

Title	Teaching and measuring critical thinking as a Mmans of improving appraisal of conspiracy theories
Authors	O'Mahony, Cian James
Publication date	2024
Original Citation	O'Mahony, C. J. 2024. Teaching and measuring critical thinking as a Mmans of improving appraisal of conspiracy theories. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2024, Cian James O'Mahony. - https://creativecommons.org/licenses/by/4.0/
Download date	2026-10-27 11:52:41
Item downloaded from	https://hdl.handle.net/10468/17487



Teaching and Measuring Critical Thinking as a Means of Improving Appraisal of Conspiracy Theories

Cian James O'Mahony

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in
Applied Psychology

School of Applied Psychology
College of Arts, Celtic Studies, and Social Sciences
University College Cork

30th December 2024

Supervisors:

Dr. Gillian Murphy

Dr. Conor Linehan

Head of School: Dr. Chris McCusker

Table of Contents

List of Figures	v
List of Tables	v
Acknowledgments	vi
Declaration of Originality	viii
Chapter 1: Introduction	1
1.1 Conspiracy Theories.....	2
1.1.1 What is a conspiracy theory? Two competing views	2
1.1.2 What can be considered an unfounded conspiracy theory?.....	6
1.1.3 Is there such a thing as a conspiracy theorist?.....	12
1.2 How can critical thinking help solve this issue?	16
1.2.1 Benefits of using critical thinking	19
1.3 Research Questions	21
1.4 Thesis Outline	22
1.5 References	25
Chapter 2: The efficacy of interventions in reducing conspiracy beliefs: a systematic review	32
2.1 Abstract	33
2.2 Introduction	34
2.2.1 Conspiracy Interventions.....	35
2.2.2 Current Study	37
2.3 Methods.....	38
2.3.1 Search Strategy	38
2.3.2 Eligibility Criteria.....	39
2.3.3 Data Management.....	41
2.3.4 Data review and screening.....	41
2.3.5 Data extraction and analysis	42
2.4 Results	44
2.4.1 Characteristics of Studies	44
2.4.2 Efficacy of Interventions	45
2.4.3 Interventions aimed at reducing conspiracy beliefs	54
2.4.4 Interventions aimed at increasing conspiracy beliefs	57
2.5 Discussion	58
2.5.1 Limitations of studies examined in the review	61
2.5.2 Limitations of the current systematic review	62
2.5.3 Implications for research / practice / design.....	63

2.6 Conclusion.....	64
2.7 References	66
Chapter 3: The Critical Thinking about Conspiracies Assessment (CTAC): Development and Validation	74
3.1 Abstract	75
3.2 Introduction	76
3.2.1 Proposing a New Approach to Measuring Conspiratorial Thinking	79
3.2.2 Present Study	83
3.3 Study 1: Scale Development	84
3.3.1 Initial Scale format	84
3.3.2 Pilot Studies.....	86
3.3.3 Final CTAC Format.....	90
3.3.4 Participant Feedback	92
3.4 Study 2: Exploratory Factor Analysis	95
3.4.1 Methods	95
3.4.2 Results	96
3.5 Study 3: Confirmatory Factor Analysis	98
3.5.1 Methods	98
3.5.2 Results	100
3.6 Study 4: CTAC Long Form.....	103
3.6.1 Methods	103
3.6.2 Exploratory Factor Analysis.....	103
3.6.3 Confirmatory Factor Analysis	104
3.6.4 Results	105
3.7 Discussion	106
3.8 References	110
Chapter 4: True discernment or blind scepticism? Comparing the effectiveness of four conspiracy belief interventions.....	119
4.1 Abstract	121
4.1.1 Potential Interventions.....	123
4.1.2 Effectively Measuring Conspiracy Beliefs.....	126
4.1.3 Overview of Present Study	128
4.2 Study 1a.....	129
4.2.1 Methods	129
4.2.2 Results	140
4.2.3 Study 1a Discussion	142
4.3 Study 1b	142
4.3.1 Participants	143

4.3.2	Materials.....	143
4.3.3	Procedure.....	143
4.3.4	Results	145
4.3.5	Study 1b Discussion	147
4.4	Study 2	148
4.4.1	Methods.....	148
4.4.2	Results	152
4.5	General Discussion.....	159
4.6	References	165
Chapter 5: Designing a serious game to increase critical appraisal of conspiracy theories		175
5.1	Abstract	176
5.2	Introduction	177
5.2.1	Why use games to challenge unfounded conspiracy beliefs?.....	178
5.2.2	Why design a novel game?.....	179
5.2.3	How do you design a useful, successful educational game?	182
5.3	Game Design.....	183
5.3.1	Initial concept.....	183
5.4	Game Evaluation	200
5.4.1	Methods.....	200
5.4.2	Results	203
5.4.3	Narrative Analysis.....	204
5.5	General Discussion.....	207
5.6	References	213
Chapter 6: Discussion		218
6.1	Thesis Overview.....	218
6.2	Contributions.....	220
6.2.1	Novel Review Comparing Efficacy of Interventions	220
6.2.2	Operationalising Conspiracy Thinking.....	221
6.2.3	Highlighting Oversights in Current Conspiracy Intervention Research.....	222
6.3	Strengths.....	223
6.4	Limitations	224
6.5	Future Directions.....	226
6.6	Conclusion.....	228
6.7	References	229

List of Figures

Figure 1.1: Binary vs spectrum conception of conspiracy beliefs.....	166
Figure 2.1: PRISMA Flow Diagram	421
Figure 2.2: Effect sizes of interventions included in review	554
Figure 3.1: Item list for final 8-item CTAC	986
Figure 3.2. Two-factor solution of the CTAC	101
Figure 4.1: Infographic stimuli used in Chapter 4.....	1332
Figure 4.2. Results of Experiment 1 in Chapter 4	1475
Figure 4.3. Main results of Experiment 2 in Chapter 4.	1575
Figure 4.4. Results of secondary outcome measures for Experiment 2 in Chapter 4.	1586
Figure 5.2: Screenshot of interactive PowerPoint prototype for 'Arguing with Your Uncle Online'.	1863
Figure 5.3: Interactive PowerPoint prototype of Arguing with Your Uncle Online....	1875
Figure 5.4: UI concept art from Life is Strange: True Colors	1886
Figure 5.5: Godot prototype of Arguing with Your Uncle Online.....	1908
Figure 5.6: Godot prototype of Arguing with Your Uncle Online showing how players can construct responses.	1908
Figure 5.7: Comparison 'Sherlock Holmes: Crime and Punishment' and 'The Return of Obra Dinn'	19492
Figure 5.8: Screenshot of 'Nashville PI'.	1964

List of Tables

Table 2.1 Effectiveness of Interventions in Included Studies	48
Table 3.1: Factor loadings for 8-item version of CTAC	96
Table 3.2: Factor loadings for 15-item version of CTAC	104
Table 4.1 Correlations for Dependent Variables	152

Acknowledgments

This thesis has been the culmination of four years of work which would not have been possible without the support of many important people in my life. I am deeply grateful to everyone who has been part of this journey.

First and foremost, I extend my heartfelt thanks to my supervisors. I am immensely grateful to Gillian Murphy for your unwavering guidance and encouragement over the last four years. Your confidence in my potential, even during times when I doubted myself, has hugely aided me throughout this project. Despite your many responsibilities, you have generously dedicated both your time and energy to lend me your guidance. Your mentorship and positivity have helped me grow both professionally and personally and I am honoured to have gotten to know you better over the last four years. Conor, it has been a privilege to work under your mentorship once again after your stellar guidance during my bachelor's degree. The lessons I learned from you then have stayed with me, and I continue to learn from your wisdom, kindness, and humour. Thank you for embracing me into the PATLAB family and for stepping in as my primary supervisor during Gillian's leave. Your support has been a guiding light throughout my PhD journey.

I would like to extend my heartfelt thanks to Aaron Bolger for undertaking the Herculean task of programming and developing the UI for *Nashville PI*. Your exceptional coding and graphic design skills are matched only by your uncanny ability to read my mind when working on this game. At times, you have felt like a third supervisor, and your invaluable advice, particularly during the final stages of my PhD, has been a source of motivation and support when I needed it most.

I would like to extend my particular gratitude to the dear friends I have made in my PhD cohort in the School of Applied Psychology, who have become some of my closest friends over the last few years. I will sincerely treasure the friendship, empathy, and support you have offered me unconditionally these past few years. I am deeply grateful to have shared this journey with you.

To all the participants who kindly dedicated their time and effort to contribute to my studies. I would not have been able to complete this PhD without your efforts.

I am overwhelmingly thankful for the support I have received from the Irish Research Council, who funded this PhD project and allowed me to wholeheartedly dedicate all my efforts to conducting this research. I would also like to thank Google who have provided me with the opportunity to present my research and graciously invited me to many of their summits here in Ireland.

I want to thank my industry mentor Rebecca Umbach, who provided me with many opportunities to share my work with her lab at Google, and who served as an inspiring mentor over the last three years of my PhD.

I would like to thank Reagan Murphy who generously helped point me in the right direction in terms of developing the psychometric skills to create the CTAC in Chapter 3 of this thesis and would like to thank Mike Murphy for helping me to refine the CTAC. I want to thank Ella Doran who helped proofread our manuscript for Chapter 4 of this thesis.

Finally, I would like to thank my parents Damien and Mary O'Mahony whose love, encouragement, and unwavering support have been a constant source of strength throughout my life. You have always believed in me, and have inspired me to pursue my dreams, and I am forever grateful for your presence in my life.

Declaration of Originality

This thesis is my own work and has not been submitted for another degree, either at University College Cork or elsewhere

Cian O'Mahony

Chapter 1: Introduction

“[I]t is not enough to identify something as a conspiracy theory to dismiss it. Anyone concerned with the truth needs to have a way of distinguishing between plausible and implausible plots. The problem is that most people don’t have the time or knowledge to work out which is which.”

(Baggini, 2017, p.
38)

The above quote by philosopher Julian Baggini illustrates the primary focus of this thesis: rather than dismissing all conspiracy theories as untrue, people should learn critical thinking skills to help them distinguish between plausible and implausible conspiracy theories. In the last decade, researchers have attempted to design various ways to measure conspiracy ideation and develop psychological interventions to reduce people’s susceptibility to believing in conspiracy theories (O’Mahony et al., 2023).

Such interventions have typically required researchers to actively fact-check and debunk new conspiracy theories (Jolley & Douglas, 2017; Orosz et al., 2016); this approach is neither scalable nor sustainable. In this thesis, I take a fundamentally different approach, by developing scalable and effective methods to teach and assess how people critically evaluate conspiracy theories rather than focusing on mere belief in conspiracy theories. This PhD project aims to examine conspiracy beliefs through the lens of critical thinking, as a means of both measuring how people think about conspiracy theories and to increase people’s ability to tell the difference between plausible and implausible conspiracy theories.

1.1 Conspiracy Theories

1.1.1 What is a conspiracy theory? Two competing views

To effectively study conspiracy theories and measure the degree to which people endorse them, we need a clear and consistent definition. First, the Oxford Languages Dictionary simply defines a *conspiracy* as “a secret plan by a group to do something unlawful or harmful” (Oxford University Press, nd). There are many documented instances of conspiracies in history that meet these criteria, such as the Pazzi Conspiracy which was a plot to overthrow the Medici family as rulers of Florence in 1478 (Hibbert, 1979), and the Watergate Scandal where members of Richard Nixon’s re-election campaign attempted to plant listening devices in the Democratic National Committee headquarters in 1972 (Goldstein, 1974). A *conspiracy* is the description of the act of conspiring itself and is distinct from a *conspiracy theory*.

Definitions of conspiracy theories vary as they are frequently conflated with forms of misinformation and fake news. For example, Brotherton et al, (2013) state that conspiracy theories are “a subset of false beliefs, and generally implicate a malevolent force (e.g., a government body or secret society) involved in orchestrating major events or providing misinformation regarding the details of events to an unwitting public, in part of a plot towards achieving a sinister goal”. Defining conspiracy theories as a subset of false beliefs places them within the general category of misinformation and makes a general assumption about whether they are true or false. By defining conspiracy theories as a set of false beliefs, Brotherton et al, (2013) are taking what is commonly referred to as the *generalist* approach to defining conspiracy theories. The generalist approach is grounded in the philosophy that all conspiracy theories, or at the very least the majority of conspiracy theories, should be dismissed *a priori* as they are

the product of irrational thinking. As such, they use the label “theory” as way of dismissing conspiracy theories as being inherently untrue and flawed by their nature.

