematic review, researchers compared 25 conspiracy interventions (O'Mahony et al., 2023). The review showed that a wide variety of approaches have been used to challenge conspiracy beliefs, including narrative interventions (Nera et al., 2018), debunking (Stojanov, 2015), and ridiculing conspiracy beliefs (Orosz et al., 2016). Most interventions were

found to be ineffective in terms of reducing conspiracy beliefs, with small to very small effects observed (O'Mahony et al., 2023). However, the review did highlight three promising interventions: analytical priming, informational inoculations, and critical thinking education, all of which had medium-to-large effects on reducing conspiracy beliefs. To date, no study has compared the effectiveness of these interventions directly, and it is difficult to fairly assess their effectiveness based on current evidence, as each respective study addressed different topics and used different measures of conspiracy belief.

1.1 Potential Interventions

The three interventions identified as promising in a recent systematic review, in ascending order of participant engagement, are analytical priming, informational inoculations and critical thinking instruction (O'Mahony et al., 2023). The evidence in support of these interventions has accumulated from both conspiracy-focused research, and studies focused on more general forms of misinformation and “fake news”. While misinformation and conspiracy beliefs are distinct (De Coninck et al., 2021), interventions aimed at tackling misinformation and conspiracy theories have the shared goal of challenging unfounded beliefs.

Analytical priming interventions aim to prime participants to adopt an analytical mindset before they encounter conspiracy-related content. Priming is carried out through a number of methods, from reading text in difficult-to-read fonts (Swami et al., 2014) to exercises that prime participants to be resistant to persuasion (Bonetto et al., 2018). Initial research suggested that priming interventions could reduce scores on unfounded conspiracy belief questionnaires, with small-to-medium effects ($d = 0.30\text{--}0.56$; Bonetto et al., 2018; Swami et al., 2014). Other analytical priming style interventions nudge participants to question the accuracy of information (Butler et al., 2024; Epstein et al., 2021; Fazio, 2020; Pennycook et al., 2020; Pennycook

& Rand, 2022a), or present participants with a general warning (Greene & Murphy, 2021). While the systematic review conducted by O'Mahony et al. (2023) identified analytical priming as a promising intervention based on studies published pre-2021, a number of studies have since emerged that have suggested the evidence for the efficacy of priming interventions is weak. Follow-up studies with greater sample sizes and statistical power have failed to replicate the findings of priming interventions, indicating that earlier evidence for their success may have been due to false positives (Janouskova et al., 2022; Roozenbeek et al., 2021; Večkalov et al., 2024). Furthermore, follow-up studies have suggested that some of the methods of priming, such as having participants read difficult-to-read font, may have failed to manipulate the intended construct (Janouskova et al., 2022). Despite their failure to replicate, priming interventions are still a common intervention method among researchers (Ceylan et al., 2023; Lee & Jang, 2023; Pennycook & Rand, 2022a, 2022b), and so we included a priming condition in this study.

Secondly, informational inoculations, often referred to as “pre-bunking”, have become a popular style of intervention for tackling various forms of misinformation, from fake news to unreasonable conspiracy theories (Banas & Miller, 2013; Banas & Rains, 2010; Maertens et al., 2020; Vivion et al., 2022; Williams & Bond, 2020). Similar to how a vaccine exposes individuals to a weakened form of a virus to build resistance, informational inoculations seek to confer resistance to misinformation before individuals encounter it directly (Banas & Rains, 2010; McGuire, 1964; Pfau et al., 2005). In controlled studies, informational inoculations have been effective in reducing unfounded conspiracy beliefs such as anti-vaccine and 9/11 conspiracy theories. For example, Banas and Miller (2013) found that participants who received informational inoculations believed in fewer 9/11 conspiracy theories than those in a control group ($d = 1.31$). Similarly, a study by Jolley and

Douglas (2017) found that when participants were asked to indicate their intention to vaccinate their fictional child against an imaginary disease, participants' vaccine intentions improved if they were presented with anti-conspiracy arguments before they were exposed to conspiracy theories. Additionally, Jolley and Douglas found these anti-vaccination informational inoculations had medium effects when compared to controls in terms of reducing anti-vaccination conspiracy beliefs ($d = 0.42$).

Many informational inoculation techniques focus on refuting a specific conspiracy theory (i.e., anti-vaccine conspiracies). However, topic-specific inoculations are difficult to scale, as they are restricted to only challenging specific beliefs (Traberg et al., 2022). To address the limitations of topic-specific inoculations, researchers have also tested whether inoculations could be used to target larger concepts, such as highlighting rhetoric techniques commonly used in misinformation. One study found that teaching participants how the tobacco industry commonly appeals to fake experts transferred to resistance against climate denial (Cook et al., 2017). These findings not only suggest that inoculations can target the processes underlying inaccurate beliefs rather than refuting specific false statements, but that the effects could transfer to different topics. This broad-spectrum resistance to misinformation has been referred to as the *blanket-of-protection* (Banas & Rains, 2010; Parker et al., 2012). This blanket-of-protection makes these ‘broad-spectrum’ inoculations much more desirable as interventions, as they can be scaled to tackle different topics.

Finally, another effective approach to tackling conspiracy beliefs is critical thinking education. A study by Dyer and Hall (2019) found that students who enrolled in a pseudoscience class that explicitly taught them to critically appraise unreasonable conspiracy theories held significantly fewer conspiracy beliefs than those who enrolled in a conventional research methods class, with a large effect ($d = 1.07$). One of the key

features of the pseudoscience class was that it required students to actively engage with the lessons and apply the skills they had learned. The students also received feedback from their professors. Due to the emphasis on participant engagement and feedback, the critical thinking education intervention used by Dyer and Hall (2019) can be thought of as a form of *active inoculation*. Unlike standard informational inoculations, active inoculations are “a two-way process through which participants engage interactively” (Kiili et al., 2024, p. 3). In the pseudoscience class, participants actively generated arguments for and against various unsubstantiated claims (Dyer & Hall, 2019). McGuire, who proposed the original theory of informational inoculations, had also argued that when participants are required to make arguments for and against the misinformation topic at hand, this would engage more cognitive processes and ultimately be a more effective intervention (McGuire & Papageorgis, 1961). Recent research has suggested that active inoculations may show more promise as a means of challenging conspiracy beliefs (Kiili et al., 2024; Roozenbeek & van der Linden, 2019).

1.2 Effectively Measuring Conspiracy Beliefs

It is difficult to compare the relative effectiveness of the different styles of conspiracy belief interventions discussed in the previous section based on existing evidence because studies in this field use a wide range of disparate methods for assessing conspiracy beliefs. Some studies ask participants to rate their belief in specific conspiracy statements on topics such as anti-vaccine sentiments (“Misrepresentation of the efficacy of vaccines is motivated by profit”, Jolley & Douglas, 2017) or 9/11 conspiracy theories (“The United States government participated in a conspiracy to perpetrate the attack on 9/11”, Banas & Miller, 2013). Other studies measure general conspiracy ideation (“A lot of important information is deliberately concealed from the public out of self-interest”, Bonetto et al., 2018; Poon et al., 2020; Stojanov, 2015).

One fundamental problem with existing forms of assessment is that they typically only include items where participants should reject any and all conspiracies, or forms of conspiratorial thinking. Therefore, a “successful” intervention is one in which participants learn to reject conspiracies. This presents a problem, as summed up by Frenken et al. (2024) because “just because it’s a conspiracy theory, doesn’t mean they’re not out to get you!” (p. 1). Conspiracies can and do happen – consider MKUltra, where the CIA conducted a number of deeply unethical human experiments for twenty years (Valentine, 2017). We argue that a successful intervention should improve participants’ ability to selectively reject *unreasonable* conspiracy theories, not simply encourage them to disavow anything that sounds conspiratorial in nature.

This is not a problem that is unique to conspiracy research. Recent research has shown that only 7% of misinformation studies assess any sort of discernment between true or false items (Murphy et al., 2023). Guay and colleagues (2022) argue that outcome measures used in misinformation interventions should include both true and false statements, allowing researchers to measure discernment rather than just measuring blind scepticism. Importantly, none of the studies identified by O'Mahony et al. (2023) that reported promising conspiracy belief intervention techniques included an assessment of discernment. It is, therefore, currently unclear whether these interventions are effective in teaching participants to identify unreasonable conspiracy theories, or whether they are merely teaching participants to reject anything that sounds conspiratorial. This is not to suggest that discernment (or a lack thereof) is an underlying process of conspiracy beliefs. However, it is important to establish whether interventions are teaching participants how to identify unfounded conspiracy theories, rather than simply encouraging a response bias to dismiss all conspiracy theories.

There has been a recent move towards

measuring belief in both plausible and implausible conspiracy items in conspiracy research. For example, Frenken et al. (2024) and Hattersley et al. (2022) included these measures in their recent studies of correlates of conspiracy belief, though neither reported discernment as an outcome measure. In the current study, we will utilise The Critical Thinking about Conspiracies (CTAC) test as our primary outcome measure. This is a psychometrically validated assessment that measures how participants respond to both implausible and plausible novel conspiracy theories (O'Mahony et al., 2024).

Another benefit of using the CTAC lies in its focus on measuring the underlying cognitive process of conspiracy thinking, as opposed to measuring belief in specific conspiracy theories or conspiracy ideation (O'Mahony et al., 2024), conferring several advantages. The first is that it moves beyond measuring whether a person believes in a particular unreasonable conspiracy theory but allows researchers to measure the justification a participant chooses for why a conspiracy is (un)reasonable. The second advantage is that focusing on the process of conspiracy thinking rather than specific conspiracy beliefs prevents the assessment from being confined by time and culture. We argue that measuring the process of conspiracy thinking will garner a greater understanding of how interventions affect (or fail to affect) conspiracy thinking.

1.3 Overview of Present Research

In this paper, we compare the efficacy of four interventions in increasing critical appraisal of conspiracy theories, across two studies. Study 1 consists of two parts. In Study 1a, we sought to examine the effects of Priming, Inoculation, and Active Inoculation interventions in increasing critical appraisal of conspiracy theories and reducing epistemically unwarranted beliefs. In Study 1b, we added an additional intervention arm (Discernment condition) to our existing sample and investigated whether it would

prevent blind scepticism of conspiracy theories. Finally, due to our primary outcome measure not meeting acceptable validity thresholds in Study 1, we attempted to replicate our findings with a more robust assessment of critical appraisal of conspiracy theories in Study 2, along with the inclusion of two more measures of conspiracy thinking.

2. STUDY 1A

We set out to compare three conspiracy interventions: a Priming condition, an Inoculation condition, and an Active Inoculation condition. We tested two hypotheses. First, we expected that participants who receive any style of conspiracy intervention will report better critical appraisal of conspiracy theories than a control condition. Second, as active involvement increases across the four conditions (Control, Nudge, Inoculation, Active Inoculation), we expected that critical appraisal of conspiracy theories will increase.

2.1 Methods

The current study was preregistered (available here: https://aspredicted.org/C83_QBD). All data and materials for this study are available here: <https://osf.io/suw57/>.

2.1.1 Design

A between-subjects design was used for this study. Participants were randomly assigned to one of four intervention conditions: Control, Priming, Inoculation, Active Inoculation. The Control participants received no intervention and proceeded to complete the rest of the survey.