In contrast, the *particularist's* approach to conspiracy theories is to define them in neutral terms, making no judgements as to whether they are true or false. Particularists, as described by Buening and Taylor (2010), suggest that we cannot make sweeping broad generalisation about conspiracy theories as a class. Rather, conspiracy theories should be judged on a case-by-case basis using the available evidence at hand (Dentith, 2023). There is an inherently deep disagreement between the generalist approach which suggests that we have *prima facie* grounds to dismiss anything labelled a ‘conspiracy theory’, and the particularist approach who refrain from approaching conspiracy theories with any judgement about how likely they are before evaluating them (Duetz, 2023). Where generalists wish to define all conspiracies as false beliefs, particularists want to define them without any reference to their likelihood.

Importantly, what is considered to be a conspiracy theory can largely vary based on time and context. Typically, conspiracy theories are hypothetical. As discussed previously there have been several instances of verified conspiracies, but these cases are usually described as “affairs”, “plots” or “scandals”. This separation of the term “conspiracy” from actual instances of conspiracy theories is evident in the language used to describe them. Some historical examples of covert plots have been labelled as conspiracies, such as the Pazzi Conspiracy in 1478 (Hibbert, 1979), the Catiline Conspiracy in 63 BCE (Beard, 2015), and the Conspiracy of the Thirty (404 BCE) (Krentz, 1982), but these labels tend to be reserved for older examples predating the modern era (post-1500). The Gunpowder Plot of 1605, was a conspiracy where a group of reactionaries planned to explode the English Parliament at Westminster by covertly planting barrels of gunpowder under a vault beneath the House of Lords (Fraser, 1996).

While this event meets the requirements of a conspiracy (i.e., a malicious group acting in secret to do harm) it is typically referred to as a “plot” rather than a “conspiracy”. Similarly, more contemporary examples are often referred to as “scandals” such as the Cambridge Analytica Scandal (Rosenberg, 2018), the Watergate Scandal (Goldstein, 1974), and the LIBOR Scandal (Cookson, 2017). Documented occurrences of pharmaceutical and medical conspiracies are often referred to by their respective trials or experiment names such as the Tuskegee Syphilis Study (Baker et al., 2005), where the CDC intentionally left approximately 400 African American men untreated for syphilis, and *Abdullah v. Pfizer, Inc.* where an illegal clinical trial was conducted to test the efficacy of a treatment of meningococcal meningitis (Wise, 2001).

Reviewing historical examples of conspiracies reveals a recurring issue: the tendency to relabel conspiracy theories as legitimate investigations in hindsight. As noted by Uscinski (2018), ‘if conspiracy theorists investigate a theory that eventually turns out to be true, that theory stops being labelled conspiracy theory’. This definition of conspiracy theories relies on hindsight bias, as a conspiracy theory is retroactively relabelled to be rational only after it has been verified. If one were to apply the generalist approach to Uscinski’s definition, one would have to conclude that before a conspiracy theory is proven true, it must be treated as though it were entirely impossible. This perspective relies on creating a circular definition of conspiracy theories. This redefinition shifts the goalposts, ensuring that only unfounded claims count as conspiracy theories, making the argument tautological—a classic example of what philosophers and logicians call the “no-true-Scotsman fallacy” (Pirie, 2006). The no-true-Scotsman fallacy is a logical error where someone redefines a term to avoid admitting that a claim has been refuted. Anthony Flaw illustrates this fallacy with this titular example:

1. Person A: "No Scotsman (a man from Scotland) puts sugar in his porridge."
2. Person B: "But I am a Scotsman (a man from Scotland) and I put sugar in my porridge."
3. Person A: "Ah, but no *true* Scotsman (a man who is not only from Scotland, but also conforms to Scottish traditions/stereotypes) puts sugar in his porridge."

Similarly, generalists define verified conspiracy theories as not being irrational once they have been confirmed, suggesting that confirmed cases of conspiracies are not *true* conspiracy theories. Indeed, Deutz (2024) also concludes that the fundamental flaw in the generalist approach is that they would also be required to reject real conspiracy theories such as the mainstream account of the September 11 attacks by assuming that conspiracy narratives are inherently suspect. Had someone proposed that the United States were lying about Iraq using Weapons of Mass Destruction prior to the truth being revealed, would that theory also be dismissed as highly improbable and inherently flawed? While there are some valid reasons to suspect that conspiracy theories are often the product of poor reasoning and little evidence (which will be explored in the next section), it is clear that there is a fundamental problem in assuming that conspiracy theories are inherently false.

In this work I take a neutral perspective on defining conspiracy theories, in line with the particularist view. I work from the definition outlined by Douglas et al., (2017), who defines conspiracy theories as narratives “that explain important events as secret plots by powerful and malevolent groups”. Douglas’ definition makes no claims about whether conspiracy theories are true or false. Following this definition, I consider several key characteristics of what is considered a conspiracy theory.

1. Conspiracy theories invoke the work of secretive and often sinister groups.

2. Conspiracy theories suggest that important events such as catastrophes are caused by deliberate human interference rather than coincidences.
3. Conspiracy theories imply the intentional concealment of information. This characteristic helps clarify the distinction between mere urban myths and conspiracy theories. While the potential existence of Bigfoot or UFOs is not in itself a conspiracy theory, suggesting that a group is intentionally hiding the existence of these phenomena would qualify as one.
4. Conspiracy theories are inherently speculative and differ from verified conspiracies. This distinction is not to suggest that they are inherently false; but that they are proposed explanations for certain events rather than established accounts.

By defining conspiracy theories under this neutral terminology, I do not argue that true conspiracy theories are a common occurrence or that widespread belief in conspiracy theories is justified. Instead, I use this terminology to avoid the circular reasoning associated with defining conspiracy theories as a set of false beliefs. Furthermore, I do not make the case that the generalists are unjustified in their suspicion of conspiracy theories in general. In the next section, I consider some of the concerns of the generalist approach to defining conspiracy theories

1.1.2 What can be considered an unfounded conspiracy theory?

In the previous section, I critiqued the generalist approach to defining conspiracy theories. While not inherently false, many popular conspiracy theories have several characteristics that suggest they are unreasonable. Indeed, some generalists such

as Keith Harris (2018) suggest that conspiracy theories themselves are not problematic, but *conspiracy theorising*, the often-flawed thinking processes associated with conspiracy theories is the primary concern. I argue that there is some merit to Harris's approach. While I argue it is irrational to dismiss all conspiracy theories as false, there are elements of conspiracy theorising (what is refer to in this thesis as *conspiracy thinking*) that are inherently flawed. The flaws associated with conspiracy thinking will be discussed more thoroughly in the following section.

Brian Keeley (1999), a philosopher, sought to make a distinction between conspiracy theories as a general concept, and particular explanations that are irrational and should not be entertained. Keeley separates conspiracy theories from what he calls "*unwarranted conspiracy theories*" as a means of making a distinction between what types of explanations one ought to either dismiss or entertain.

Unwarranted conspiracy beliefs have several additional characteristics that set them apart from the neutral definition of conspiracy theories:

1. *Unwarranted conspiracy theories are an explanation that contradict some official or established account*

Keeley suggests that unwarranted conspiracy theories are almost always contrarian, contradicting any official narratives. Keeley's point is best illustrated by Edward Snowden's accounts of the public reception of his leaking of sensitive NSA documents in 2013. In an article published in the *Guardian*, Snowden (2021) noted that when he leaked classified files to show that the NSA was engaging in a conspiracy to spy on the public, the conspiracy theories that garnered most attention were the ones that contradicted Snowden's narrative. Some conspiracy theories suggested he was a Russian spy, others suggested that the CIA set him up to infiltrate the NSA. Despite

presenting the American public with clear evidence that their intelligence services were conspiring to violate the privacy of everyday citizens, Snowden was met with countless contrarian conspiracy theories that sought to contradict the established account being accepted by the general population. Snowden's experience illustrates Keeley's point that unfounded conspiracy theories tend to always contradict the general popular consensus about what happened during a particular event, even when a verified conspiracy has taken place. Similarly, the September 11 attacks were by definition a conspiracy to commit an act of terrorism on US soil. However, most popular discussions regarding the 9/11 attacks contradict most, if not all, of the official narratives surrounding the attack (Banas & Miller, 2013).

2. The intentions behind the conspiracy theory are always nefarious

While Keeley notes that the nefarious aspect of conspiracy theories may be obvious, he claims that many real conspiracy theories consist of people attempting to achieve positive benevolent goals. However, I argue that this factor is not particularly distinct to unwarranted conspiracy theories, as even the neutral definition of conspiracy theories suggests groups conspiring, with often harmful or malicious intentions.

3. Unfounded conspiracy theories tie together seemingly unrelated events

Keeley notes that a common characteristic of unwarranted conspiracy theories is that they often tie together unrelated events to use as evidence to support their claims. He uses the example of the Oklahoma City Bombing that occurred in 1995, which was a domestic terrorist attack carried out by former Gulf War veterans. In the aftermath of the attack, many conspiracy theories surface that tied the bombings in Oklahoma with the Sarin Gas attacks that had occurred on the other side of the world in Japan a month

later (Keeley, 1999). Despite the two events seemingly having no relation to one another, people were insistent that these two catastrophes were linked. Similarly, a more recent example can be seen in the response to the Sandy Hook Elementary School shootings that occurred in 2012, where political pundit Paul Joseph Watson drew similarities between the young-adult novel *The Hunger Games* and the shootings, claiming that 23 children were killed in the novel, while 20 children were killed in the shooting. Watson interprets these loose similarities as evidence of some grander plot that contradicts the official narrative surrounding the Sandy Hook shootings (Gawker, 2012).

4. *The truth behind these conspiracy theories are well-guarded secrets, even if the conspirators are well-known figures in the public eye*

This factor is based on the assumption of those who support conspiracy theories believe that if the truth about a conspiracy were exposed, it would completely unravel the conspiracy. Those who endorse conspiracy theories believe that this is why conspirators prioritise keeping their plans and evidence of their involvement secret, for if their plans were exposed that would be all that is needed for the conspiracy to collapse.

5. *Unwarranted conspiracy theories rely on “errant data”.*

Keeley suggests that unwarranted conspiracy theories rely heavily on what he calls “errant data” (1999). He suggests errant data exists in two forms: unaccounted-for-data and contradictory data. Unaccounted-for-data does not contradict established data but addresses the gaps that the official account does not address. An example of unaccounted data is “the Umbrella Man” who was an individual spotted numerous times

on the day that John F Kennedy was assassinated. The man was seen carrying an umbrella despite the sunny conditions and as a result, has become an object of speculation. Some conspiracy theories suggest that the Umbrella Man was in some way involved in Kennedy's assassination. Because the Warren Commission did not address the presence or role of the Umbrella Man, his existence has been used as unaccounted-for data to support several unwarranted conspiracy theories. Contradictory data is data that if proven to be true would contradict the official account. For example, if evidence surfaced that Covid-19 did not originate from an animal, but was a biological weapon leaked from a lab, this evidence would completely contradict the official narrative surrounding the origins of the virus.

Finally, another identifying characteristic that distinguishes unwarranted conspiracy theories from general conspiracy theories is how they are structured as a 'theory'. Philosophy researcher Julia Deutz (2023) suggests that many disagreements over the validity of conspiracy theories are less to do with the specific arguments or evidence used to support them, but more so due to a deep disagreement about what people mean when they say that a conspiracy theory is a 'theory'. The term theory is used to refer to three generally distinct concepts: (i) an *established account*, which is considered a widely accepted explanation that survived rigorous scrutiny and is considered true by most experts (such as the theory of gravity or evolution), (ii) a *hypothesis*, which is a testable and educated speculation that is developed in an informed attempt to discover the truth, and (iii) a *hunch*, which is uninformed conjecture that has not been subjected to careful reflection and are subjectively motivated rather than based on factual reasons, without considering alternative explanations.

Deutz (2023) suggests people who are sympathetic to conspiracy narratives and those who are sceptical of conspiracy narratives differ in how they define what a ‘theory’ is. People who are generally sceptical of conspiracy theories view official accounts as an ‘established account’ and that conspiracy theories are nothing more than hunches. People who are generally accepting of conspiracy narratives tend to think of both the official account and conspiracy theories as equally valid hypotheses. Deutz suggests that the different ways that the two opposing schools of thought define ‘theories’ are responsible for most of the disagreements.

Deutz's framework does not necessarily make any judgements regarding which perspective is more valid but uses the framework to illustrate the differing perspectives of those who generally support conspiracy theories and those who reject them. However, I suggest that this framework allows us to identify the fundamental flaw of unfounded conspiracy theories. Unfounded conspiracy theories are theories that present themselves as educated conjectures informed by solid data. In actuality, unfounded conspiracy theories are hunches that are uninformed conjectures that are founded on poor evidence and intuition that do not justify the conclusions they propose.

A useful distinction can be made between conspiracy theories defined broadly as a concept, and what I will refer to from here on as *unfounded conspiracy theories*, which are narratives that falsely present themselves as hypotheses when in fact, they are hunches founded on errant data. Using Keeley’s framework; I refer to conspiracy theories that one ought not to believe in given the available evidence at hand as unfounded conspiracy theories. The concept of unfounded conspiracy theories does not suggest these conspiracy theories are untrue, but that given the available evidence, one does not have sufficient reason to believe in them. This thesis concerns itself with challenging the problematic thinking associated with unfounded conspiracy theories.

1.1.3 Is there such a thing as a conspiracy theorist?

A common belief about conspiracy theories is that there is a clear distinction between those who believe in them and those who do not. Those who believe in conspiracy theories are considered 'conspiracy theorists'. Articles such as 'My husband became a conspiracy theorist. Would our marriage survive?' (Howe, 2024) or 'How to talk to conspiracy theorists—and still be kind' (Newman, 2020) would indicate that there are two categories of people: conspiracy theorists and 'normal' people. This perception of conspiracy theorists as individuals who are separated from reality is reinforced further by conspiracy research that aimed to pathologise conspiracy beliefs (Radnitz & Underwood, 2017; J. Van Prooijen & Douglas, 2018). Previous conspiracy research sought to understand which factors contributed to people endorsing conspiracy theories and found that paranoia, delusions, and schizotypy personality disorder were predictors of conspiracy belief (Darwin et al., 2011; March & Springer, 2019). Such associations suggested that conspiracy beliefs were reserved for paranoid individuals who were incapable of rational judgment (Hofstadter, 1964). Such a conclusion led researchers to conclude that conspiracy theories were created and endorsed by fringe members of society, solidifying the idea that there is a clear line between those who we would define as a conspiracy theorist and those who are not.

Recent research has contradicted the assumption that there is a clear distinction between conspiracy theorists and 'normal' people. On average, most people believe in at least one conspiracy theory (Bowes et al., 2023), and a cross-sectional study during Covid-19 found that at least a third of participants believed in at least one Covid-19 conspiracy theory (Earnshaw et al., 2020). In another study Oliver and Wood (2014) found that over half of participants endorsed at least one of seven conspiracy theories

regarding chemtrails (the concept that the vapour trails left by aircrafts are actually harmful chemicals).

If everybody believes in conspiracy theories to some extent, does this suggest that the concept of a conspiracy theorist is a myth? There is a body of research suggesting that people have a general tendency to endorse conspiracy theories (or not). For example, research has found that one of the strongest predictors for belief in a conspiracy theory is belief in other conspiracy theories, even if the topics and content are entirely unrelated (Goertzel, 1994; Wood et al., 2012). Goertzel suggested that belief in conspiracy theories is part of a larger *monological belief system*, which makes a given person more likely to endorse conspiracy narratives, regardless of the content. Various researchers have created assessments that measure the degree to which people generally endorse conspiracy theories such as the Generic Conspiracist Beliefs Scale (Brotherton et al., 2013) and the Conspiracy Mentality Questionnaire (Bruder et al., 2013). Research using these assessments would suggest that the general tendency to endorse conspiracy theories is an individual difference that varies within the population. Rather than conspiracy theories being something that only the paranoid and delusional endorse, they are narratives that everybody entertains on a spectrum. This spectrum would still suggest that there are individuals with high conspiracy ideation who are likely to entertain most conspiracy theories they encounter. Such individuals may meet the criteria for the stereotypical "tin foil hat-wearing conspiracy theorist".

While there is evidence to suggest that conspiracy theories form part of a monological belief system, there is considerable variation in which conspiracy theories people who engage in conspiracy content endorse or reject. Topic modelling research of the popular conspiracy-related subreddit r/conspiracy theory found that there were many factions of members who endorsed specific conspiracy theories but outright rejected

others. For example, many users endorsed conspiracy theories that vaccines caused autism but were not convinced by narratives that suggested the government was hiding the existence of Bigfoot (Klein et al, 2018). In fact, US Patriots and Anti-Imperialist subgroups were at complete odds with one another. Despite not sharing the same first-order beliefs about which conspiracy theories are true, the authors suggested that the members of this community may share second-order beliefs (i.e., institutions and industry do not have your best interests at heart, powerful institutions are lying to you about the truth). These results suggest that while conspiracy theorists may not all share the same beliefs, they do share a general scepticism of established accounts and powerful institutions (Klein et al., 2018). Additionally, the authors did note a specific subgroup they labelled as 'true believers' who did in fact endorse most conspiracy theories they encountered but only consisted of 5% of all users on the forum (Klein et al., 2018). These true believers may fit the conventional concept of the "tin foil hat-wearing conspiracy theorist" but only make up a small minority of the spectrum of those who endorse conspiracy theories. These findings suggest that contrary to the belief that there is a clear distinction between conspiracy theorists and 'normal' people, conspiracy theories can negatively impact anyone. Therefore, if we are interested in mitigating the damages that may be associated with indulging unfounded conspiracy theories, it is important not to focus primarily on this false distinction. Instead, we should consider that the negative impact of conspiracy theories as potentially applying to anyone.

One of the core elements that underlies the hard distinction between a 'conspiracy theorist' and a 'normal' individual is the assumption made by generalists that it is never rational to endorse a conspiracy theory. Therefore, rejecting a conspiracy theory for any reason is seen as entirely rational. As discussed earlier, this approach is flawed as it relies on circular reasoning and does not leave room for the possibility of

plausible conspiracy theories. I suggest that similar to individuals with the tendency to accept all conspiracy theories, there is also a tendency for people to reject conspiracy theories regardless of the evidence for or against them. I will refer to this tendency as blind scepticism. While a generalist may divide conspiracy beliefs into two distinct categories, I propose that conspiracy ideation can be viewed on a spectrum from total acceptance of any conspiracy to total rejection of any conspiracy theory as can be seen in Figure 1.1.



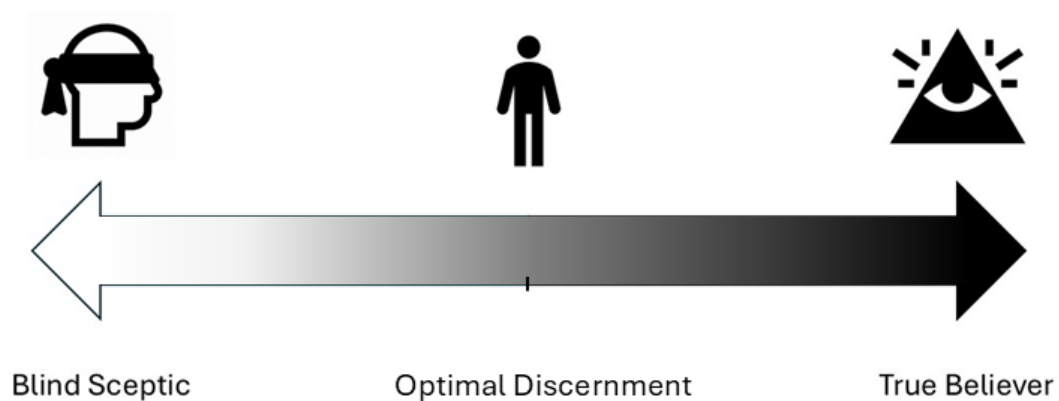


Figure 1.1: (Top) Binary approach to conceptualising conspiracy theorists. (Bottom) Spectrum showing general undiscerning acceptance of conspiracy theories to general blind scepticism of conspiracy theories.

1.2 How can critical thinking help solve this issue?

People who label themselves as conspiracy theorists often consider themselves "free thinkers" who can see past the lies perpetuated by the system. The majority of empirical literature would contradict this statement, suggesting that conspiracy beliefs are associated with lower levels of critical thinking and cognitive ability (Lantian et al., 2021; Stasielowicz, 2022), along with scientific literacy (Allred & Bolton, 2024).

Evidence strongly suggests that a lack of critical thinking is associated with the endorsement of conspiracy theories. Van Prooijen, (2019) proposes the gullible conspiracy hypothesis which suggests that rather than conspiracy theories being the product of rational scientific reasoning, they are the product of flawed logic and gullibility. Van Prooijen's reasoning is founded upon the observation that belief in conspiracy theories is associate with several key characteristics:

a) conspiracy beliefs are associated with other unfounded beliefs

Belief in conspiracy theories has been associated with belief in paranormal occurrences, and pseudoscientific claims (Bensley & Lilienfeld, 2020; Lobato et al., 2014). Furthermore, additional research has suggested that those who score higher in conspiracy belief are more likely to believe in supernatural phenomena such as PSI, and precognition (Darwin et al., 2011). Conspiracy beliefs are also correlated with a higher tendency to reject scientific research, believe misconceptions about psychology, and support poor research practices (Rutjens & Većkalov, 2022). Finally, conspiracy theorists are more likely to believe in pseudo-profound statements, complex sentences that create the impression of conveying a deep message, but ultimately have no underlying meaning (Pennycook et al., 2015). If belief in conspiracy theories were associated with critical thinking you would expect those who believed in conspiracy theories would be less likely to endorse unsubstantiated beliefs. However, the evidence suggests the contrary.

b) conspiracy beliefs are tied to cognitive biases and cognitive styles

Belief in conspiracies also has been shown to be correlated with several cognitive biases and cognitive styles. Belief in conspiracy theories is related to people overestimating the probability of certain events occurring. For instance, those who believe in conspiracy theories are more susceptible to the cognitive bias known as the “conjunction fallacy” (Brotherton & French, 2014). The conjunction fallacy is where a person overestimates the likelihood of two or more events co-occurring. For instance, some US citizens claim that President Joe Biden rigged the 2020 election by orchestrating a coordinated effort involving voters, poll workers, election security, and all U.S. courts. According to the conjunction fallacy, this explanation is actually less likely than any single one of these groups acting alone (Stall & Patroli, 2022). However, proponents find the explanation

more convincing because the detailed narrative creates a false sense of plausibility, even though the probability of so many actors colluding without error or exposure is extremely low. Those who believe in conspiracy theories are also more susceptible to the jumping-to-conclusions bias, where people fail to recognize the need for more information before drawing conclusions. In one study, participants were told about two lakes. In Lake A, 60% of the fish were white, and 40% were black. In Lake B, 40% of the fish were white, and 60% were black. Participants were told they would see images from one of the lakes and would be asked to guess which lake the images came from. Participants could ask to see more photos or draw their conclusion (Pytlik et al., 2020). Those who endorsed more conspiracy beliefs viewed fewer photos before stating their conclusions regarding which lake the photos came from.

c) conspiracy beliefs are exacerbated by emotion.

Belief in conspiracy theories is often exacerbated by an emotional state of mind. Van Prooijen (2019) notes that those who engage in a System 1 form of thinking (i.e. intuitive and emotional) as opposed to System 2 thinking (deliberative and analytical) are more likely to believe in conspiracy theories. Feels of powerlessness (Abalakina-Paap et al., 1999), lack of control (Van Prooijen & Acker, 2015; Whitson & Galinsky, 2008), anxiety (Grzesiak-Feldman, 2013) and perceived ostracism (Poon et al., 2020) are all associated with an increase in conspiracy beliefs.

The above evidence would suggest that critical thinking reduces belief in conspiracy theories. As I will discuss several times over the coming chapters, critical thinking interventions have successfully reduced someone's tendency to believe in conspiracy theories. I propose that examining conspiracy theories through the lens of critical thinking has several advantages, as it allows us to examine the thinking processes associated with endorsing unfounded conspiracy beliefs.

1.2.1 Benefits of using critical thinking

Critical thinking has long been considered a sought-after skill in education, but its definition has varied. Some have referred to critical thinking as “the intellectually disciplined process of actively and skilfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide belief and action” (Scriven & Paul, 1997). Later researchers such as Halpern (2003) describe critical thinking as “thinking that is purposeful, reasoned, and goal directed” in the context of “solving problems, formulating inferences, calculating likelihoods, and making decisions”. Lantian et al. (2021) argue that critical thinking provides a pathway to challenging conspiracy beliefs, which they consider a high-level skill that involves “rationality, cognitive skills, and analytical thinking”. Specifically, they suggest that the use of critical thinking to engage in appraising arguments and to formulate responses is crucial in terms of challenging unfounded conspiracy theories. As I have discussed previously, unfounded conspiracy theories are often based on a series of unwarranted assumptions or fallacious arguments (conjunction fallacy etc). Lantian et al argue that critical thinking provides people with the ability to identify inferential relations in arguments, spot when unwarranted assumptions are made, and judge the credibility and plausibility of multiple explanations when posed with a conspiratorial narrative (2021).