2.1.2 Participants

Power analysis using GPower version 3.1.9.7 indicated that 1095 participants were required to detect a Cohen's f effect size of 0.1 with 80% power, with a one-way ANOVA. An initial sample of 960 participants were recruited through Prolific, an online participant recruitment service (www.Prolific.com), along with 96 from the student mailing list of an Irish University, and

40 participants from the Reddit *r/sample size* online discussion forum. According to our pre-registered analysis plan, nine participants were removed from the dataset as they failed two attention checks. We replaced these participants by recruiting an additional nine participants from Prolific. We then excluded 63 participants for failing comprehension checks on the intervention material, in line with our preregistration. The final analysis was conducted on 1032 participants (625 women; $M_{age} = 41.68$, $SD_{age} = 13.86$). Most participants were from the United Kingdom (72%), with the remaining participants from Ireland (9%), the United States (5%), Australia (5%), Canada (4%), and a mix of European countries (5%).

2.1.3 Materials

Priming. The Priming condition was designed to warn participants about the dangers of believing in unfounded conspiracy theories. We chose to use a nudge message over other methods of priming analytical thinking, as previous research has critiqued the validity of task-based priming interventions such as using difficult-to-read fonts in surveys (Janouskova et al., 2022). Furthermore, we found that nudge messages were more aligned with priming strategies that are commonly used by government initiatives such as the UK government's "Don't feed the Beast" campaign to tackle Covid-19 misinformation (U.K. Government, 2020). The Priming intervention that we used in the study was modelled on academic research and government campaigns that utilise simple warnings to challenge unwarranted beliefs, such as fake news and misinformation (i.e., Greene & Murphy, 2021; Media Literacy Ireland, 2020; U.K. Government, 2020) and presented participants with a simple warning paragraph that told participants that conspiracy theories can often be misleading. The warning message in the Priming condition was:

"Conspiracy theories are unverified claims or explanations that seek to connect unrelated events or actions in a way that implies secret

or covert coordination. They often involve powerful organizations or individuals conspiring to achieve a hidden agenda. For instance, a common conspiracy theory is that the COVID-19 pandemic is a hoax created by governments and pharmaceutical companies to control and manipulate the population.

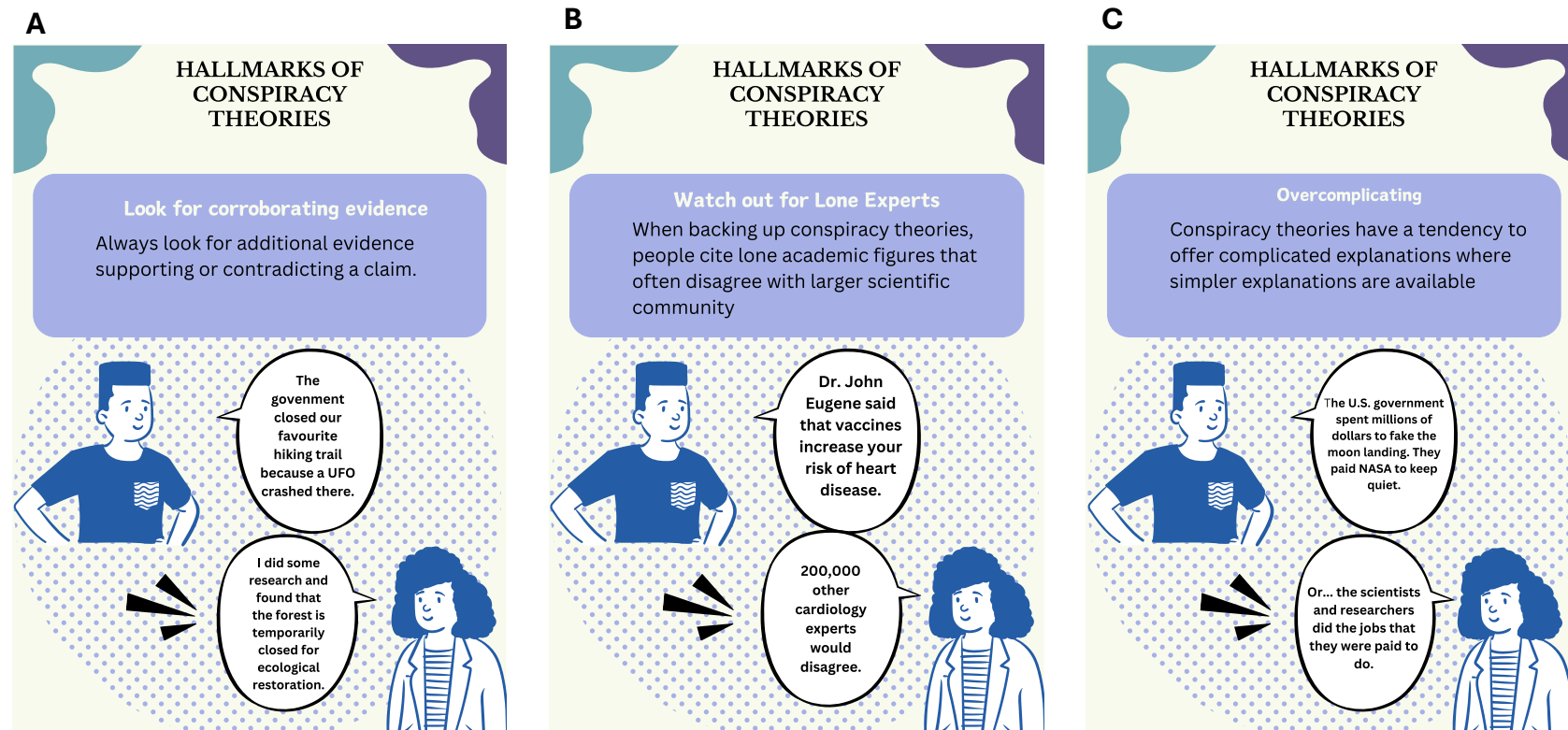
Conspiracy theories are often presented in a compelling manner and can be quite convincing, appealing to our natural fondness for stories. Conspiracy theories are not necessarily false, as real examples of conspiracies do exist. However, they can be misleading and lead people to doubt factual information or scientific evidence. While conspiracy theories are entertaining, and may contain some kernel of truth, most are not supported by logic. As such, they should not be taken at face value".

Inoculation. The design of the Inoculation intervention was informed by previous research, using both a threat and refutational component (Banas & Miller, 2013; Cook et al., 2017; van der Linden et al., 2017). We opted to use a broad-spectrum inoculation over one that debunks specific conspiracy theories, as we were interested in teaching participants about the underlying logical flaws that are common in conspiracy beliefs, rather than debunking specific conspiracy theories. Furthermore, our dependent variables cover a range of various conspiracy theory topics. Therefore, specific inoculations would not be suited for this specific study.

The Inoculation condition presented the warning paragraph from the Priming condition but was also followed by five infographics that outlined common mistakes that conspiracy theorists make (see Figure 1). These five infographics were based on the overarching themes found in previous informational inoculations along with logical fallacies typically associated with conspiracy theories. Topics in the inoculation related to (i) false experts (Banas & Miller, 2013; Cook et al., 2017), (ii) overcomplicating explanations and making unnecessary assumptions

Figure 1

Three of the five inoculation infographics used in the Inoculation and Active conditions, relating to (a) corroborating evidence, (b) lone experts, and (c) overcomplication. The two other infographics related to identifying the “who benefits” fallacy and logical leaps can be seen in our Supplementary Online Materials.



(Banas & Miller, 2013; Stall & Petrocelli, 2023), (iii) identifying logical leaps (Pytlík et al., 2020; Roozenbeek et al., 2022), (iv) the “who benefits” fallacy (Wagner-Egger, 2022), and (v) looking for corroborating evidence (Stall & Petrocelli, 2023). While many informational inoculations use passages of text as part of the refutational component (Banas & Miller, 2013), other studies have condensed these messages into infographic formats (Basol et al., 2021). We chose infographics as our medium to make the messages in the inoculations more accessible over dense paragraphs of text.

Active Inoculation. The Active condition was identical to the Inoculation condition but included a short quiz after the presentation of the infographics. We aimed to condense the active engagement and feedback mechanic used by Dyer and Hall (2019) by utilising a quiz to allow participants to practice applying the skills they had learned. The quiz allowed participants to apply the lessons from the infographics to practical scenarios, with the intention of reinforcing their learning and skill application (Bälter et al., 2013; Förster et al., 2018). Participants answered five multiple-choice questions. An example quiz item can be seen below:

“A new line of laptops has been reported to explode while charging. People online speculate that their competitors must have planted someone in the company to sabotage these laptops. What do you think is the main weakness of this theory?”

- a) This theory relies on the “who benefits” fallacy*
- b) It is unlikely that a rival company would do something so bold*
- c) Rival companies would have more subtle ways of beating their competition*
- d) This theory relies on vague terminology to justify its argument”*

The correct answer here is a), the conspiracy relies on the assumption that the company's competitors are responsible solely on the basis that they would stand to benefit from this incident. Participants received feedback that said, *“This conspiracy theory is flawed as it presumes just because the rival company would benefit from sabotaging the laptops, this automatically makes them guilty”*.

Critical Thinking About Conspiracies Assessment (CTAC). The CTAC is an eight-item assessment of critical appraisal of conspiracy theories (O'Mahony et al., 2024). The assessment consists of vignettes that describe hypothetical conspiracy theories. Participants are provided with four responses to choose from and are instructed to select the answer that they believe best reflects what they think is the most reasonable interpretation of this conspiracy. Participants are awarded one point for each correct item and receive no points for incorrect answers. Scores ranged from a possible zero to eight (min=1, max=8). The mean score was 5.96 ($SD = 1.68$). The scale showed low reliability, Cronbach's $\alpha = .54$. The implausible items had acceptable reliability ($\alpha = .64$) while plausible items had low reliability ($\alpha = .49$). An example item can be seen below:

Researchers in a prominent university have been investigating means of creating clean energy that will be five times more effective than current fossil fuel energy sources. Last week, a section of the university caught fire. Among the regions affected by the fire was the lab conducting research on clean energy. Your family is discussing the incident, when your dad suggests that the fire was an act of arson, committed on behalf of some major oil companies. He says that the first thing you should ask yourself in these situations is ‘who benefits?’”
What do you think of this reasoning?

- a) This reasoning is flawed. An action being useful to someone does not automatically imply they committed it.*

b) This reasoning is flawed. Oil companies wouldn't want to draw bad press on themselves.

c) This reasoning is valid. Accidents of this nature are rare. The first step in identifying the cause is to see who would benefit from this disaster.

d) This reasoning is valid. We know Big Oil can be ruthless in silencing opposition, so it is likely this is what happened here.

The correct answer is a) as the conspiracy is founded on a logical fallacy referred to as *cui bono* (or the “who benefits” fallacy). The logical fallacy suggests that oil companies are guilty purely on the basis that they would benefit from the arson attack. The item assesses participants' skills in argument analysis, by identifying flawed premises in arguments. The answer b) produces the correct conclusion, but it misses the primary problem with the conspiracy claim. Both c) and d) reflect poor reasoning, missing the most obvious problem with this claim. While answer b) may not necessarily be wrong, it misses the primary logical fallacy that is underlying this conspiracy theory. The purpose of the CTAC is that participants are awarded for identifying the *primary* problem with each scenario. This is similar to other measures of critical thinking (Lawson et al., 2015) and is particularly suitable for assessing beliefs about conspiracy theories, where there are often multiple, technically true interpretations of ambiguous information (Wood, 2017).

It should be noted that the CTAC is quite similar in content and style to the infographics and quiz used in the Inoculation and Active interventions, though, participants do not receive any instructional feedback as they complete the CTAC. This was an intentional decision, as the interventions were designed to teach the relevant critical thinking skills that are also assessed by the CTAC. Crucially though, the CTAC scenarios are drawn from a range of different topics and contexts, so participants must

generalise the skills learned in the intervention stage in order to apply them successfully to the novel contexts and scenarios included in the CTAC. In this way, we argue that any improvement in CTAC scores after completing these interventions would represent true learning that is also likely to transfer to real-life conspiracy theories.

Index of Epistemically Unwarranted Beliefs.

The IEUB is a 37-item Likert scale that measures the extent to which people endorse unfounded statements. Items on the scale cover a range of topics, including conspiracy theories, paranormal phenomena, and pseudoscientific claims and are rated on a scale of (1 = *I'm sure this is false*, to 5 = *I'm sure this is true*). Example items are “Houses can be inhabited by the spirits of those who died in unusual ways” and “The white streaks left in the sky behind airplanes contain chemicals released purposely on the population for nefarious purposes”.

We used an adapted version of the IEUB that consisted of 20 statements (see full materials on our OSF page). We omitted items from the original scale that referred to topics that are not well known outside of the USA, such as “A predatory animal lives unknown to biology in the southwestern US and Latin America and attacks livestock and drains their blood”. There were also five factual statements such as “The seasons of the Earth are caused by the tilt of the Earth on its axis as it orbits the Sun”. The five factual statements were excluded from the final total score but were used for signal detection analyses. Higher IEUB scores indicate higher unwarranted beliefs. An overall mean score for participants was calculated for the purpose of analysis with a range of one to five (min = 1.00, max = 4.20). The overall mean score was 2.26 (*SD* = 0.64). The scale showed strong reliability, Cronbach's α = .88.

Self-reported Enjoyment. Participants were asked to rate how much they enjoyed the survey. Due to the variable length of the different intervention groups, and various levels of

participant engagement these required, we sought to measure whether intervention group would influence perceived enjoyment of the survey. Participants were asked “How much did you enjoy taking this survey?” (1 = *Not at all*, 5 = *Extremely*). Mean enjoyment was 3.73 out of a possible total of 5.

2.1.4 Procedure

The study was advertised as an experiment about how “people use critical thinking to evaluate online information”. After providing informed consent and demographic information, participants were randomly assigned to one of the four conditions: Control, Priming, Inoculation, and Active. Afterwards, participants completed the CTAC, followed by the IEUB. Items for the CTAC and IEUB were presented in random order. The CTAC response options were also randomised. Participants then rated how much they enjoyed the survey and how likely they would be to recommend the survey to a friend (see Supplementary Materials, <https://osf.io/suw57/>). Finally, participants were debriefed, and the true purpose of the survey was revealed.

Participants received a number of attention checks throughout the study, depending on which group they were assigned to (for a full overview of attention checks used in this paper, see Supplementary Materials). We excluded participants who failed two or more attention checks. This is in line with Prolific's policy for rejecting poor-quality submissions and these participants were immediately replaced (though note that we did not specify this in our preregistration and these exclusions did not affect our sample size as the participants were replaced). As outlined in our preregistration, we also excluded any participants who failed one or more of the infographic attention checks in the Inoculation or Active conditions, as this suggested that they did not properly read the infographic material and, therefore, did not complete the intervention.

The percentage of failed attention checks per

condition were as follows: Control (1%), Priming (4%), Inoculation (12%), Active (11%). A one-way ANOVA found that there were significant differences between the total failed attention checks across the intervention groups ($F(3, 1046) = 1.38$, $p < .001$, $\eta^2 = 0.03$). Those in the Active condition failed significantly more attention checks than those in the Priming ($p = .015$) and Control ($p < .001$) conditions. Similarly, those in the Inoculation condition failed significantly more attention checks than those in the Priming ($p = .006$) and Control ($p < .001$) conditions. These results were expected, as participants in the Active and Inoculation conditions received more attention checks than the Priming and Control conditions. Furthermore, the difficulty of the Active and Inoculation checks was higher than those in the Priming and Control conditions. We accounted for the number of attention checks per condition by calculating the percentage of overall checks failed (rather than the total number) and found that deviations between intervention groups persisted ($F(3, 1046) = 7.79$, $p < .001$, $\eta^2 = 0.02$). These findings would suggest that the difficulty of the attention checks in the Active and Inoculation conditions may account for the differences in failure rates as opposed to the number of attention checks in each condition. Importantly, it should be noted that when comparing the attention check that was consistent between groups, we found that no participant failed attention checks.

To rule out any possible effects of attention checks on the nature of our samples across conditions, we report the preregistered analyses in the current paper (with attention check failures excluded) but also report analysis of the full sample (without these exclusions) in Supplementary Materials. While the overall pattern of results is broadly similar, there are some differences. These differences are elaborated on in the Supplementary Materials.

2.1.5 Ethics

The study procedure was approved by our institution's research ethics committee. In

accordance with the evidence-based debriefing practices outlined by Murphy and Greene (2023) and Greene et al. (2023), participants were fully debriefed at the end of the study and told that all of the conspiracy items used in the survey were fictional and created by the researchers, and that they should not take these into account when making decisions in the future.

2.2 Results

2.2.1 Did the Interventions Improve Critical Appraisal of Conspiracy Theories?

Reliability analysis revealed that the CTAC had an alpha level below the acceptable range for valid interpretation ($\alpha = .54$), therefore, we have omitted the results of the CTAC analysis from the manuscript and have instead reported them in more detail within Supplementary Materials while briefly summarizing them here. Our preregistered analysis suggested that there was no main effect of the intervention group on overall CTAC scores. We then conducted exploratory analyses where we separately assessed plausible and implausible CTAC items. Our analysis indicated the possibility that more engaging interventions (Inoculation and Active) *increased* critical appraisal of implausible conspiracy theories (where the correct answer is to reject the conspiracy) but potentially *reduced* critical appraisal of plausible conspiracy theories (where the correct answer is to endorse the conspiracy). However, the low reliability of the overall CTAC assessment and its subscales prevented any valid interpretation of these results (see Study 2 for the replication of these findings with a more reliable version of the CTAC).

2.2.2 Did the Interventions Decrease Epistemically Unwarranted Beliefs?

Per our preregistration, one participant was removed for scoring 2.5 standard deviations above the mean and one participant did not answer any IEUB items. A one-way ANOVA revealed a significant difference in IEUB scores between the intervention groups ($F(3, 1026) = 5.00$, $p = .002$, $\eta^2 = 0.01$). A Tukey's HSD test for

multiple comparisons revealed that participants held significantly fewer epistemically unwarranted beliefs in the Active condition relative to the Priming ($p = .025$, $d = -0.23$) and Control condition ($p = .008$, $d = -0.27$). Similarly, there were fewer unwarranted beliefs reported in the Inoculation condition than in the Control condition ($p = .026$, $d = -0.25$). There were no differences found between any of the other pairs.

As an exploratory robustness check, an ANCOVA was conducted to examine the effect of interventions on IEUB scores while controlling for age, gender, and nationality. The ANCOVA revealed a significant main effect of intervention type on IEUB scores after controlling for age, gender, and nationality, ($F(3, 986) = 4.25$, $p = .005$, $\eta^2 = 0.01$). The results of the multiple comparison analysis were broadly the same as the main ANOVA analysis. The full ANCOVA analysis can be found in Supplementary Materials.

2.3 Study 1a Discussion

Our results indicated that interventions had a significant effect on epistemically unwarranted beliefs, where participants who received more engaging interventions (Inoculation and Active) believed in fewer unwarranted beliefs than controls. Furthermore, our tentative analysis of the CTAC scores (reported in detail in the Supplementary Materials) suggested that the more engaging interventions seemed to improve critical appraisal of implausible conspiracy theories but might have also contributed to increased scepticism of plausible conspiracy theories. We considered whether explicitly teaching participants about the importance of discernment would improve their ability to appraise both implausible and plausible conspiracy theories. We gathered further data for a new condition (Discernment) to test this hypothesis.

3. STUDY 1B

Study 1b was preregistered (available here: https://aspredicted.org/QQG_FHY). We

collected an additional condition and ran analyses to compare performance in this novel condition to the data from the Study 1a conditions. In this analysis, we tested the same hypotheses as Study 1a, though here we preregistered that we would separately assess the Plausible and Implausible CTAC subscales.

3.1 Methods

3.1.1 Participants

The analysis for Study 1b utilised the data collected in Study 1a for the Control, Priming, Inoculation, and Active Inoculation interventions. We used Prolific's screening function to exclude participants who participated in Study 1a from taking the survey. We recruited an additional 201 and after removing four participants for failing attention checks we had a total of 187 participants (126 women, $M_{age} = 38.05$, $SD_{age} = 13.02$) for the added Discernment condition. The majority of participants were from the United Kingdom (60%), and the remaining participants were from Canada (24%), New Zealand (5%), the United States (5%), Australia (4%), and Ireland (2%).

3.1.2 Materials

Critical Thinking About Conspiracies (CTAC)

Assessment. The CTAC was once again used to measure critical appraisal of novel conspiracy theories. The internal reliability was again low for the overall assessment ($\alpha = .55$) and for the Plausible ($\alpha = .50$) subscale. The Implausible subscale had acceptable reliability ($\alpha = .63$).

Index of Epistemically Unwarranted Beliefs.

The IEUB was once again used to measure unfounded beliefs. The internal reliability was excellent ($\alpha = .85$).

3.1.3 Procedure

Participants completed an intervention identical to the Active condition in Study 1a. However, after the practice quiz and feedback, they were asked to answer one further question. In this new item participants were presented with a practice example where the correct conclusion was to suggest that the conspiracy was

plausible. The new item can be seen below:

A Senator has been accused of accepting bribes from the CEO of a prominent pharmaceutical company. He has denied this claim, stating "I have never met them in my life". Yet, investigative journalists published photos of him meeting with that very CEO 3 months ago. Is this suspicious?

- a) *Yes, the photos serve as corroborating evidence that he is lying about his association with the CEO.*
- b) *No. This conspiracy theory relies on the "who benefits" fallacy.*
- c) *No. This conspiracy theory is very unlikely.*
- d) *Yes. Politicians should never be trusted*

The correct answer is a). The photographs serve as corroborating evidence, which is a skill that participants were taught in the infographic material. Afterwards, whether they answered the item correctly or not, participants were told that this specific item was actually plausible and warned about blindly dismissing conspiracy theories without using reasoning:

"A common side-effect of learning to think critically about conspiracy theories is that it can make people overly sceptical of conspiracies, even ones that could possibly happen. It is important to understand that while many conspiracy theories may be unlikely, some do occur, so it is important to always consider the available information and not automatically dismiss claims".

3.1.4 Attention Checks

For the added Discernment condition, participants received one infographic comprehension check, in contrast to the Inoculation and Active conditions who received two infographic comprehension checks. Participants in the Discernment condition were instead asked to clarify the takeaway of the Discernment Message they

were shown at the end of the quiz.