There are several applied benefits to using critical thinking as a means of both measuring how people think about conspiracy theories and teaching critical thinking skills to challenge unfounded conspiracy beliefs. The first advantage of using applying critical thinking as a lens to study conspiracy thinking is that it is a more concrete and tangible skill to teach participants. As we discussed earlier, conspiracy theories thrive in uncertainty, and we have no way of establishing a priori if a conspiracy theory is likely

to be true or not. However, it is much easier to identify flawed reasoning that is being used to support a conspiracy theory. The focus on challenging the thinking processes underlying belief in conspiracy theories aligns with the view of the generalist researcher Harris discussed earlier, where he notes that the problem is not with conspiracy theories but with conspiracy thinking. The advantage of using critical thinking is that it allows us to challenge and measuring flawed conspiracy thinking.

The second advantage is that critical thinking is a scalable means of measuring and challenging conspiracy beliefs. As I will argue in Chapters 2 and 3, current conspiracy measures and interventions are either too specific (challenging specific conspiracy beliefs such as anti-vaccine), or too broad (asks people about their general suspicions of the government). The problem with interventions that challenge specific topics is that they must constantly develop new interventions once a new conspiracy has emerged. Measuring and teaching critical thinking allows us to avoid this problem by examining thinking processes that are demonstrably false and to teach skills that people can apply to novel conspiracy theories.

The last benefit of measuring and teaching critical thinking is that it offers a more democratic means of measuring and teaching susceptibility to unfounded conspiracy theories. By telling participants in our studies which conspiracy theories they should accept or reject, we as researchers are acting as arbiters of truth. Particularly since conspiracy theories are by nature ambiguous, we cannot with certainty say that these conspiracy theories are false. Whereas it is a lot more feasible to identify established logical fallacies (i.e., conjunction fallacy, jumping-to-conclusions bias etc) and flawed reasoning, and to instruct participants how to avoid these errors in judgement. I propose that moving away from debunking specific conspiracy theories

and moving towards identifying and addressing problematic thinking patterns offers a much more democratic and scalable means of tackling unfounded conspiracy theories.

1.3 Research Questions

The research questions underlying this thesis are centred around the following question: how can we measure and teach critical thinking skills as a means of improving critical discernment between plausible and implausible conspiracy theories? For this purpose, this thesis consists of four studies that aim to address a core gap in the existing literature. First, there has been no review to date that has comprehensively compared the efficacy of psychological interventions to challenge conspiracy beliefs. To design interventions to increase critical appraisal of conspiracy theories, there needs to be a review to see how both interventions and measurements for conspiracy ideation have been handled to date. Second, in order to teach critical appraisal of conspiracy theories, there needs to be a way of measuring whether participants are learning those skills. In Study 2 I design a novel assessment to achieve this goal. Third, there needs to be an experimental study to establish whether critical appraisal of conspiracy theories can be increased by way of promising interventions. Based on the most effective interventions identified in Study 1, there needs to be a direct comparison to identify which interventions are the most effective at increasing critical appraisal of conspiracy theories. Finally, I want to ascertain whether the principles we learned from Study 3 can be condensed into a serious game that would likely be more enjoyable to take part in and motivate more people to enrol in this intervention.

Study 1

- a. How effective are existing interventions at reducing conspiracy beliefs?

- b. How effective are current measures of conspiracy thinking?

Study 2

- a. How can critical appraisal of conspiracy theories be conceptualised and measured reliably?
- b. How well does this novel assessment correlate with general conspiracy ideations and critical thinking?

Study 3

- a. How do potential conspiracy interventions compare in terms of increasing critical appraisal of conspiracy theories?
- b. Does encouraging reasoning over blind scepticism improve discernment between plausible and implausible conspiracy theories?

Study 4

- a. Can a serious game effectively encourage critical evaluation of conspiracy theories?

1.4 Thesis Outline

This thesis consists of four papers which examine the interaction between critical thinking skills and conspiracy beliefs. Two of these papers have been published, one is currently under review, and the remaining paper is being prepared as two papers for publication after submission of the thesis. Each chapter contains a foreword which details the motivation behind the study, the contribution of each author, and the publication status of the paper.

Chapter 2 consists of a systematic review of the efficacy of previous interventions at reducing conspiracy beliefs, along with examining the various approaches used to measure conspiracy thinking. This paper has been published in *PLOS One*.

Chapter 3 presents a novel method of measuring conspiracy thinking, that focuses on measuring how people apply critical thinking skills to conspiracy theories. This assessment was designed to address the oversights highlighted in Chapter 2 of current conspiracy thinking measures which focused primarily on belief rather than the underlying cognitive processes of conspiratorial thinking. This paper has been submitted for publication in *Applied Cognitive Psychology* and is currently under review.

Chapter 4 directly compares the efficacy of several potential interventions in terms of increasing critical appraisal of conspiracy theories, building upon the most promising conspiracy interventions identified in Chapter 2. This study is novel in that it is the first to directly compare different conspiracy intervention types experimentally in terms of their ability to reduce susceptibility to implausible conspiracy theories. The findings indicate that current conspiracy interventions may be increasing blind scepticism, and that interventions that encourage participants to use reasoning instead of blindly dismissing conspiracy theories may help distinguish between plausible and implausible conspiracy theories.

This paper has been published in *Advances in Psychology*.

Chapter 5 further explores the efficacy of conspiracy interventions by condensing the principles of our previous successful interventions into an online serious game that teaches players how to apply critical thinking skills to appraise both plausible and implausible conspiracy theories.

Finally, Chapter 6 discusses the findings of the previous chapters with reference to the literature on critical thinking and conspiracy thinking. This chapter discusses both the theoretical and practical implications of these findings, along with recommendations for future investigations.

I have included additional papers I have worked on during the duration of my PhD project in the Appendices of this thesis. In Appendix 1 I have included a systematic review that assesses the current literature on the role of executive function on the misinformation effect. In Appendix 2 I include a paper I wrote with my colleague John Twomey that is based on the findings of systematic review in Chapter 2. In this paper, we discuss the implications deepfake media may have in terms of fuelling deepfake conspiracy theories and outline several guidelines for future research that aims to design interventions to challenge deepfake conspiracy theories. This paper has been accepted for publication in *ACM Interactions*.

1.5 References

- Abalakina-Paap, M., Stephan, W. G., Craig, T., & Gregory, W. L. (1999). Beliefs in Conspiracies. *Political Psychology*, 20(3), 637–647.
- Allred, N., & Bolton, L. E. (2024). Conspiracy Beliefs and Consumption: The Role of Scientific Literacy. *Journal of Consumer Research*, 51(4), 656–678.
<https://doi.org/10.1093/jcr/ucae024>
- Baggini, J. (2017). *A short history of truth: Consolations for a post-truth world*. Quercus.
- Baker, S. M., Brawley, O. W., & Marks, L. S. (2005). Effects of untreated syphilis in the negro male, 1932 to 1972: A closure comes to the Tuskegee study, 2004. *Urology*, 65(6), 1259–1262. <https://doi.org/10.1016/j.urology.2004.10.023>
- 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>
- Beard, M. (2015). *SPQR: A history of ancient Rome* (First edition). Liveright Publishing Corporation, a Division of W.W. Norton & Company.
- Bensley, D. A., & Lilienfeld, S. O. (2020). Assessing belief in unsubstantiated claims. *Scholarship of Teaching and Learning in Psychology*, 6(3), 198–211.
<https://doi.org/10.1037/stl0000218>
- Bowes, S. M., Costello, T. H., & Tasimi, A. (2023). The conspiratorial mind: A meta-analytic review of motivational and personological correlates. *Psychological Bulletin*, 149(5–6), 259–293. <https://doi.org/10.1037/bul0000392>

- 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>
- Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring Individual Differences in Generic Beliefs in Conspiracy Theories Across Cultures: Conspiracy Mentality Questionnaire. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00225>
- Buenting, J., & Taylor, J. (2010). Conspiracy Theories and Fortuitous Data. *Philosophy of the Social Sciences*, 40(4), 567–578. <https://doi.org/10.1177/0048393109350750>
- Cookson, C. (2017, January 18). LIBOR scandal: The bankers who fixed the world's most important number. *The Guardian*. <https://www.theguardian.com/business/2017/jan/18/libor-scandal-the-bankers-who-fixed-the-worlds-most-important-number>
- Darwin, H., Neave, N., & Holmes, J. (2011). Belief in conspiracy theories. The role of paranormal belief, paranoid ideation and schizotypy. *Personality and Individual Differences*, 50(8), 1289–1293. <https://doi.org/10.1016/j.paid.2011.02.027>
- Dentith, M. R. X. (2023). Some Conspiracy Theories. *Social Epistemology*, 37(4), 522–534. <https://doi.org/10.1080/02691728.2023.2173539>
- 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>

- Earnshaw, V. A., Eaton, L. A., Kalichman, S. C., Brousseau, N. M., Hill, E. C., & Fox, A. B. (2020). COVID-19 conspiracy beliefs, health behaviors, and policy support. *Translational Behavioral Medicine*, 10(4), 850–856.
- Fraser, A. (1996). *Faith and treason: The story of the Gunpowder Plot*. Doubleday.
- Gawker. (2012, December 14). *The insane Sandy Hook conspiracy theories that are already flooding Facebook and Twitter*.
- Goertzel, T. (1994). Belief in Conspiracy Theories. *Political Psychology*, 15(4), 731. <https://doi.org/10.2307/3791630>
- Goldstein, J. L. (Ed.). (1974). *Watergate: Chronology of a crisis. 2: Compiled by Janice L. Goldstein*. Congressional Quarterly.
- Grzesiak-Feldman, M. (2013). The Effect of High-Anxiety Situations on Conspiracy Thinking. *Current Psychology*, 32(1), 100–118. <https://doi.org/10.1007/s12144-013-9165-6>
- Halpern, D. (2012). *HCTA Halpern Critical Thinking Assessment* [dataset]. <https://doi.org/10.1037/t10940-000>
- Harris, K. (2018). What's Epistemically Wrong with Conspiracy Theorising? *Royal Institute of Philosophy Supplement*, 84, 235–257. <https://doi.org/10.1017/S1358246118000619>
- Hibbert, C. (1979). *The rise and fall of the house of medici*. Penguin.
- Hofstadter, R. (1964). The Paranoid Style in American Politics. *Harper's Magazine*, 229(1374), 77–86.
- Howe, L. (2024). My husband became a conspiracy theorist. Would our marriage survive? *The Guardian*. <https://www.theguardian.com/lifeandstyle/2024/nov/09/my-husband-the-conspiracy-theorist>

- 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>
- Keeley, B. (1999). Of Conspiracy Theories. *Journal of Philosophy*, 96(3), 109–126. <https://doi.org/10.2139/ssrn.1084585>
- Klein, C., Clutton, P., & Polito, V. (2018). Topic Modeling Reveals Distinct Interests within an Online Conspiracy Forum. *Frontiers in Psychology*, 9, 189. <https://doi.org/10.3389/fpsyg.2018.00189>
- Krentz, P. (1982). *The Thirty at Athens*. Cornell university press.
- Lantian, A., Bagneux, V., Delouvé, S., & Gauvrit, N. (2021). Maybe a free thinker but not a critical one: High conspiracy belief is associated with low critical thinking ability. *Applied Cognitive Psychology*, 35(3), 674–684. <https://doi.org/10.1002/acp.3790>
- Lobato, E., Mendoza, J., Sims, V., & Chin, M. (2014). Examining the Relationship Between Conspiracy Theories, Paranormal Beliefs, and Pseudoscience Acceptance Among a University Population. *Applied Cognitive Psychology*, 28(5), 617–625. <https://doi.org/10.1002/acp.3042>
- March, E., & Springer, J. (2019). Belief in conspiracy theories: The predictive role of schizotypy, Machiavellianism, and primary psychopathy. *PLOS ONE*, 14(12), e0225964. <https://doi.org/10.1371/journal.pone.0225964>
- Newman, L. H. (2020). How to talk to conspiracy theorists—And still be kind. *MIT Technology Review*. <https://www.technologyreview.com/2020/07/15/1004950/how-to-talk-to-conspiracy-theorists-and-still-be-kind/>

- Oliver, J. E., & Wood, T. J. (2014). Conspiracy Theories and the Paranoid Style(s) of Mass Opinion. *American Journal of Political Science*, 58(4), 952–966.
<https://doi.org/10.1111/ajps.12084>
- 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>
- Orosz, G., Krekó, P., Paskuj, B., Tóth-Király, I., Bőthe, B., & Roland-Lévy, C. (2016). Changing Conspiracy Beliefs through Rationality and Ridiculing. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01525>
- Oxford University Press. (2024). *Conspiracy*. <https://languages.oup.com/>
- Pennycook, G., Allan Cheyne, J., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making*, 10(6), 549–563. <https://doi.org/10.1017/S1930297500006999>
- Pirie, M. (2006). *How to win every argument: The use and abuse of logic*. Continuum.
- 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>
- Radnitz, S., & Underwood, P. (2017). Is Belief in Conspiracy Theories Pathological? A Survey Experiment on the Cognitive Roots of Extreme Suspicion. *British Journal of Political Science*, 47(1), 113–129.
<https://doi.org/10.1017/S0007123414000556>

Rosenberg, Y. (2018, April 4). Cambridge Analytica scandal fallout: How the political data firm exploited Facebook data. *The New York Times*.

<https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>

Rutjens, B. T., & Većkalov, B. (2022). Conspiracy beliefs and science rejection.

Current Opinion in Psychology, 46, 101392.

<https://doi.org/10.1016/j.copsyc.2022.101392>

Scriven, M., & Paul, R. (1987). Defining critical thinking. Paper presented at: 8th

Annual International Conference on critical thinking and Education Reform.

Retrieved from <https://web.archive.org/web/20190930032113/>.

<https://www.criticalthinking.org/pages/defining-critical-thinking/766>

Snowden, E. (2021, July 1). Why do conspiracy theories flourish? Because the truth is too hard to handle. *The Guardian*.

<https://www.theguardian.com/commentisfree/2021/jul/01/edward-snowdon-conspiracy-theories-belief-powerlessness>

Stasielowicz, L. (2022). Who believes in conspiracy theories? A meta-analysis on personality correlates. *Journal of Research in Personality*, 98, 104229.

<https://doi.org/10.1016/j.jrp.2022.104229>

Uscinski, J. E. (Ed.). (2018). *Conspiracy Theories and the People Who Believe Them* (1st ed.). Oxford University Press.

<https://doi.org/10.1093/oso/9780190844073.001.0001>

Van Prooijen, J., & Acker, M. (2015). The Influence of Control on Belief in Conspiracy Theories: Conceptual and Applied Extensions. *Applied Cognitive Psychology*, 29(5), 753–761. <https://doi.org/10.1002/acp.3161>

- Van Prooijen, J., & Douglas, K. M. (2018). Belief in conspiracy theories: Basic principles of an emerging research domain. *European Journal of Social Psychology*, 48(7), 897–908. <https://doi.org/10.1002/ejsp.2530>
- Van Prooijen, J.-W. (2019). Belief in Conspiracy Theories. In J. P. Forgas & R. F. Baumeister (Eds.), *The Social Psychology of Gullibility* (1st ed., pp. 319–332). Routledge. <https://doi.org/10.4324/9780429203787-17>
- Whitson, J. A., & Galinsky, A. D. (2008). Lacking Control Increases Illusory Pattern Perception. *Science*, 322(5898), 115–117. <https://doi.org/10.1126/science.1159845>
- Wise, J. (2001). Pfizer accused of testing new drug without ethical approval. *BMJ*, 322(7280), 194–194. <https://doi.org/10.1136/bmj.322.7280.194>
- 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>

Chapter 2: The efficacy of interventions in reducing conspiracy beliefs: a systematic review

This paper was published in *PLOS One*.

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>

Study Motivation

This systematic review explores the efficacy of current interventions in terms of addressing unfounded conspiracy beliefs. The review also examines the limitations of current intervention methods along with examining how conspiracy ideation is currently being assessed.

Author Contributions

COM conducted the searches, screened papers, and performed the narrative analysis. COM wrote the manuscript and revised it in response to feedback from CL and GM. MB assisted in screening papers as a second reviewer and provided feedback on the manuscript. All authors approved the final version of the manuscript

2.1 Abstract

Conspiracy beliefs have become a topic of increasing interest among behavioural researchers. While holding conspiracy beliefs has been associated with several detrimental social, personal, and health consequences, little research has been dedicated to systematically reviewing the methods that could reduce conspiracy beliefs. We conducted a systematic review to identify and assess interventions that have sought to counter conspiracy beliefs. Out of 25 studies (total N = 7,179), we found that while the majority of interventions were ineffective in terms of changing conspiracy beliefs, several interventions were particularly effective. Interventions that fostered an analytical mindset or taught critical thinking skills were found to be the most effective in terms of changing conspiracy beliefs. Our findings are important as we develop future research to combat conspiracy beliefs.

Keywords: conspiracy beliefs; interventions; critical thinking, analytical thinking

2.2 Introduction

Conspiracy beliefs are defined as a set of beliefs that “explain important events as secret plots by powerful and malevolent groups” (Douglas et al, 2017, p. 538). For example, various groups of individuals hold the beliefs that the 1969 moon landing was a hoax staged by the United States government, and that the cure to cancer has been discovered but is kept secret from the public (Goldacre, 2012; Plait, 2002). Holding conspiracy beliefs is often associated with negative personal, social, and health-related consequences. For example, conspiracy beliefs are associated with reluctance to receive a COVID-19 vaccine and reduced adherence to public health regulations (Allington et al., 2021; Bierwiazzonek et al., 2020, 2022; Imhoff et al., 2021; Klimiuk et al., 2021; Knobel et al., 2022), extremist and violent behavior (Bartlett & Miller, 2010; Jolley & Paterson, 2020). As much of this research shows only associations between conspiracy beliefs and negative outcomes, further research is needed to draw any conclusions regarding causality. Nevertheless, as there are possible negative implications associated with conspiracy beliefs at both personal and societal levels, there is a need for interventions to lower the likelihood that people will engage uncritically with such theories.

One of the primary characteristics of a conspiracy theory is that they are unfalsifiable. Specifically, any attempt to demonstrate that the claims made by conspiracy theories are false, could be perceived as evidence of a cover-up of the “truth”, and institutions that attempt to debunk conspiracies are perceived as accomplices attempting to conceal such conspiracies (Grimes, 2016; Lewandowsky et al., 2013). The effectiveness of counterarguments on conspiracy beliefs is mixed. Several studies have found that counterarguments could change the conspiracy beliefs of participants (Swami et al., 2013). However, other studies found that counterarguments had no effect of conspiracy attitudes (Jolley & Douglas, 2017; Stojanov, 2015). Additionally, some research has found that counterarguments in some particular circumstances can

paradoxically strengthen conspiracy beliefs (Nyhan et al., 2013, 2014; Nyhan & Reifler, 2015). As such, simply arguing against conspiracy beliefs may not be a sufficient way of changing them.

2.2.1 Conspiracy Interventions

The following section introduces and defines the various types of interventions that have been designed to influence conspiracy beliefs. It should be noted that much of the following research surrounding conspiracy interventions were designed based on evidence accumulated by studies that focus on misinformation and fake news in addition to conspiracy belief. Such papers frequently refer to misinformation, fake news, and conspiracy beliefs interchangeably when citing evidence to support the utility of a particular intervention. While the authors of this paper believe that misinformation, fake news, and conspiracy beliefs are distinct phenomena, the common characteristic of the studies in this area is that they attempt to change beliefs that persist, even when sufficient counterevidence exists.

The first category of conspiracy interventions are *informational inoculations*. The theory behind this intervention suggests that much like a vaccine protects against disease, an informational inoculation consisting of a pre-emptive debunking (Lewandowsky & van der Linden, 2021). The inoculation consists of giving participants a weakened form of the conspiracy argument along with points that refute the claims made in the argument. So, whereas a direct counterargument or debunking attempts to confront conspiracy beliefs after people have been exposed to conspiracy theories, an informational inoculation confronts conspiracy beliefs before exposure. Inoculations generally consist of two components: 1) a forewarning that the recipient of the inoculation is going to be exposed to misinformation, and 2) a pre-emptive refutation of the misinformation.

Previous experiments have shown these inoculations are effective with regards to decreasing belief in fake news and political gossip, as well as conspiracy beliefs (Banas & Miller, 2013; Lewandowsky & van der Linden, 2021; Roozenbeek & van der Linden, 2019; van der Linden et al., 2020). A meta-analysis has also shown that inoculations overall represent an effective means of challenging belief persistence, in situations where participants would otherwise be unwilling to change their beliefs despite contradictory evidence (Banas & Rains, 2010).

Another common type of intervention for changing conspiracy beliefs is *priming*. Priming interventions experimentally manipulate the psychological states of participants before they are asked about their conspiracy beliefs. For example, Swami et al [20b], found that priming participants to engage in analytical thinking significantly changed the likelihood that they would agree with conspiracy beliefs when tested. While priming interventions have shared features (i.e. they expose participants to a task or stimulus with the intention of subsequently influencing how they score on conspiracy measures), they are a heterogeneous group. Some priming interventions seek to alter one's cognition by administering tasks that increase focus and analytical thinking (Swami et al., 2014). Other priming interventions sought to influence one's affective state, for example altering their feelings of ostracism (Poon et al., 2020). Some researchers consider priming to be another form of informational inoculation, as they also consist of treatments that are administered before participants are exposed to conspiracy theories (Bonetto et al., 2018).

Other approaches attempt to use the persuasive power of stories to confront conspiracy beliefs, often referred to as *narrative persuasion* (Lazić & Žeželj, 2021). The hypothesis is that the narrative and anecdotal nature of conspiracy theories are more appealing than the detached nature of scientific evidence. These claims have been

supported by the fact that personal anecdotes both in favour and against vaccines were found to be far more persuasive than scientific evidence (Xu, 2019).

Conspiracy beliefs have been associated with a number of cognitive biases and logical fallacies such as the conjunction fallacy (Brotherton & French, 2014), jumping-to-conclusions bias (Pytlik et al., 2020), and illusionary pattern perception (van Prooijen et al., 2018). These biases have been attributed to the unfounded conclusions drawn by those who hold conspiracy beliefs. A number of the interventions similar to those listed above, such as fostering effortful thought, have been successful in reducing susceptibility to some of the cognitive biases associated with conspiracy belief (e.g. conjunction fallacy) (Scherer et al., 2017). However, it is difficult to draw conclusions regarding how these interventions change the underlying cognitive processes behind conspiracy belief, as many of these studies exclusively measure the extent to which people endorse conspiracy theories.

2.2.2 Current Study

Systematic reviews within the area of conspiracy research have typically focused on identifying and understanding the individual differences and predictors of conspiracy belief (Douglas et al., 2017; Mulukom et al., 2020). There is yet to be a comprehensive review of the efficacy of various conspiracy interventions. While one systematic review focused on examining conspiracy interventions, it was solely focused on the use of narrative persuasion in the specific case of anti-vaccine conspiracy theories (Lazić & Žeželj, 2021). As there are growing calls to effectively combat widespread conspiracy beliefs, it is pertinent that we review the evidence for existing interventions. The COVID-19 pandemic and associated conspiracy theories highlighted the potentially harmful

public health threats posed by unsubstantiated beliefs . By reviewing the evidence, we may gain a greater understanding of how to challenge conspiracy beliefs.

2.3 Methods

The protocol was written in accordance to the PRISMA-P 2015 Statement guidelines (Moher et al., 2015). PRISMA is a 27-point checklist of the preferred reporting items for a systematic review (see S1 File for PRISMA checklist). The authors pre-registered this review at PROSPERO (https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021256263).

2.3.1 Search Strategy

Three electronic databases were searched (PubMed, PsychINFO, Scopus). PubMed was chosen out of consideration for health-related publications related to interventions challenging medical conspiracy theories such as anti-vaccine narratives and Big Pharma myths. PsychINFO was chosen as we felt it contains the widest database of psychological interventions. Other psychology databases, like Web of Science's psychology subset, are less specialized in clinical and social interventions. Finally, we included Scopus in order to catch any interdisciplinary papers that would not be identified through PubMed or PsychINFO. The search strategy of this review used two search categories; conspiracy theories, and reduction/intervention. The search strings were any combination of (conspirac* OR 'unsubstantiated beliefs' OR 'implausible beliefs' OR 'unsubstantiated claims' OR 'unfounded beliefs' OR truther) AND (intervention OR reduc* OR chang* OR alter* OR experiment OR persua*). The list of keywords was determined after repeated prechecking.

With the concern of missing potentially eligible papers, no filtering was applied in the database searches (subject, publication dates) to avoid the possibility of the scope of the view becoming too narrow. There was no limitation applied to the publication dates of papers in the electronic search. Searches were conducted on 3rd June 2021.

2.3.2 Eligibility Criteria

The PICO model (Richardson et al., 1995) was used to determine the eligibility criteria for the current review. During a preliminary review of the literature, we found that very few papers had been published on the topic of conspiracy interventions. Furthermore, literature concerning conspiracy beliefs was found to use various terms to refer to such beliefs (e.g., ‘empirically unfounded beliefs’, ‘conspiratorial ideation’ etc.). Lastly, the research designs of studies researching conspiracy interventions vary considerably. As such, the authors concluded that using a strict inclusion/exclusion criteria would exclude potentially useful papers from this review, and opted to use a broad inclusion criteria to mitigate this possibility. We chose to focus exclusively on adult populations. Ethically, research involving minors often requires additional considerations, such as parental consent and heightened safeguards, which could limit the scope and feasibility of our review. Methodologically, focusing on adults enhances the generalisability of these findings, as interventions designed for this population are more commonly studied and applicable across diverse contexts.