A total of 7% of the Discernment group failed one attention check. A one-way ANOVA found that there were no differences between the number of attention checks failed within the Discernment condition, and the Active ($p = .413$), Inoculation ($p = .268$), Priming ($p = .824$), and Control ($p = .075$) condition

3.2 Results

3.2.1 Did the Discernment Intervention Improve Critical Appraisal of Conspiracy Theories?

As was the case with the results from Study 1a, the reliability of the CTAC assessment was below the acceptable threshold for valid interpretation ($\alpha = .55$). Therefore, a general narrative summary of the analysis has been provided in place of the results, and more detailed analyses can be found in Supplementary Materials. In brief, those in the Discernment condition appeared to report better critical appraisal of implausible conspiracy theories relative to the Control group. Additionally, those in the Discernment condition appeared to perform better than other intervention groups at critically appraising plausible conspiracy theories. The effects of the Discernment condition were in contrast to the Inoculation and Active interventions, which improved critical appraisal of conspiracy theories but increased scepticism of plausible conspiracy theories. See Study 2 for a replication of these analyses with a more reliable scale.

3.2.2 Did the Discernment Intervention Decrease Epistemically Unwarranted Beliefs?

Per our preregistration, one participant was removed for scoring 2.5 standard deviations above the mean and one participant did not answer any IEUB items. As seen in Figure 2, a one-way ANOVA revealed that there was a main effect of intervention group on IEUB scores ($F(4, 1212) = 3.89, p = .004, \eta^2 < 0.01$). However, multiple comparisons revealed that this main effect was not driven by the Discernment

condition, as there were no significant differences between the Discernment condition and the Active ($p = .924$), Inoculation ($p = .975$), Priming ($p = .486$) and Control ($p = .268$) conditions.

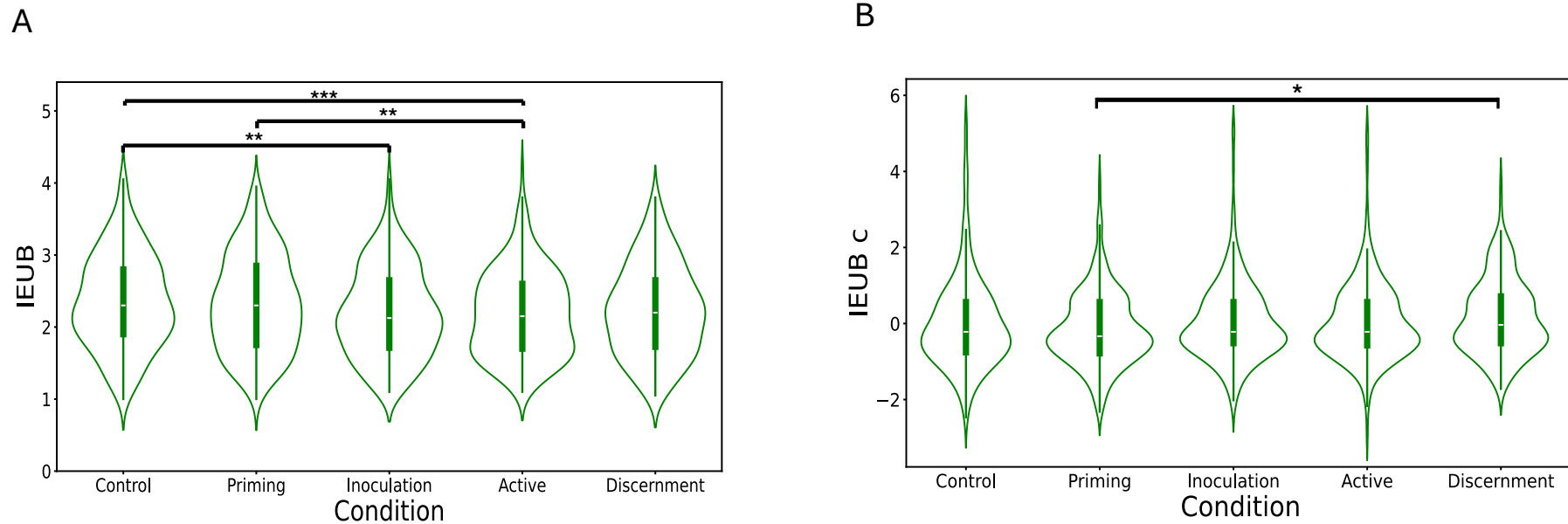
As a robustness check, an exploratory ANCOVA was also conducted to examine the effect of interventions on IEUB scores while controlling for age, gender, and nationality (variables that might have differed between Studies 1a and b). The addition of these covariates did not change the results, finding a significant effect of intervention type on IEUB scores, ($F(4, 1171) = 3.21, p = .012, \eta^2 = 0.01$). The full ANCOVA is reported in Supplementary Materials.

Importantly, to examine any response bias that the interventions may have encouraged, we drew on signal detection theory to calculate c , which allowed us to calculate the threshold for veracity that participants required before judging a statement as factual. Positive values indicate that a participant is less likely to judge statements as factual regardless of their veracity, and negative values indicate that a participant is more likely to judge statements as factual regardless of their veracity. We calculated c by subtracting the z -score of participants' erroneous beliefs in unwarranted statements from the z -score of their accurate beliefs in factual statements, dividing the difference by two, and then multiplying the result by a negative one.

A one-way ANOVA of c scores indicated a significant effect of intervention group ($F(4, 1211) = 2.67, p = .031, \eta^2 < 0.01$). A post-hoc Tukey's test revealed that those in the Discernment group were significantly more likely to have a bias towards dismissing all statements in the IEUB as false in comparison to the Priming group ($p = .013, d = .14$), which shows a bias towards accepting all statements in the IEUB as factual. Importantly, neither of the c scores in either the Discernment group ($p = .495$) or the Priming group ($p = .431$) were significantly different from the Control condition. There were no differences between the Control condition and the

Figure 2

Violin plots showing the kernel densities of A) epistemically unwarranted beliefs, and B) *c* scores for epistemically unwarranted beliefs to estimate response bias



Note: * $p < .05$, ** $p < .01$, *** $p < .001$; Lines denote significant differences.



Active ($p = 1.00$) or Inoculation ($p = 1.00$) conditions. Furthermore, an exploratory ANCOVA revealed that the effects of intervention groups on IEUB c scores were not significant when controlling for nationality, age, and gender ($F(4, 1170) = 2.28, p = .059, \eta^2 < 0.01$).

3.3 Study 1b Discussion

While the reliability of the CTAC was still below acceptable thresholds for valid interpretation, our analysis tentatively suggested that encouraging discernment could improve participants' overall ability to critically appraise conspiracy theories. Participants in the Discernment condition appeared to more accurately appraise conspiracy theories than controls while still indicating improved appraisal of implausible conspiracy theories as seen in Active and Inoculation conditions. These results were promising as they suggested the possibility that the Discernment intervention could improve appraisal of implausible conspiracy theories in a manner similar to the Active and Inoculation interventions while preventing encouragement of blind scepticism. Furthermore, signal detection analysis conducted on the IEUB to test for response bias suggested that the Discernment condition was not merely encouraging a bias to accept or reject conspiracies but possibly teaching participants to distinguish between plausible and implausible stimuli. However, the CTAC did not meet acceptable thresholds for valid interpretation of the findings. We therefore opted to see whether these results could be replicated using a more robust and valid version of the CTAC in Study 2.

4. STUDY 2

We conducted a second study using a 15-item version of the CTAC, which had been previously shown to have higher reliability ($\alpha = 0.75$) than its eight-item counterpart (O'Mahony et al., 2024) to replicate the findings of Study 1. We included two additional measures of conspiracy thinking to assess the effectiveness of the interventions more broadly. The Generic

Conspiracist Belief Scale (Brotherton et al., 2013) was used to measure participants' general tendency to entertain conspiracy beliefs. The Judgments of (Im)Plausible Conspiracy Theories (Frenken et al., 2024) was used to measure how participants assessed the likelihood of plausible and implausible conspiracy theories but focused primarily on likelihood judgments in contrast to critical appraisal as measured by the CTAC. This study was not preregistered as it was a replication of Study 1.

4.1 Methods

4.1.1 Participants

Power analysis using GPower version 3.1.9.7 indicated that 540 participants were required to detect a Cohen's f effect size of 0.15 with 80% power, with a one-way ANOVA. We expected a larger effect size based on the findings of Study 1. We recruited 554 participants from the participant recruitment website Prolific (353 women, $M_{age} = 41.26, SD_{age} = 13.31$). Most participants were from the United Kingdom (67%), and the remaining participants were from Canada (17%), the United States (6%), Ireland (5%), Australia (4%), and New Zealand (1%). Seven participants failed two attention checks and were removed from the sample, leaving 547 participants in the final analysis.

4.1.2 Materials

Critical Thinking About Conspiracies Assessment (CTAC) 15-item Long-Form. The 15-item long-form version of the CTAC was used in this study (O'Mahony et al., 2024; see Supplementary Materials). Like the eight-item CTAC used in Study 1, this version presents participants with 15 vignettes describing hypothetical scenarios (seven plausible and eight implausible). The mean score was 10.91 ($SD = 2.96$) out of a possible 15. The scale showed good reliability in the current study (Cronbach's $\alpha = .76$). The implausible subscale showed good reliability ($\alpha = .77$), and the plausible subscale showed acceptable reliability ($\alpha = .62$).

Generic Conspiracist Belief Scale (GCBS). The

Generic Conspiracist Beliefs scale (Brotherton et al., 2013) is a 15-item Likert scale that presents various statements about conspiracy theories (e.g., “A small, secret group of people is responsible for making all major world decisions, such as going to war”). We included the GCBS to assess how the interventions might affect conspiracy ideation, which was not measured in our previous study. Scores for each item range from one to five (1 = *Definitely not true*, 5 = *Definitely true*). Higher scores are indicative of higher conspiracy belief. Mean scores were calculated for each participant, with scores ranging from a possible 1 to 5 (min = 1.00, max = 4.67). The mean score was 2.43 ($SD = 0.80$). The internal reliability was excellent ($\alpha = .93$).

Conspiracy Judgements. The Judgments of (Im)Plausible Conspiracy Theories (JICT; Frenken et al., 2024) assessment was used as an additional measure of discernment between plausible and implausible conspiracy theories. The overall assessment consists of six items. The Plausible ($\alpha = .75$) and Implausible ($\alpha = .71$) subscales consist of three items each and the overall reliability was acceptable ($\alpha = .78$). Participants were told to imagine a fictional country named Lumaria, described as an industrialised and democratic state with diverse institutions. Participants then rated the likelihood of six theories being true on a Likert scale (1 = *Very Unlikely*, 7 = *Very likely*). An implausible example is: “*The Lumarian government, facing declining support, staged a false flag terrorist attack on a crowded train station with the help of its intelligence service. As the election approached, tensions rose and the government used fear to rally the population against a perceived terrorist threat in order to maintain its grip on power*”. A plausible example is “*The powerful technology company ‘TechWave Enterprises’ uses advanced surveillance technology to spy on users and manipulate their behavior. This includes collecting and analyzing user data without their knowledge or consent, using that data to create targeted advertising, or even influencing political or*

social outcomes through targeted messaging”. Mean scores were calculated for each participant, ranging from 1 to 7 (min = 2.67, max = 6.67). The overall mean score was 4.75 ($SD = 0.73$). The Implausible and Plausible subscales were found to negatively correlate with one another ($r = -.45$, $p < .001$). Moreover, the JICT was only reliable when higher scores indicated higher ratings of plausibility on both subscales – in contrast to reverse scoring the Implausible scale so that higher scores indicate more correct answers. This would suggest that participants generally either endorsed or rejected all conspiracy theories regardless of their plausibility.