The inclusion criteria for this review consisted of the following:

Population:

- Adults above the age of 18
- Participants without clinical/psychiatric afflictions

Intervention:

- Includes a measure of conspiracy belief (both general and specific measures were accepted).
- Experimental condition/stimulus used with the intention of influencing the measurement of conspiracy belief.

Comparator:

- A control condition either between or within-groups

Outcomes:

- Difference in conspiracy belief between experimental and control conditions

The exclusion criteria for this review consisted of the following:

Population:

- Participants under 18 years of age
- Clinical populations

Intervention:

- Correlation studies/observational measures.

Comparator:

- Studies with no control/baseline condition for the experimental intervention condition

Outcomes:

- No outcomes were excluded from this study.

We additionally excluded any papers that dealt with primarily qualitative data.

2.3.3 Data Management

References were collected from PsychInfo, PubMed, and Scopus. References were then imported into Endnote, where they were screened for possible duplicates. Once the duplicates were removed, the remaining references were exported to Covidence.

2.3.4 Data review and screening

The screening process was conducted using the Covidence systematic review management online software. The screening process consisted of two waves., Title and Abstract screening, and Full-text screening, in accordance to PRISMA standards (Moher et al., 2009). Two reviewers (COM & MB) screened the papers for both the Title/Abstract, and Full Text data screenings. Conflicting votes were resolved by a third reviewer (GM) who made the final decision regarding the inclusion of such papers. The interrater reliability was assessed using Cohen's Kappa (k). The level of agreement was found to be fair ($k = 0.40$) for title and abstract screening, and substantial for full-text screening ($k = 0.62$). Initially, 16 papers passed the full-text screening. However, due to the broad nature of the inclusion criteria, a number of studies were deemed eligible but did not have sufficient data for analysis. As such, a second examination of the remaining papers was conducted to see whether each paper had sufficient data to analyse. Studies that reported neither sufficient data to calculate the Cohen's d , nor statistical significance, were excluded from the analysis. One paper (Einstein & Glick, 2015) was excluded as the paper reported its numerical findings solely in the form of graphs, making it difficult to extract exact values. Two papers were excluded as no relevant information could be obtained for our analysis. After the papers were screened for sufficient data, 13 articles in total were included in the final analysis. The PRISMA diagram (Figure 2.1) shows each stage of this process.

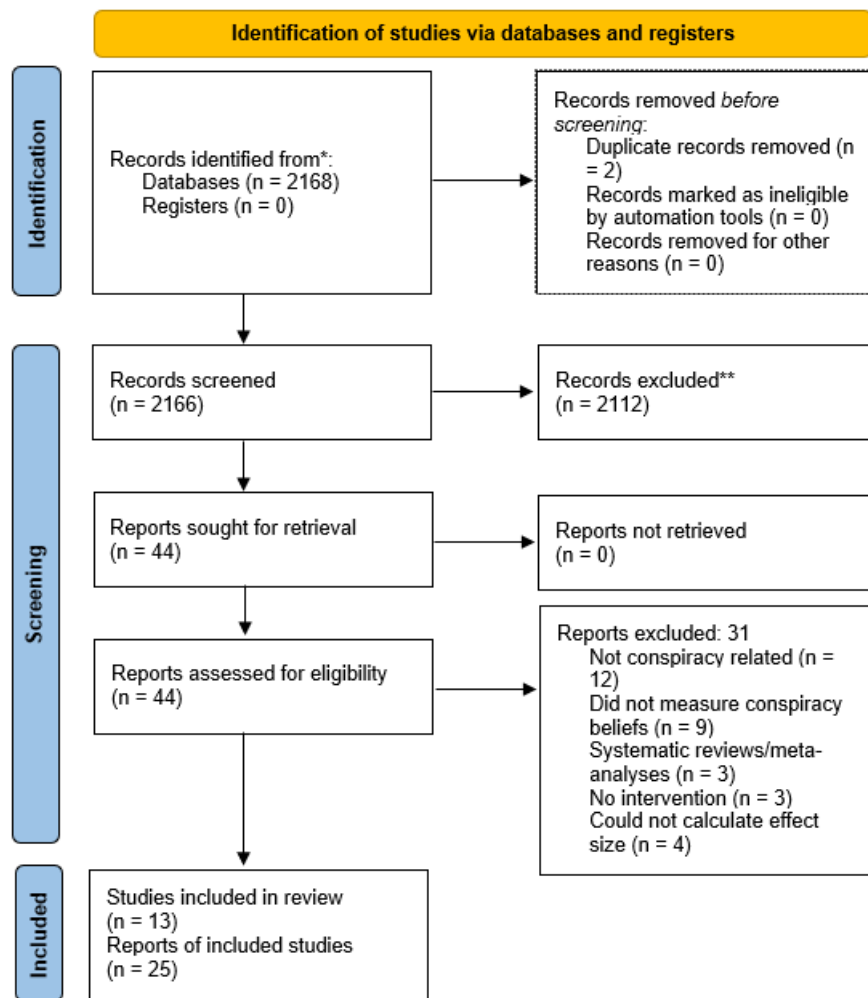


Figure 2.1: PRISMA Flow Diagram

2.3.5 Data extraction and analysis

A number of papers included in the review consisted of more than a single study, leading to a total of 24 studies being included in the final review. The following data was extracted from the studies: (1) Publication characteristics such as author names, journal, and publication dates; (2) Demographic characteristics, such as mean age, country, target population; and (3) Intervention characteristics – type of intervention used, conspiracy measure used, whether the conspiracy measure was specific of generic.

From the 25 studies included in our review, we extracted 48 intervention comparisons (k) in total (see Table 1.). The outcome measure of interest was the Cohen's d effect size. For all papers that used paired group designs, the Cohen's d had already been reported. Where the Cohen's d was not reported in a study, we calculated the d value for the independent group designs using the mean difference and the pooled standard deviation (Murphy et al., 2014). The Cohen's d was calculated for each comparison using Microsoft Excel. Where there was insufficient data to calculate the d value, statistical significance was used to indicate whether the intervention successfully induced any change in belief. Quality assessment was conducted using the Risk of Bias Tool 2 (Sterne et al., 2019, see S3 Figure). We selected the Risk of Bias 2 (RoB 2) tool due to its comprehensive and well-documented framework, which ensures a structured and transparent assessment of study quality. While RoB 2 is primarily used in medical research, its rigorous evaluation criteria, particularly for randomized controlled trials, are equally applicable to psychological interventions. Compared to other bias assessment tools commonly used in psychology research, such as the Jadad Scale or the PEDro Scale, RoB 2 provides a more detailed assessment of potential biases, including deviations from intended interventions, missing outcome data, and selective reporting. Given the novelty of this analysis, applying RoB 2 allows for a more robust and interdisciplinary evaluation, aligning psychological research with high methodological standards used in health sciences. To avoid repetition, the Risk of Bias was applied to the 25 studies as opposed to each comparator, as each comparator was subject to same randomisation process and measurement within their respective studies.

We conducted a narrative synthesis rather than a meta-analysis or GRADE assessment due to the considerable heterogeneity among the included studies in terms of intervention types, outcome measures, and methodologies. Additionally, as this is the first

systematic review to examine interventions targeting conspiracy beliefs, there was no prior framework or standardised approach to build upon, further supporting the appropriateness of a narrative approach

2.4 Results

2.4.1 Characteristics of Studies

Out of the 13 reports included, the majority were conducted in English-speaking countries; US (3), UK (3). The other studies were conducted in France (2), Hungary (1), Macedonia (1), Hong Kong (1), and Belgium (1). It should be noted that since the majority of studies were conducted online, a number of the studies did not use samples from their countries of origin. Out of the three studies conducted in the UK, two consisted of samples from the United States. Furthermore, the lone study conducted in Hong Kong also used participants from the United States. All studies were published between 2013 and 2020. As highlighted by the Risk of Bias assessment (see S3 Fig), most studies were not preregistered. All papers were peer-reviewed and had English language versions available.

This review aimed to identify the various interventions that have been studied as a means of challenging conspiracy beliefs. We identified a number of different intervention types within the review. The first type of interventions were informational inoculations. These inoculation interventions had two different variations; one drew attention to the factual inaccuracies of conspiracy theories, and one that described logical fallacies of thought process underpinning conspiracy beliefs (Banas & Miller, 2013; Jolley & Douglas, 2017). The second type of intervention identified was priming. As with inoculation interventions, priming interventions also had several variations. One variation used priming as means of manipulating participants sense of control (Stojanov et al., 2020). These control priming interventions were typically coupled with a second priming

task to examine how feelings of control interacted with other variables. Another variation of priming interventions consisted of inducing an analytical mindset in participants by having them complete certain tasks that required mental acuity (e.g., reading difficult-to-read fonts) (Swami et al., 2014). The third type of intervention identified were counterargument interventions. These interventions took the approach of providing counterarguments to conspiracy beliefs. A number of variations included arguments that appealed to the participants sense of empathy by describing the negative consequences conspiracy beliefs can have (Orosz et al., 2016). Another attempted to argue against conspiracy beliefs by ridiculing those who believe in them (Orosz et al., 2016). Finally, a number of standalone interventions were found in our sample, which did not fit into any of the intervention types described above, such as an educational intervention and a labeling intervention. An overview of the characteristics of each study in the review is provided in S2 Table.

2.4.2 Efficacy of Interventions

2.4.2.1 Overview of included interventions

A complete overview of the 48 comparisons analysed within this review, along with their the effect sizes and *p*-values is provided in Table 2.1. All of the comparisons analysed in this review used self-reported measures of conspiracy belief. Conspiracy belief measures fall into one of two categories; general measures and specific measures. General measures measure an individual's general disposition to conspiracy ideation based on the extent in which they endorse conspiratorial themes (e.g., that certain events are the result of small malicious group who control the world) (Brotherton et al., 2013). Specific measures assess the extent in which people endorse specific conspiracy theories, such as conspiracy theory suggesting that the September 11 attacks were orchestrated by the US government (Swami et al., 2010). As these two types of measures vary, it should be noted that the

implications of decreasing scores on a specific measure compared to a general measure may differ (Imhoff et al., 2022; Sutton & Douglas, 2020). Furthermore, the correlates for both types of measures differ. An intervention that changes how one endorses anti-vaccination conspiracy theories may not necessarily affect how they score on a measure of general conspiracy ideation. As noted previously (Imhoff et al., 2022), general measures of conspiracy beliefs are more stable and less malleable than their specific measure counterparts.

Approximately 60% of studies used some form of measure that examined generalised conspiracy belief and ideation ($k = 28$), while the rest used measures that assessed belief in specific conspiracy theories ($k = 6$). Only a small number of studies utilised a within-group design ($k = 5$) meaning the majority of comparisons used cross-sectional designs. Overall, 21 (~40%) interventions successfully changed conspiracy beliefs when compared to control, baseline, or neutral groups. However, only a small number of these interventions demonstrated large effect sizes ($k = 3$). Furthermore, medium effects were only found with a handful of interventions ($k = 2$). The remaining studies were found to have only small, or very small effects.

Table 2.1 Effectiveness of Interventions in Included Studies

Author	Intervention	Design	Grouping	Aware of Intervention	Conspiracy measure	Type of Measure	<i>p-value</i>	<i>Cohen's d</i>
(Whitson et al., 2019)	Prevention Regulatory Focus	Experimental	Between-groups	Unclear	Cooperative Congressional Election Study (Ansolabeher, 2013)	General	0.132	0.132
(Whitson et al., 2019)	Promotion Regulatory Focus	Experimental	Between-groups	Unclear	Cooperative Congressional Election Study (Ansolabeher, 2013)	General	0.014*	0.37
(Whitson et al., 2019)	Increasing Perceived Control (Promotion)	Experimental	Between-groups	Unclear	Cooperative Congressional Election Study (Ansolabeher, 2013)	Specific	0.017*	0.49
(Whitson et al., 2019)	Increasing Perceived Control (Prevention)	Experimental	Between-groups	Unclear	Cooperative Congressional Election Study (Ansolabeher, 2013)	Specific	0.534	0.11
(Adam-Troian et al., 2019)	Rationality Priming	Experimental	Between-groups	Yes	Conspiracist Mentality Questionnaire (Lantian, Muller, Nurra, & Douglas, 2016).	General	NR	-0.016
(Orosz et al., 2016)	Ridiculing Beliefs	Experimental	Within-groups	Yes	Conspiracy Assessment Tool (CAT)	General	0.001**	0.11
(Orosz et al., 2016)	Rational counterarguments	Experimental	Within-groups	Yes	Conspiracy Assessment Tool (CAT)	General	0.001**	0.13
(Orosz et al., 2016)	Empathetic counterarguments	Experimental	Within-groups	Yes	Conspiracy Assessment Tool (CAT)	General	0.09	0.05
(Orosz et al., 2016)	Ridiculing Beliefs	Experimental	Between-groups	Yes	Conspiracy Assessment Tool (CAT)	General	0.059	0.2
(Orosz et al., 2016)	Rational counterarguments	Experimental	Between-groups	Yes	Conspiracy Assessment Tool (CAT)	General	0.011*	0.27
(Orosz et al., 2016)	Empathetic counterarguments	Experimental	Between-groups	Yes	Conspiracy Assessment Tool (CAT)	General	0.333	0.1

(Banas & Miller, 2013)	Fact-based Inoculation	Experimental	Between-groups	No	Generic attitudes (Burgoon, Cohen, Miller, & Montgomery, 1978)	Specific	<0.001	1.311
------------------------	------------------------	--------------	----------------	----	--	----------	--------	-------

Table 2.1 continued

Author	Intervention	Design	Grouping	Aware of Intervention	Conspiracy measure	Type of Measure	<i>p</i> -value	Cohen's <i>d</i>
(Banas & Miller, 2013)	Logic-based Inoculation	Experimental	Between-groups	No	Generic attitudes (Burgoon, Cohen, Miller, & Montgomery, 1978)	Specific	<0.001	0.909
(Banas & Miller, 2013)	Metainoculation (fact based) ²	Experimental	Between-groups	No	Generic attitudes (Burgoon, Cohen, Miller, & Montgomery, 1978)	Specific	NR	0.58
(Banas & Miller, 2013)	Metainoculation (logic based) ²	Experimental	Between-groups	No	Generic attitudes (Burgoon, Cohen, Miller, & Montgomery, 1978)	Specific	NR	0.58
(Swami et al., 2014)	Analytical priming	Experimental	Between-groups	Unclear	Belief in Conspiracy Theories Inventory (BCTI Swami et al., 2010, 2011).	General	0.018*	0.46
(Swami et al., 2014)	Analytical priming	Experimental	Between-groups	Unclear	Belief in Conspiracy Theories Inventory (BCTI Swami et al., 2010, 2011).	General	0.001**	0.49
(Swami et al., 2014)	Analytical priming (Font)	Experimental	Between-groups	Unclear	7/7 bombings questionnaire	Specific	0.046*	0.34
(Swami et al., 2014)	Analytical priming (Font)	Experimental	Between-groups	No	Generic Conspiracist Beliefs scale (GCBS; Brotherton, French, & Pickering, 2013)	General	0.014*	0.42
(Jolley & Douglas, 2017)	Pro-conspiracy arguments ¹	Quasi-Experimental	Between-groups	Yes	Anti-vaccine Conspiracy Beliefs Measure	Specific	0.001**	-0.66
(Jolley & Douglas, 2017)	Anti-conspiracy arguments	Quasi-Experimental	Between-groups	Yes	Anti-vaccine Conspiracy Beliefs Measure	Specific	0.017*	0.42
(Jolley & Douglas, 2017)	Anti-conspiracy/conspiracy arguments	Quasi-Experimental	Between-groups	Yes	Anti-vaccine Conspiracy Beliefs Measure	Specific	0.047*	-0.1
(Jolley & Douglas, 2017)	Conspiracy/anti-conspiracy arguments ¹	Quasi-Experimental	Between-groups	Yes	Anti-vaccine Conspiracy Beliefs Measure	Specific	0.263	-0.39
(Jolley & Douglas, 2017)	Anti-conspiracy/conspiracy arguments	Quasi-Experimental	Between-groups	Yes	Anti-vaccine Conspiracy Beliefs Measure	Specific	0.1	0.62

Table 2.1 continued

Author	Intervention	Design	Grouping	Aware of Intervention	Conspiracy measure	Type of Measure	<i>p</i> -value	<i>Cohen's d</i>
(Jolley & Douglas, 2017)	Conspiracy/anti-conspiracy arguments ¹	Quasi-Experimental	Between-groups	Yes	Anti-vaccine Conspiracy Beliefs Measure	Specific	0.001**	0.31
(Wood, 2016)	Conspiracy labelling	Experimental	Between-groups	No	Generic Conspiracist Beliefs scale (GCBS; Brotherton, French, & Pickering, 2013)	General	>0.05	NR
(Wood, 2016)	Conspiracy labelling	Experimental	Between-groups	No	Historical items	General	>0.05	-0.1
(Wood, 2016)	Conspiracy labelling	Experimental	Between-groups	No	Adaptation of Wood et al, 2012	Specific	>0.05	0.07
(Bonetto et al., 2018)	Priming Resistance to Persuasion	Experimental	Between-groups	No	Generic Conspiracist Beliefs scale (GCBS; Brotherton, French, & Pickering, 2013)	General	0.012*	0.56
(Bonetto et al., 2018)	Priming Resistance to Persuasion	Experimental	Between-groups	No	Generic Conspiracist Beliefs scale (GCBS; Brotherton, French, & Pickering, 2013)	General	0.029*	0.3
(Bonetto et al., 2018)	Priming Resistance to Persuasion	Experimental	Between-groups	No	Generic Conspiracist Beliefs scale (GCBS; Brotherton, French, & Pickering, 2013)	General	0.004*	0.37
(Stojanov, 2015)	Debunking	Experimental	Between-groups	Yes	Generic Conspiracist Beliefs scale (GCBS; Brotherton, French, & Pickering, 2013)	General	0.07	0.29
(Stojanov, 2015)	Debunking, motives, fallacy	Experimental	Between-groups	Yes	Generic Conspiracist Beliefs scale (GCBS; Brotherton, French, & Pickering, 2013)	General	<0.05*	0.37
(Stojanov et al., 2020)	High control priming	Experimental	Between-groups	No	Conspiracy Theory Ideation subscale of Conspiracy Mentality Scale (Stojanov & Halberstadt)	General	NR	0.11
(Stojanov et al., 2020)	Low control priming ¹	Experimental	Between-groups	No	Conspiracy Theory Ideation subscale of Conspiracy Mentality Scale (Stojanov & Halberstadt)	General	0.07	0.27
(Stojanov et al., 2020)	High control priming	Experimental	Between-groups	Unclear	Conspiracy Theory Ideation subscale of Conspiracy Mentality Scale (Stojanov & Halberstadt)	General	>0.05	-0.13

Table 2.1 continued

Author	Intervention	Design	Grouping	Aware of Intervention	Conspiracy measure	Type of Measure	<i>p</i> -value	<i>Cohen's d</i>
(Stojanov et al., 2020)	Low control priming ¹	Experimental	Between-groups	Unclear	Conspiracy Theory Ideation subscale of Conspiracy Mentality Scale (Stojanov & Halberstadt)	General	>0.05	-0.15
(Stojanov et al., 2020)	Low control priming ¹	Experimental	Between-groups	No	Conspiracy Theory Ideation subscale of Conspiracy Mentality Scale (Stojanov & Halberstadt)	General	0.35	-0.13
(Stojanov et al., 2020)	Low control priming ¹	Experimental	Between-groups	No	Belief in Conspiracy Theories Inventory (BCTI Swami et al., 2010, 2011).	Specific	0.88	0.02
(Poon et al., 2020)	Ostracism priming ¹	Experimental	Between-groups	No	Conspiracy beliefs of 14 political events	Specific	0.02*	NR
(Poon et al., 2020)	Pain priming ¹	Experimental	Between-groups	No	Conspiracy beliefs of 14 political events	Specific	0.78	NR
(Poon et al., 2020)	Ostracism priming ¹	Experimental	Between-groups	No	Conspiracy beliefs of 6 political events	Specific	0.03*	NR
(Poon et al., 2020)	Ostracism priming/self-affirmation ³	Experimental	Between-groups	No	10-item version of the Generic Conspiracist Beliefs scale (GCBS; Brotherton, French, & Pickering, 2013),	General	0.001**	NR
(Poon et al., 2020)	Ostracism priming/no affirmation ¹	Experimental	Between-groups	No	10-item version of the Generic Conspiracist Beliefs scale (GCBS; Brotherton, French, & Pickering, 2013),	General	0.89	NR
(Nera et al., 2018)	Narrative persuasion	Experimental	Between-groups	Unclear	Self-made 2-item measure	General	0.086	NR
(Nera et al., 2018)	Narrative persuasion	Experimental	Between-groups	Unclear	Conspiracist Mentality Questionnaire (Lantian, Muller, Nurra, & Douglas, 2016).	General	0.47	NR
(Dyer & Hall, 2019)	Pseudoscience class	Experimental	Within-groups	Unclear	Conspiracy belief subscale of the Paranormal Beliefs Scale (Tobacyk, 2004).	Specific	< 0.0001**	1.07
(Dyer & Hall, 2019)	Research Methods class	Experimental	Within-groups	Unclear	Conspiracy belief subscale of the Paranormal Beliefs Scale (Tobacyk, 2004).	Specific	0.14	NR

NR = Not Reported; Design = the experimental design used in the intervention; Experimental = the study used two independent groups; Quasi-experimental = the study used the same participants for both conditions; Aware of Intervention = whether participants were aware of the true nature of the experiment; Yes =

participants were aware of the interventions true intentions; No = the true nature of the interventions was not revealed to participants; Unclear = the authors do not establish whether the participants were aware or unaware of the intervention.; Type of Measure = whether the conspiracy measure used measured broad conspiracy ideation, or examined a specific set of beliefs; General = the scale used measured general conspiracy ideation; Specific = the scale measured belief a specific set of conspiracy theories; * indicates $p < 0.05$; ** indicates $p < 0.01$.¹ = these interventions were designed to increase conspiracy beliefs, as a comparison for the other interventions. ² = these metainoculations were designed to nullify the effects of the inoculation interventions used in prior studies , not to reduce conspiracy beliefs. ³ = these self-affirmations were designed to nullify the effects of ostracism on increasing conspiracy beliefs.

2.4.3 Interventions aimed at reducing conspiracy beliefs

The effect sizes of interventions aimed at reducing conspiracy beliefs are reported in Figure 2.2. Four interventions aimed at reducing conspiracy beliefs were omitted from the graph as they only reported statistical significance, and the Cohen's d could not be calculated. These interventions were a single conspiracy labelling intervention ($k = 1$), narrative persuasion interventions ($k = 2$), and a research methods module ($k = 1$). This section also omits any interventions that aimed to increase conspiracy beliefs.

Approximately half of the examined interventions consisted of *priming* based tasks ($k = 15$). The majority of these interventions demonstrated a significant change in conspiracy beliefs ($k = 10$). These effects were all either small or very small. Participants who were primed to be less susceptible to persuasion tactics showed significantly lower conspiracy beliefs when compared to controls among three experimental comparisons. These effects were shown to range from small to medium ($d = 0.3-0.56$). Interventions that primed participants to engage in analytical thinking resulted in primed participants having lower conspiracy belief than controls ($k = 4$). However, the effects of these differences were small. Other priming interventions focused on manipulating participants' sense of control. Priming participants to feel like they had a higher sense of control had mixed results, either increasing or decreasing conspiracy beliefs with very small effects ($d = -0.13 - 0.11$) (Stojanov et al., 2020).

About a sixth of interventions in the analysis used inoculation methods ($k = 5$). All were successful at reducing conspiracy beliefs, relative to controls, all with either medium or large effects. Inoculations that identified the factual inaccuracies of conspiracy beliefs were found to be the most effective of all the interventions in the review ($d = 1.3$) (Banas & Miller, 2013). Jolley and Douglas (Jolley & Douglas, 2017) only reported medium effects for inoculation methods in terms of reducing conspiracy belief measures. Inoculations that demonstrated the

logical fallacies of conspiracy beliefs were found to be the second most effective intervention ($d = 0.9$) (Banas & Miller, 2013).