4.1.3 Procedure

Participants completed a study that was almost identical to that used in Study 1. After being randomly assigned to one of the 5 intervention groups (Control, Priming, Inoculation, Active, Discernment), participants completed the CTAC, followed by the GCBS and the JICT. Items for the CTAC, GCBS, and the JICT assessment were presented in random order. The CTAC responses were also randomised.

4.1.4 Attention Checks

The attention checks used in Study 2 were the same as those used in Study 1. However, participants were only excluded from the analysis if they failed two or more of any attention checks. Participants who failed two or more were replaced through Prolific and thus were not included in the dataset. The percentage of failed attention checks per condition were: Control (1%), Priming (2%), Inoculation (7%), Active (4%), Discernment (8%). A one-way ANOVA found that there were significant differences between the total failed attention checks across the intervention groups ($F(4, 542) = 2.92$, $p = .021$, $\eta^2 = 0.02$). However, multiple comparisons revealed that there were no significant differences between any of the comparisons, with only the Discernment condition reporting marginally higher attention check failures than the Control condition ($p = .053$). We accounted for the

number of attention checks per condition and calculated the percentage of overall checks failed and found that there were no significant differences between groups ($F(4, 542) = 2.09, p = .081, \eta^2 = 0.02$). These findings suggest that the different number of attention checks were responsible for this slight deviation between groups.

4.2 Results

As shown in Table 1, there were small to moderate correlations found between the different outcome measures used in this study. Higher critical appraisal of conspiracy theories (CTAC) was associated with lower conspiracy ideation (GCBS) and more accurate likelihood judgments of plausible and implausible conspiracy theories (JICT). More accurate likelihood judgments (JICT) were also associated with lower levels of general conspiracy ideation (GCBS).

A total of ten participants were excluded from the CTAC analysis for scoring 2.5 standard deviations above or below the mean. There was a significant difference in overall CTAC scores among the interventions as assessed by a one-way ANOVA ($F(4, 532) = 5.05, p < .001, \eta^2 = 0.04$). As can be seen in Figure 3, a Tukey's *HSD* Test for multiple comparisons found that participants were significantly more accurate at critically appraising conspiracy theories in the Discernment condition than those in the Control ($p = .004, d = 0.48$) or Priming ($p = .001, d = 0.56$) conditions. Other comparisons were not significant.

We then compared the intervention groups on

their ability to critically appraise implausible and plausible conspiracy theories. For the Implausible subscale (that is, where participants *should* reject the conspiracy), a one-way ANOVA indicated a significant effect of condition ($F(4, 532) = 13.93, p < .001, \eta^2 = 0.09$). A post-hoc Tukey's *HSD* Test for multiple comparisons found that participants were significantly more accurate at critically appraising implausible conspiracy theories in the Discernment condition than in the Control ($p < .001, d = 0.61$) and Priming conditions ($p < .001, d = 0.72$). However, those in the Discernment condition were not significantly more accurate than those in the Active ($p = 1.00$) and Inoculation conditions ($p = .995$). Those in the Active condition were also significantly more accurate than those in the Control ($p < .001, d = 0.60$) and Priming ($p < .001, d = 0.71$) conditions. Similarly, those in the Inoculation condition were significantly more accurate than those in the Control ($p < .001, d = 0.54$) and Priming conditions ($p < .001, d = 0.65$). There was no significant difference between the Active and Inoculation ($p = .991$) conditions, nor between the Priming and Control conditions ($p = .943$).

For the Plausible CTAC subscale (where the correct answer is to *endorse* the possibility of a conspiracy), there was a main effect of intervention group ($F(4, 532) = 3.75, p = .005, \eta^2 = 0.03$). A post-hoc Tukey's test showed that those in the Discernment condition were significantly more accurate at appraising plausible conspiracy theories than those in the Active condition ($p = .005, d = 0.47$). Additionally, those in the Priming condition were significantly more accurate

Table 1
Correlations for Dependent Variables

Measures	1	2	3	4	5	6
1. CTAC						
2. CTAC Implausible	0.83**					
3. CTAC Plausible	0.74**	0.24**				
4. JICT	0.34**	0.30**	0.23**			
5. JICT Implausible	0.35**	0.49**	0.03	0.52**		
6. JICT Plausible	-0.001	-0.18**	0.22**	0.52**	-0.46**	
7. GCBS	-0.42**	-0.57**	-0.05	-0.27**	-0.72**	0.44**

Note. ** $p < .001$

than those in the Active condition ($p = .038$, $d = 0.39$). No other comparisons were statistically significant.

We measured the discernment rates of participants in terms of their ability to critically distinguish between implausible and plausible conspiracy theories. We computed a d -prime (d') as a measure of critical sensitivity to the difference between implausible and plausible conspiracy theories. This was calculated using z scores of accurately appraising implausible conspiracy theories ("hits") and subtracting the z scores of participants erroneously dismissing plausible conspiracy theories ("false alarms"). This is based on the premise of participants correctly appraising implausible conspiracy theories while not overshooting and erroneously dismissing plausible conspiracy theories. A one-way ANOVA indicated a significant effect of intervention group ($F(4, 532) = 4.32$, $p < .002$, $\eta^2 = 0.03$). A post-hoc Tukey's test revealed that those in the Discernment condition demonstrated better discernment than those in the Priming ($p = .003$, $d = 0.52$) and Control ($p = .009$, $d = 0.45$) conditions. There were no significant differences between any other comparisons.

As a robustness check, four ANCOVAs were conducted to examine the effect of the interventions on each CTAC outcome while controlling for age, gender, and nationality. All results were broadly the same as the ANOVAs and are reported in full in Supplementary Materials. In summary, while controlling for these covariates, we still found a significant effect of intervention type on overall CTAC scores ($F(4, 525) = 5.51$, $p < .001$, $\eta^2 = 0.04$), on Plausible CTAC scores ($F(4, 525) = 3.54$, $p = .007$, $\eta^2 = 0.03$), on Implausible CTAC scores ($F(4, 525) = 14.71$, $p < .001$, $\eta^2 = 0.10$), and on CTAC d' scores ($F(4, 525) = 4.64$, $p = .001$, $\eta^2 = 0.03$). Multiple comparisons revealed that results were broadly the same as the main analysis.

4.2.1 Did the Interventions Decrease General Conspiracy Ideation?

There was a significant effect of

the intervention group on GCBS scores ($F(4, 542) = 5.05$, $p < .001$, $\eta^2 = 0.04$). As seen in Figure 4, participants in the Active condition reported significantly lower conspiracy ideation than those in the Priming condition ($p = .001$, $d = -0.53$). Similarly, those in the Inoculation condition reported significantly lower conspiracy ideation than those in the Priming condition ($p < .001$, $d = -0.52$). There were no significant differences amongst any of the other comparisons.

An ANCOVA was conducted to examine the effect of interventions on GCBS scores while controlling for age, gender, and nationality. The ANCOVA revealed a significant effect of intervention type, $F(4, 535) = 5.47$, $p < .001$, $\eta^2 = 0.04$. Results were the same as the main ANOVA analysis.

4.2.2 Did the Interventions Improve Likelihood Ratings of Plausible and Implausible Conspiracy Theories?

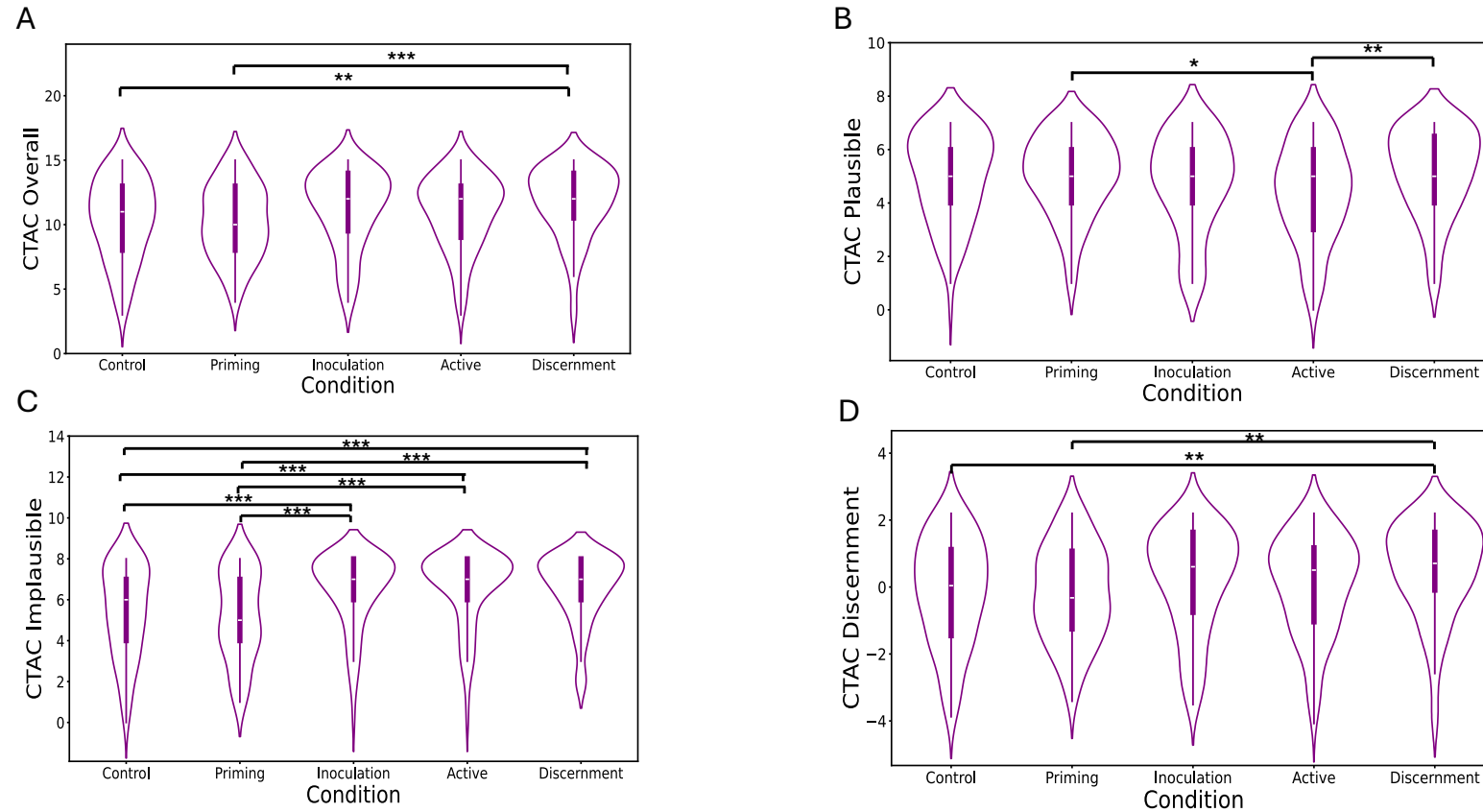
There was no significant effect of intervention group on the JICT ($F(4, 542) = 1.26$, $p = .287$, $\eta^2 = 0.01$). There were no significant differences between the Discernment ($M = 4.66$, $SD = .76$), Active ($M = 4.82$, $SD = 0.75$), Inoculation ($M = 4.71$, $SD = 0.72$), Priming ($M = 4.70$, $SD = 0.61$) and Control ($M = 4.83$, $SD = 0.78$) conditions.

We then separately assessed intervention effects on appraisal of the scale's implausible and plausible conspiracy theories subscales. There was a significant main effect of intervention group on ratings of implausible conspiracy theories ($F(4, 542) = 3.00$, $p = .018$, $\eta^2 = 0.02$). Participants in the Active condition were significantly more inclined to correctly rate implausible conspiracy theories as being unlikely than those in the Priming condition ($p = .022$, $d = 0.43$). Similarly, participants in the Inoculation condition were significantly more inclined to correctly rate implausible conspiracy theories as being unlikely than those in the Priming condition ($p = .040$, $d = 0.38$).

There was also a main effect of intervention group on appraisal of plausible conspiracies

Figure 3

Violin plots showing the Study 2 kernel density distribution of A) overall critical appraisal of conspiracy theories (CTAC Total), B) critical appraisal of plausible conspiracy theories (CTAC Plausible), C) critical appraisal of implausible conspiracy theories (CTAC Implausible), and D) critical discernment between plausible and implausible conspiracy theories (CTAC d').

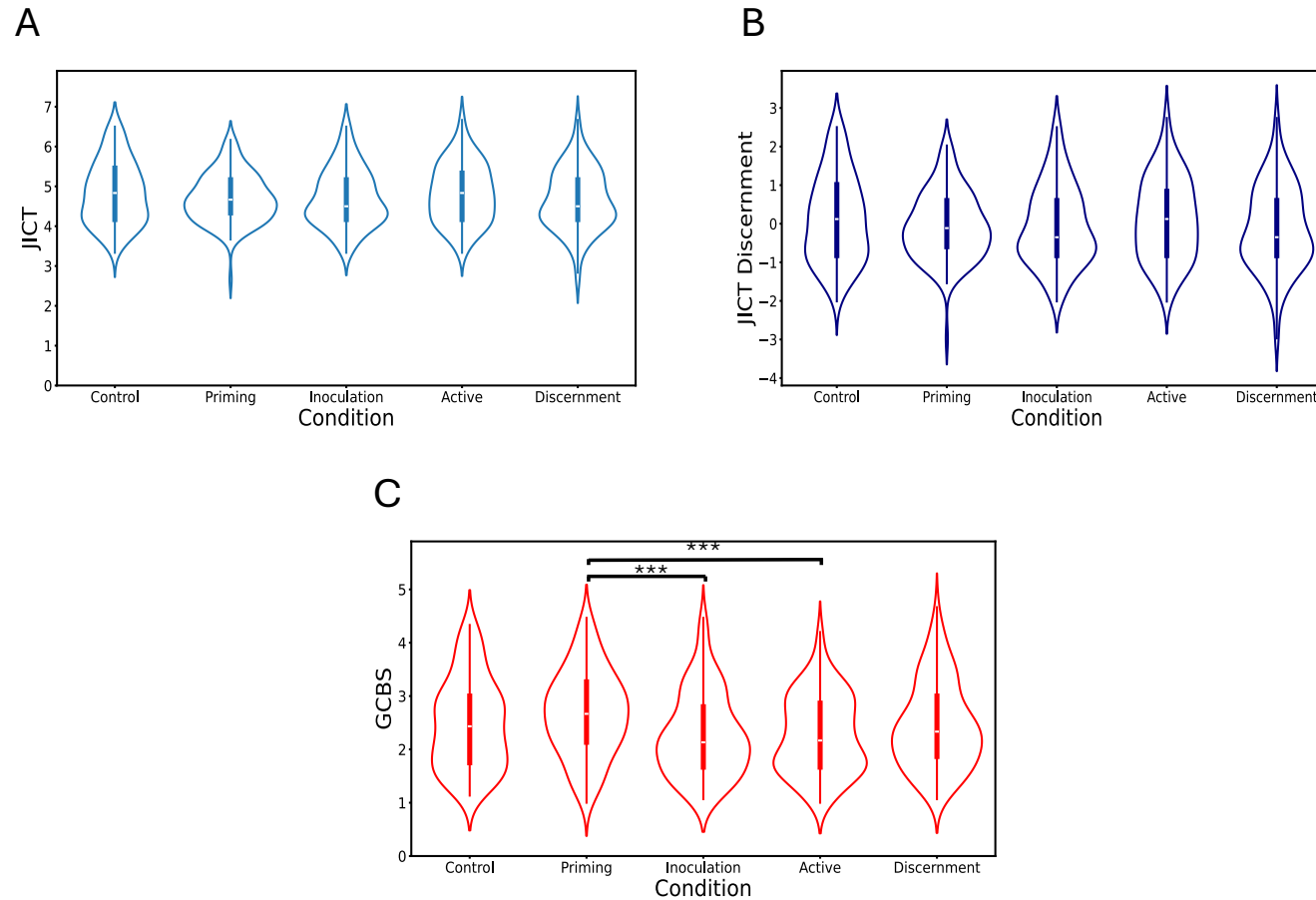


Note: * $p < .05$, ** $p < .01$, *** $p < .001$; Lines denote significant differences.



Figure 4

Violin plots showing the kernel density distribution of A) conspiracy judgements of the likelihood of plausible and implausible conspiracy theories (JICT), B) discernment between plausible and implausible conspiracy theories (JICT d'), and C) general conspiracy ideation (GCBS).



Note: * $p < .05$, ** $p < .01$, *** $p < .001$; Lines denote significant differences.



($F(4, 542) = 3.00, p = .018, \eta^2 = 0.02$). Those in the Inoculation condition were significantly less inclined to rate plausible conspiracy theories as being likely than those in the Priming condition ($p = .049, d = -0.37$) and also marginally less likely than those in the Control condition to rate plausible conspiracy theories as being likely ($p = .053, d = -0.38$). Overall, participants in the Inoculation condition were more blindly sceptical of plausible conspiracy theories, similar to the effects of the interventions on the CTAC. No other comparisons were significant.

We calculated d' for the JICT to calculate the extent to which participants discerned between the likelihood of plausible and implausible conspiracy theories. A one-way ANOVA found no significant differences between d' scores on the JICT among any intervention group ($F(4, 532) = 1.59, p = .175, \eta^2 = 0.01$).

As a robustness check, four ANCOVAs were conducted to examine whether our findings still held if we controlled for age, gender and nationality. The results remained the same, with no effect of condition on overall JICT scores ($F(4, 535) = 1.41, p = .230, \eta^2 = 0.01$), a significant effect on Implausible JICT scores ($F(4, 535) = 3.00, p = .018, \eta^2 = 0.02$), a significant effect on Plausible JICT scores, ($F(4, 535) = 2.84, p = .024, \eta^2 = 0.02$), and no significant effect on JICT d' scores ($F(4, 535) = 1.41, p = .230, \eta^2 = 0.01$). All analyses are reported in full in Supplementary Materials.

5. GENERAL DISCUSSION

Conspiracy theory research - especially intervention studies - have suffered from ambiguities in both definitions and measurements. Based on the findings from the current study, we echo previous calls for teaching critical thinking skills as the most appropriate intervention type. We argue for more accurate distinctions to be made between plausible and implausible conspiracy theories in how we define and discuss conspiracy beliefs (Duetz, 2023), how we measure them empirically

(O'Mahony et al., 2024), and how we design interventions (Guay et al., 2022). To the best of our knowledge, this paper is the first to examine the relative effectiveness of several intervention types designed to increase critical appraisal of conspiracy theories, focusing on the processes that underpin conspiracy beliefs rather than just assessing the outcome. Overall, we consider our findings positive and encouraging, as we detected significant (though moderate) effects of very brief interventions (~5 minutes in length) on participants' ability to critically evaluate novel conspiracy theories. This highlights the potential malleability of these sometimes unreasonable and problematic beliefs and encourages further research in this area.

Our results indicate that the interventions varied in effectiveness. Priming interventions were found to be no more effective than controls at improving critical appraisal of conspiracy theories or reducing conspiracy and unfounded beliefs. Indeed, priming interventions were found to have no effect on any outcome measures when compared to controls. This is consistent with recent literature indicating that priming interventions are not effective (Epstein et al., 2021; Roozenbeek et al., 2021; Večkalov et al., 2024). Importantly, it should be noted that the priming intervention used in the current study was modelled on real-life misinformation campaigns that have been run by government agencies, such as the "Don't Feed the Beast" campaign (U.K. Government, 2020). Our results suggest these interventions are ineffective at reducing unreasonable conspiracy beliefs. While we had expected that the active inoculation would be more effective than the passive inoculation, this was not the case, suggesting that the addition of a quiz where participants applied the inoculation lessons did not improve learning outcomes. This findings contrasts with previous research indicating an improvement on learning outcomes when interactive quizzes are used in the learning process (Bälter et al., 2013; Förster et al., 2018). It may have been a result of the brief nature of our intervention, or

the lack of a delay between the treatment and assessment (Banas & Rains, 2010; McGuire, 1964). In particular, early inoculation research found that the effects of active inoculations were only discernible after a two-day delay, in comparison to passive inoculations, which were much more instantaneous (Rogers & Thistlethwaite, 1969). It is possible that active inoculations may have a sleeper effect, and due to the short time interval of our study, the booster effects of the active inoculation were not detected.

Our analysis produced the novel finding that many of the interventions that are well-established in the research literature have either no effect or a negative effect on participants' ability to correctly reason about *plausible* conspiracy theories. Our findings suggest that many of interventions designed to tackle unfounded conspiracy theories are effective in reducing susceptibility to implausible conspiracy theories but have no statistically significant effect in terms of increasing critical appraisal of plausible conspiracy theories. These conclusions are consistent with findings from general misinformation literature that indicates that interventions aimed at reducing susceptibility to fake news often makes participants overly sceptical of factual news headlines as well (Hoes et al., 2023; van der Meer et al., 2023). Interventions that increase blind scepticism may do more harm than good, and this raises important questions regarding how we should intervene in conspiracy beliefs, but also how we measure them (Guay et al., 2022; Hameleers, 2023). Our findings suggest that previously published interventions (which assess belief in exclusively implausible conspiracy theories or general conspiracy ideation), were perhaps increasing participants' tendency to dismiss all conspiracy theories regardless of their plausibility (Banas & Miller, 2013; Jolley & Douglas, 2017). To date, this has been invisible to researchers due to the nature of the measures employed (O'Mahony et al., 2023). We encourage researchers to consider discernment as an important outcome

measure in future conspiracy belief studies.

We also tested a novel discernment condition, which was unlike any intervention found in a prior systematic review of the topic, (O'Mahony et al., 2023) and found it increased overall discernment between plausible and implausible conspiracy theories. The Discernment condition showed success across three different measures in comparison to control groups (CTAC overall, CTAC Implausible, CTAC d') with Active and Inoculations having significant effects on two measures (IEUB, CTAC Implausible). Relative to the more standard inoculation interventions, teaching discernment improved appraisal of plausible conspiracy theories while having no detrimental effect on appraisal of implausible conspiracy theories, improving overall discernment. Indeed, it should be noted that those in the discernment condition were no more accurate at appraising plausible conspiracies than those in the control group. This suggests that teaching discernment does not undo the healthy scepticism conferred by the inoculation conditions but rather has an additive effect to prevent the blindly sceptical "over-correction" that such interventions may cause. Signal detection analysis using the CTAC d' and IEUB c scores also suggested that the discernment condition was not merely encouraging a response bias to dismiss or accept conspiracy theories but was improving participants' ability to critically distinguish between plausible and implausible conspiracy theories. Importantly, our results indicated that enjoyment did not differ between the intervention groups despite differences in length and engagement required from the participant. While participants only rated moderate enjoyment in all conditions, these findings indicate that the variability in length and complexity of these interventions had no negative effect on participants' enjoyment of these interventions. Therefore, more engaging interventions may not run the risk of decreasing participants' overall enjoyment of these interventions (Galesic & Bosnjak, 2009).

Our analysis also examined the effect of the interventions on other measures of conspiracy thinking and unfounded beliefs. The results indicated that both passive and active interventions were successful in reducing epistemically unwarranted beliefs, which is consistent with previous research that utilised much more intensive intervention methods (Dyer & Hall, 2019). These findings are promising, suggesting that shorter, scalable interventions may be capable of challenging unwarranted beliefs. Furthermore, these findings may also suggest that broad interventions that target conspiracy beliefs may translate to tackling adjacent topics such as pseudoscience, which is consistent with previous work (Cook et al., 2017). However, our brief interventions did not have significant effects on a measure of generic conspiracy ideation (i.e., the GCBS), which is inconsistent with previous research (Mason et al., 2024).

Our interventions did not consistently improve ratings of conspiracy plausibility, though to our knowledge, this was the first study to use the JICT as an intervention outcome measure (Frenken et al., 2024). This is perhaps not so surprising, as the conspiracies described in the JICT items are classed as correct or incorrect based on the subject matter (e.g., Big Tech conspiracies were plausible, whereas secret religious organisations were implausible). This is a different approach to the CTAC, where plausibility was based entirely on the information available to the participant. Interventions improving participants' scores on the CTAC but not the JICT may suggest that these interventions improved participants' ability to change how they critically evaluated the evidence supporting or rejecting a novel conspiracy theory, but did not ultimately alter their estimations of likelihood judgements of specific novel conspiracy theories. Alternatively, the observed results may be due to the nature of the interventions used. The discernment intervention in particular was quite similar in format and content to the CTAC itself. Therefore, it may be that the skills acquired in the discernment

intervention translated to the CTAC, but not to the JICT. We recommend further study in this regard to disentangle the effectiveness of various kinds of discernment-based interventions using a wide variety of outcome measures.

Overall, we tentatively suggest that the different effects of the interventions on the various conspiracy thinking measures may be explained by the fact that these measures assess related but distinct concepts. We hypothesise that these interventions were effective at improving participants' ability to think rationally about novel conspiracy theories (measured via the CTAC) and evaluate general misinformation (IEUB) but did not effectively sway their underlying propensity to think in a conspiratorial manner (GCBS), or their attitudes towards the plausibility of conspiracy theories (JICT). We hope our research might encourage further studies on the various mechanisms that underpin these potentially distinct constructs in the hope of finding interventions that effectively target unreasonable conspiracy beliefs.

One important limitation of this study is that we used a cross-sectional design and did not measure any longitudinal effects, a limitation common to conspiracy intervention studies (O'Mahony et al., 2023). The second limitation of our study is that we primarily sampled Western countries, with the majority of our sample residing in the United Kingdom. Recent research has suggested that there is a lack of diverse samples in conspiracy and misinformation research (O'Mahony et al., 2023; Murphy et al., 2023). However, this was a deliberate trade-off, as the CTAC assessment has only been validated in English, and due to the length of the items in assessment, we felt that native English-speaking skills were a requirement for this research. Future studies should examine the generalisability of the results observed here, both across cultures and over a longer period. Finally, we only tested broad-spectrum inoculations in this study. Therefore, the efficacy of other inoculations, such as topic-specific

inoculations, need to be evaluated (Banas & Miller, 2013; Jolley & Douglas, 2017).

In summary, while our study found that established active and passive inoculations are effective at increasing critical appraisal of implausible conspiracy theories and reducing epistemically unwarranted beliefs, it also highlights the threat of these inoculations promoting blind scepticism. We propose a novel discernment intervention that is effective in promoting critical appraisal of implausible conspiracy theories while having a protective effect that mitigates the negative effects of typical interventions on evaluations of plausible conspiracy theories. We recommend that future studies focus on both measuring discernment of conspiracy theories as an outcome measure and designing interventions that encourage discernment over blind scepticism.

6. CONFLICTS OF INTEREST

The authors declare no competing interests.

7. AUTHOR CONTRIBUTIONS

C.OM., G.M., and C.L. designed the study. C.OM., collected and analysed the data. C.OM., G.M., and C.L., drafted the manuscript. C.OM. created the visuals and tables. All authors provided critical revision.

8. ACKNOWLEDGEMENTS

This research received funding as part of COM Grant Number: EPSPG/2021/212 from the Irish Research Council. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

9. REPRODUCIBILITY STATEMENT

All materials and data for the current project are available at the Open-Science Framework (<https://osf.io/suw57/>).

REFERENCES

- Bälter, O., Enström, E., & Klingenberg, B. (2013). The effect of short formative diagnostic web quizzes with minimal feedback. *Computers & Education*, 60(1), 234–242.
<https://doi.org/10.1016/j.compedu.2012.08.014>
- Banas, J. A., & Miller, G. (2013). Inducing Resistance to Conspiracy Theory Propaganda: Testing Inoculation and Metainoculation Strategies: Metainoculation Strategies. *Human Communication Research*, 39(2), 184–207.
<https://doi.org/10.1111/hcre.12000>
- 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., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. V. D. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society*, 8(1), 205395172110138.
<https://doi.org/10.1177/20539517211013868>
- Bonetto, E., Troian, J., Varet, F., Lo Monaco, G., & Girandola, F. (2018). Priming Resistance to Persuasion decreases adherence to Conspiracy Theories. *Social Influence*, 13(3), 125–136.
<https://doi.org/10.1080/15534510.2018.1471415>
- Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring Belief in Conspiracy Theories: The Generic Conspiracist Beliefs Scale. *Frontiers in Psychology*, 4.
<https://doi.org/10.3389/fpsyg.2013.00279>
- Butler, L. H., Prike, T., & Ecker, U. K. H. (2024). Nudge-based misinformation interventions are effective in information environments with low misinformation prevalence. *Scientific Reports*, 14(1), 11495.
<https://doi.org/10.1038/s41598-024-62286-7>
- Ceylan, G., Anderson, I. A., & Wood, W. (2023). Sharing of misinformation is habitual, not just lazy or biased. *Proceedings of the National Academy of Sciences*, 120(4), e2216614120.
<https://doi.org/10.1073/pnas.2216614120>
- Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through

inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE*, 12(5), e0175799. <https://doi.org/10.1371/journal.pone.0175799>

- De Coninck, D., Frissen, T., Matthijs, K., d'Haenens, L., Lits, G., Champagne-Poirier, O., Carignan, M.-E., David, M. D., Pignard-Cheynel, N., Salerno, S., & G  n  reux, M. (2021). Beliefs in Conspiracy Theories and Misinformation About COVID-19: Comparative Perspectives on the Role of Anxiety, Depression and Exposure to and Trust in Information Sources. *Frontiers in Psychology*, 12, 646394. <https://doi.org/10.3389/fpsyg.2021.646394>
- Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The Psychology of Conspiracy Theories. *Current Directions in Psychological Science*, 26(6), 538–542. <https://doi.org/10.1177/0963721417718261>
- Duetz, J. C. M. (2023). What Does It Mean for a Conspiracy Theory to Be a 'Theory'? *Social Epistemology*, 37(4), 438–453. <https://doi.org/10.1080/02691728.2023.2172697>
- Dyer, K. D., & Hall, R. E. (2019). Effect of Critical Thinking Education on Epistemically Unwarranted Beliefs in College Students. *Research in Higher Education*, 60(3), 293–314. <https://doi.org/10.1007/s1162-018-9513-3>
- Epstein, Z., Berinsky, A. J., Cole, R., Gully, A., Pennycook, G., & Rand, D. G. (2021). Developing an accuracy-prompt toolkit to reduce COVID-19 misinformation online. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-71>
- Erisen, C. (2022). Psychological foundations and behavioral consequences of COVID-19 conspiracy theory beliefs: The Turkish case. *International Political Science Review*. <https://doi.org/10.1177/01925121221084625>
- 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>
- F  rster, M., Weiser, C., & Maur, A. (2018). How feedback provided by voluntary electronic quizzes affects learning outcomes of university students in large classes. *Computers & Education*, 121, 100–114. <https://doi.org/10.1016/j.compedu.2018.02.012>
- Frenken, M., Reusch, A., & Imhoff, R. (2024). "Just Because It's a Conspiracy Theory Doesn't Mean They're Not Out to Get You": Differentiating the Correlates of Judgments of Plausible Versus Implausible Conspiracy Theories. *Social Psychological and Personality Science*, 19485506241240506. <https://doi.org/10.1177/19485506241240506>
- Galesic, M., & Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, 73(2), 349–360. <https://doi.org/10.1093/poq/nfp031>
- Grawitch, M. J., & Lavigne, K. (2021). Do Attitudes, Trust, and Acceptance of Pseudoscience and Conspiracy Theories Predict COVID-19 Vaccination Status? (rayyan-968513171). <https://doi.org/10.31234/osf.io/tg7xr>
- Greene, C. M., De Saint Laurent, C., Murphy, G., Prike, T., Hegarty, K., & Ecker, U. K. H. (2023). Best Practices for Ethical Conduct of Misinformation Research: A Scoping Review and Critical Commentary. *European Psychologist*, 28(3), 139–150. <https://doi.org/10.1027/1016-9040/a000491>
- Greene, C. M., & Murphy, G. (2021). Quantifying the effects of fake news on behavior: Evidence from a study of COVID-19 misinformation. *Journal of Experimental Psychology: Applied*. <https://doi.org/10.1037/xap0000371>
- Guay, B., Berinsky, A., Pennycook, G., & Rand, D. G. (2022). How To Think About Whether Misinformation Interventions Work. <https://doi.org/10.31234/osf.io/gv8qx>
- Hameleers, M. (2023). The (Un)Intended Consequences of Emphasizing the Threats of Mis- and Disinformation. *Media and Communication*, 11(2). <https://doi.org/10.17645/mac.v11i2.6301>
- Hattersley, M., Brown, G. D. A., Michael, J., &

- Ludvig, E. A. (2022). Of tinfoil hats and thinking caps: Reasoning is more strongly related to implausible than plausible conspiracy beliefs. *Cognition*, 218, 104956. <https://doi.org/10.1016/j.cognition.2021.104956>
- Hoes, E., Aitken, B., Zhang, J., Gackowski, T., & Wojcieszak, M. (2023). *Prominent Misinformation Interventions Reduce Misperceptions but Increase Skepticism*. <https://doi.org/10.31234/osf.io/zmpdu>
 - Janouskova, A., Kocyan, J., Simova, M., Uvirova, K., Zahradnickova, K., Vaculik, M., & Prochazka, J. (2022). The effect of font readability on the Moses illusion: A replication study. *Consciousness and Cognition*, 99, 103284. <https://doi.org/10.1016/j.concog.2022.103284>
 - Jolley, D., & Douglas, K. M. (2017). Prevention is better than cure: Addressing anti-vaccine conspiracy theories. *Journal of Applied Social Psychology*, 47(8), 459–469. <https://doi.org/10.1111/jasp.12453>
 - Jolley, D., Marques, M. D., & Cookson, D. (2022). Shining a spotlight on the dangerous consequences of conspiracy theories. *Current Opinion in Psychology*, 47, 101363. <https://doi.org/10.1016/j.copsyc.2022.101363>
 - Kiili, K., Siuko, J., & Ninaus, M. (2024). Tackling misinformation with games: A systematic literature review. *Interactive Learning Environments*, 1–16. <https://doi.org/10.1080/10494820.2023.2299999>
 - Lawson, T. J., Jordan-Fleming, M. K., & Bodle, J. H. (2015). Measuring Psychological Critical Thinking: An Update. *Teaching of Psychology*, 42(3), 248–253. <https://doi.org/10.1177/0098628315587624>
 - Lee, E.-J., & Jang, J. (2023). How Political Identity and Misinformation Priming Affect Truth Judgments and Sharing Intention of Partisan News. *Digital Journalism*, 11(1), 226–245. <https://doi.org/10.1080/21670811.2022.2163413>
 - Maertens, R., Anseel, F., & Van Der Linden, S. (2020). Combatting climate change misinformation: Evidence for longevity of inoculation and consensus messaging effects. *Journal of Environmental Psychology*, 70, 101455. <https://doi.org/10.1016/j.jenvp.2020.101455>
 - Mason, A. M., Compton, J., Tice, E., Peterson, B., Lewis, I., Glenn, T., & Combs, T. (2024). Analyzing the Prophylactic and Therapeutic Role of Inoculation to Facilitate Resistance to Conspiracy Theory Beliefs. *Communication Reports*, 37(1), 13–27. <https://doi.org/10.1080/08934215.2023.2256803>
 - McGuire, W. J. (1964). Some Contemporary Approaches. In *Advances in Experimental Social Psychology* (Vol. 1, pp. 191–229). Elsevier. [https://doi.org/10.1016/S0065-2601\(08\)60052-0](https://doi.org/10.1016/S0065-2601(08)60052-0)
 - McGuire, W. J., & Papageorgis, D. (1961). The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *The Journal of Abnormal and Social Psychology*, 62(2), 327–337. <https://doi.org/10.1037/h0042026>
 - Media Literacy Ireland (2020). Be Media Smart. <https://www.bemediasmart.ie/>
 - Murphy, G., De Saint Laurent, C., Reynolds, M., Aftab, O., Hegarty, K., Sun, Y., & Greene, C. M. (2023). What do we study when we study misinformation? A scoping review of experimental research (2016–2022). *Harvard Kennedy School Misinformation Review*, 4(6). <https://doi.org/10.37016/mr-2020-130>
 - Murphy, G., & Greene, C. M. (2023). Conducting ethical misinformation research: Deception, dialogue, and debriefing. *Current Opinion in Psychology*, 54, 101713. <https://doi.org/10.1016/j.copsyc.2023.101713>
 - Nera, K., Pantazi, M., & Klein, O. (2018). “These Are Just Stories, Mulder”: Exposure to Conspiracist Fiction Does Not Produce Narrative Persuasion. *Frontiers in Psychology*, 9, 684. <https://doi.org/10.3389/fpsyg.2018.00684>
 - O'Mahony, C., Brassil, M., Murphy, G., & Linehan, C. (2023). The efficacy of interventions in reducing belief in conspiracy

- theories: A systematic review. *PLOS ONE*, 18(4), e0280902. <https://doi.org/10.1371/journal.pone.0280902>
- O'Mahony, C., Linehan, C., & Murphy, G. (2024). *The Critical Thinking about Conspiracies (CTAC) Test: Development and Validation*. <https://doi.org/10.31234/osf.io/xvvh3>
 - Orosz, G., Krekó, P., Paskuj, B., Tóth-Király, I., Bóthe, B., & Roland-Lévy, C. (2016). Changing Conspiracy Beliefs through Rationality and Ridiculing. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01525>
 - Parker, K. A., Ivanov, B., & Compton, J. (2012). Inoculation's Efficacy With Young Adults' Risky Behaviors: Can Inoculation Confer Cross-Protection Over Related but Untreated Issues? *Health Communication*, 27(3), 223–233. <https://doi.org/10.1080/10410236.2011.575541>
 - 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. *Psychological Science*, 31(7), 770–780. <https://doi.org/10.1177/0956797620939054>
 - Pennycook, G., & Rand, D. G. (2022a). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*, 13(1), 2333. <https://doi.org/10.1038/s41467-022-30073-5>
 - Pennycook, G., & Rand, D. G. (2022b). Nudging Social Media toward Accuracy. *The ANNALS of the American Academy of Political and Social Science*, 700(1), 152–164. <https://doi.org/10.1177/00027162221092342>
 - Pfau, M., Ivanov, B., Houston, B., Haigh, M., Sims, J., Gilchrist, E., Russell, J., Wigley, S., Eckstein, J., & Richert, N. (2005). Inoculation and Mental Processing: The Instrumental Role of Associative Networks in the Process of Resistance to Counterattitudinal Influence. *Communication Monographs*, 72(4), 414–441. <https://doi.org/10.1080/03637750500322578>
 - Poon, K.-T., Chen, Z., & Wong, W.-Y. (2020). Beliefs in Conspiracy Theories Following Ostracism. *Personality and Social Psychology Bulletin*, 46(8), 1234–1246. <https://doi.org/10.1177/0146167219898944>
 - Pytlik, N., Soll, D., & Mehl, S. (2020). Thinking Preferences and Conspiracy Belief: Intuitive Thinking and the Jumping to Conclusions-Bias as a Basis for the Belief in Conspiracy Theories. *Frontiers in Psychiatry*, 11, 568942. <https://doi.org/10.3389/fpsyg.2020.568942>
 - Rogers, R. W., & Thistlethwaite, D. L. (1969). An analysis of active and passive defenses in inducing resistance to persuasion. *Journal of Personality and Social Psychology*, 11(4), 301–308. <https://doi.org/10.1037/h0027354>
 - Roozenbeek, J., Freeman, A. L. J., & Van Der Linden, S. (2021). How Accurate Are Accuracy-Nudge Interventions? A Preregistered Direct Replication of Pennycook et al. (2020). *Psychological Science*, 32(7), 1169–1178. <https://doi.org/10.1177/09567976211024535>
 - Roozenbeek, J., & Van Der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(1), 65. <https://doi.org/10.1057/s41599-019-0279-9>
 - Roozenbeek, J., Van Der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, 8(34), eabo6254. <https://doi.org/10.1126/sciadv.abo6254>
 - Sallam, M., Dababseh, D., Eid, H., Hasan, H., Taim, D., Al-Mahzoum, K., Al-Haidar, A., Yaseen, A., Ababneh, N. A., Assaf, A., Bakri, F. G., Matar, S., & Mahafzah, A. (2021). Low COVID-19 Vaccine Acceptance Is Correlated with Conspiracy Beliefs among University Students in Jordan. *International Journal of Environmental Research and Public Health*, 18(5). <https://doi.org/10.3390/ijerph18052407>
 - Stall, L. M., & Petrocelli, J. V. (2023). Countering conspiracy theory beliefs: Understanding the conjunction fallacy and considering disconfirming evidence. *Applied Cognitive Psychology*, 37(2), 266–276.

<https://doi.org/10.1002/acp.3998>

- Stojanov, A. (2015). Reducing conspiracy theory beliefs. *Psihologija*, 48(3), 251–266. <https://doi.org/10.2298/PSI1503251S>
- Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition*, 133(3), 572–585. <https://doi.org/10.1016/j.cognition.2014.08.006>
- Traberg, C. S., Roozenbeek, J., & Van Der Linden, S. (2022). Psychological Inoculation against Misinformation: Current Evidence and Future Directions. *The ANNALS of the American Academy of Political and Social Science*, 700(1), 136–151. <https://doi.org/10.1177/00027162221087936>
- U.K. Government (2020). Government cracks down on spread of false coronavirus information online. <https://www.gov.uk/government/news/government-cracks-down-on-spread-of-false-coronavirus-information-online>
- Valentine, D. (2017). *The CIA as organized crime: How illegal operations corrupt America and the world*. Clarity Press, Inc.
- 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 Meer, T. G. L. A., Hameleers, M., & Ohme, J. (2023). Can Fighting Misinformation Have a Negative Spillover Effect? How Warnings for the Threat of Misinformation Can Decrease General News Credibility. *Journalism Studies*, 24(6), 803–823. <https://doi.org/10.1080/1461670X.2023.2187652>
- Van Prooijen, J.-W., Krouwel, A. P. M., & Pollet, T. V. (2015). Political Extremism Predicts Belief in Conspiracy Theories. *Social Psychological and Personality Science*, 6(5), 570–578. <https://doi.org/10.1177/1948550614567356>
- Većkalov, B., Gligorić, V., & Petrović, M. B. (2024). No evidence that priming analytic thinking reduces belief in conspiracy theories: A Registered Report of high-powered direct replications of Study 2 and Study 4 from. *Journal of Experimental Social Psychology*, 110, 104549. <https://doi.org/10.1016/j.jesp.2023.104549>
- Vivion, M., Anassour Laouan Sidi, E., Betsch, C., Dionne, M., Dubé, E., Driedger, S. M., Gagnon, D., Graham, J., Greyson, D., Hamel, D., Lewandowsky, S., MacDonald, N., Malo, B., Meyer, S. B., Schmid, P., Steenbeek, A., van der Linden, S., Verger, P., Witteman, H. O., ... Canadian Immunization Research Network (CIRN). (2022). Prebunking messaging to inoculate against COVID-19 vaccine misinformation: An effective strategy for public health. *Journal of Communication in Healthcare*. <https://doi.org/10.1080/17538068.2022.2044606>
- Wagner-Egger, P. (2022). *The Noises of Conspiracy: Psychology of Beliefs in Conspiracy Theories*. <https://doi.org/10.31234/osf.io/fv52e>
- Williams, M. N., & Bond, C. M. C. (2020). A preregistered replication of “Inoculating the public against misinformation about climate change.” *Journal of Environmental Psychology*, 70, 101456. <https://doi.org/10.1016/j.jenvp.2020.101456>
- Wood, M. J. (2017). Conspiracy suspicions as a proxy for beliefs in conspiracy theories: Implications for theory and measurement. *British Journal of Psychology*, 108(3), 507–527. <https://doi.org/10.1111/bjop.12231>
- Wood, M. J., Douglas, K. M., & Sutton, R. M. (2012). Dead and Alive: Beliefs in Contradictory Conspiracy Theories. *Social Psychological and Personality Science*, 3(6), 767–773. <https://doi.org/10.1177/1948550611434786>