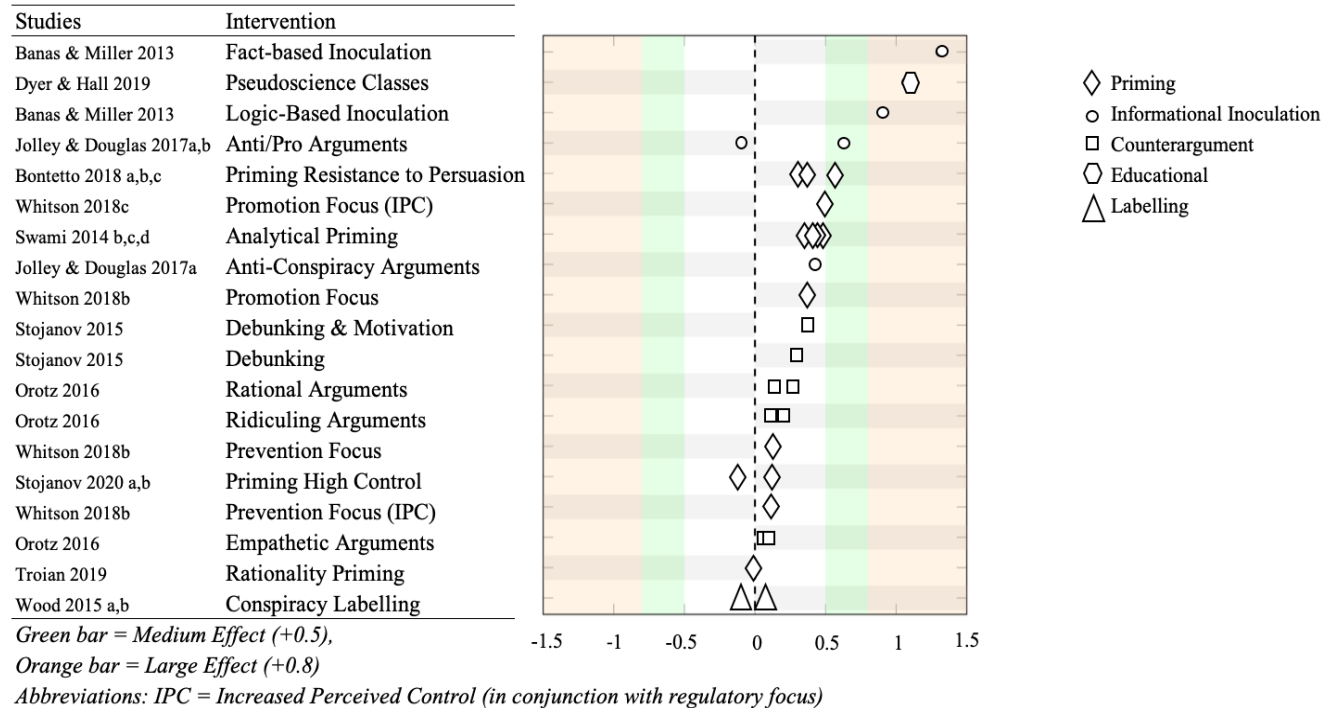


Figure 2.2: Cohen's d effect sizes interventions aimed at reducing conspiracy beliefs

Note: This figure has been adapted from (Marshall & Linehan, 2021).

In contrast to inoculations, traditional counterarguments only produced small and very small effects. Rational counterarguments that described the factual inaccuracies of conspiracy theories were found to have only very small to small effects ($d = 0.13 - 0.27$) (Orosz et al., 2016). Stojanov (Stojanov, 2015), reported similar results, with counterarguments only producing small effects in terms of changing conspiracy beliefs ($d = 0.29$). These counterarguments were found to be slightly more effective, when they included an extra component that described to participants both the motives of conspiracy theories, and the logical fallacies in the conspiratorial thinking ($d = 0.37$). Counterarguments that appealed to participant's sense of empathy, outlining the damages that can result from conspiracy beliefs

was found to have very small effects ($d = 0.05-0.1$), thus being one of the least effective interventions in the sample (Orosz et al., 2016). Finally, counterarguments that attempted to ridicule those who held conspiracy beliefs also produced only very small effects in terms of reducing conspiracy beliefs ($d = 0.11-0.2$) (Orosz et al., 2016).

We also found a number of standalone measures that utilised novel interventions to confront conspiracy belief. One study used an in-person 3-month educational course that aimed at explicitly teaching students how to differentiate between proper scientific and pseudoscientific practices. At the end of the term, the pseudoscience class reported significantly lower conspiracy belief in comparison to a typical university research methods class (Dyer & Hall, 2019). The effect was also found to be large ($d = 1.07$). Furthermore, this was one of the few studies that was within-subjects. It was the sole study to use a longitudinal design (3-month interval between pre and post measures). Another study found that labelling conspiracy statements as “conspiracy theories” in comparison to “ideas”, or “political scandals”, had no effect in changing belief in conspiracy theories ($p = >0.05$, $d = -0.1 - 0.07$) (Wood, 2016).

A handful of studies reported no Cohen’s d effects, nor provided sufficient information to calculate them. As per the protocol, in this circumstance we reported the statistical significance of these interventions. One study investigating the effects of narrative persuasion on conspiracy belief was found to have no significant effects (Nera et al., 2018).

Finally, our analysis revealed a number of “backfiring” interventions; those that aimed at reducing conspiracy belief but paradoxically increased conspiracy beliefs. While labeling conspiracy theories did not produce any significant effects, one study found that it had a small effect in increasing conspiracy beliefs ($d = -0.1$). Likewise, priming participants to feel like they had a higher sense of control increased conspiracy beliefs in one study ($d = -0.13$). Anti-conspiracy inoculations followed by pro-conspiracy arguments, while one of the most effective

interventions found in this review at reducing conspiracy beliefs, also increased conspiracy beliefs in one study albeit with very a small effect ($d = -0.1$).

2.4.4 Interventions aimed at increasing conspiracy beliefs

In addition to the interventions shown in Figure 2.2, which were all designed to reduce conspiracy belief, the review also found a number of interventions that were not intended to reduce conspiracy belief (and are therefore not shown in Figure 2.2). Most of these interventions were used as a comparison to interventions within the same study intending to reduce conspiracy beliefs. One study used informational inoculations, where participants were presented with arguments supporting conspiracy beliefs before they were shown conspiratorial content (Jolley & Douglas, 2017). This interventions was found to have a medium effect in increasing conspiracy beliefs ($d = -0.66$). Within the same article, two interventions similarly used pro-conspiracy inoculations but were then followed by anti-conspiracy arguments to test whether the order in which these inoculation were presented would influence their effects. The results for these two interventions were mixed ranging from a small effect in increasing conspiracy beliefs and a small effect in decreasing conspiracy beliefs ($d = -0.39 - 0.31$). Another study primed participants to feel lower levels of perceived control with the intention of increasing conspiracy beliefs ($k = 4$). Two low control priming interventions were found to have very small to very small effects in increasing conspiracy beliefs ($d = -0.13 - -0.15$). However, the remaining two low control priming interventions had small to very small effects in decreasing conspiracy beliefs ($d = 0.02 - 0.27$). Finally, one study tested the effects of meta-inoculations, where researchers told participants what inoculations were, with the intention of investigating whether the initial effects of informational inoculations could be negated. Meta-inoculations were another form of inoculations that were found to be reasonably effective in changing conspiracy beliefs, with medium effects, and thus were able to moderately negate the effects of anti-conspiracy inoculations ($d = 0.58$). Another study found that when participants

were primed to feel ostracised, they were found to be more likely to hold conspiracy beliefs than control groups ($p = 0.02$) (Poon et al., 2020). However, it was also reported that when participants who were in the ostracism condition were instructed to think of values that were important to them, this was shown to significantly negate the effects of the ostracism priming (Poon et al., 2020). The study reported no effect sizes for these interventions.

2.5 Discussion

The aim of this review was to assess how researchers are currently developing interventions to reduce susceptibility to conspiracy beliefs. We found that only half of the interventions examined induced any change in the degree in which participants agreed with conspiracy statements. Furthermore, only a handful of interventions produced medium to large effects. These findings suggest that most existing conspiracy interventions are ineffective in terms of changing conspiracy beliefs. This finding is in line with previous research that suggests that conspiratorial thinking may be underpinned by inherent susceptibilities (Enders et al., 2021) and may not be easily refuted with standard counterarguments (Lewandowsky et al., 2013; Nyhan et al., 2013; Nyhan & Reifler, 2015).

The few successful interventions in our sample shared a number of common characteristics. The majority of these interventions consisted of treatments that were given *before* participants were exposed to conspiracy statements. The first category consisted informational inoculations, of which the two interventions with the largest observed effect sizes belong. These findings suggest that the most effective conspiracy interventions are those take place before participants have been exposed to conspiracy beliefs. Pre-emptively refuting the inaccuracies of conspiracy beliefs was much more effective than traditional counterarguments that were given after participants had encountered conspiracy theories. Traditional counterarguments and debunking strategies in comparison only produced very small effects,

whereas inoculation strategies ranged from medium to large effects. However, one pitfall of inoculation-based interventions is that their ability to reduce conspiracy belief was easily counteracted. For participants who were warned about the inoculation interventions, and told there would be an attempt to change their conspiracy beliefs, the effects of the inoculation treatment were rendered negligible (Banas & Miller, 2013). These warnings were referred to as *metainoculations*. All that was required to make these interventions useless was to warn participants about them beforehand. In essence, the effectiveness of inoculation strategies can be used to counteract their own effects. The effects of these meta-inoculations were found to be in the medium range.

The second category of successful interventions consisted of priming interventions, particularly those that attempted to instill critical thinking skills in their participants. Swami et al (Swami et al., 2014), found that by having participants evaluate pieces text in fonts that were difficult to read, they were more likely to engage in analytical thinking. Other research has shown that participants that were primed to feel more in control had lower conspiracy beliefs when compared to the baseline group. Recent research has also found that participants who were primed to have a higher resistance to persuasion were shown to be less susceptible to conspiracy beliefs than those in control group (Bonetto et al., 2018). These findings are noteworthy when considering recent research regarding how people engage with conspiracy theories. Previous research has indicated that conspiracy beliefs are the product of intuitive, emotional thinking and that deliberation and focused analysis is associated with lower susceptibility to conspiracy beliefs (Bago et al., 2022; Forgas & Baumeister, 2019). As such, fostering critical and analytical thinking in participants through priming might possibility motivate them to push past superficial evaluations, and examine the content of conspiracy beliefs. While the possibility of priming members of the public to engage with conspiracy beliefs more critically may be promising, the feasibility of such a priming intervention outside

a controlled lab environment has yet to be ascertained. Whether one priming session would be sufficient to foster critical evaluation of conspiracy theories has not yet been demonstrated. It should also be noted that all the analytical priming interventions included in this review originate from one 4-paper study (Swami et al., 2014). More independent research may be needed to replicate the results of this intervention. Furthermore, there are some drawbacks associated with the manipulation checks used during these interventions. Participants' susceptibility to the Moses Illusion was used as a manipulation check to assess whether reading difficult-to-read fonts successfully elicited analytical processing. However, previous research has failed to replicate whether reading in a difficult font has a significant effect on susceptibility to the Moses Illusion (Janouskova et al., 2022).

Among the interventions which reported medium-to-large effects, there was one intervention which was characteristically different to the rest. This intervention made use of a university module to teach students about the differentiation between good scientific practice and pseudoscience (Dyer & Hall, 2019). After taking the module, students endorsed far fewer conspiracy theories. This was the sole study in this review with a large effect size that did not administer the intervention before participants were exposed to conspiracy statements.

It should be noted that the three intervention types that were most effective within this review are difficult to administer in a real-world setting. Informational inoculations need to be administered to participants before they encounter a conspiracy which would be difficult to maintain in practice. Priming interventions are difficult to maintain as people encounter a variety of new sources every day. Finally, educational interventions require a substantial amount of effort and commitment from both the educator and participants. None of these interventions will provide an easy solution to conspiracy beliefs practically speaking. As such, future research should examine whether the mechanisms of long-form interventions such as the three-month university pseudoscience module (Dyer & Hall, 2019), can be condensed into

a short-form counterpart. Research in this area may help bridge the gap between effective but unscalable interventions and interventions that can realistically be implemented in everyday life. In order to efficiently counter misinformation spread by conspiracy beliefs, future interventions will need to focus on developing a solution that will be more easily implemented.

2.5.1 Limitations of studies examined in the review

There were a number of limitations found in the studies examined in this review. The main limitations identified are as follows: focusing only on one form of research design, lack of cross-culture sampling, and issues with the measurement of conspiracy beliefs. Firstly, one of the potential limitations of the findings of this review is the approach taken to research design in the studies included in our sample. The majority of the studies in this sample used cross-sectional designs. Where within-groups comparisons were used, it was over a relatively small interval. As such, it cannot be ascertained whether the beneficial effects of the interventions found in this study will remain consistent after a period of time. The interventions only examined short-term effects. Therefore, it is difficult to conclude whether even the most effective intervention found in this review will be of much use in tackling conspiracy beliefs outside the lab. This limitation also applies to interventions that intended to temporarily alter psychological states of participants (e.g., sense of control). As these studies aim at only temporarily changing the psychological states of their participants, it's to be expected that any effects observed would only last until the effects of experimental manipulation wore off. Furthermore, as identified by Swire-Thompson et al (Swire-Thompson et al., 2020), one of the main issues with cross-sectional studies that exclusively use post-intervention measures of belief is that there is no way of establishing whether the groups tested were matched at baseline. The conspiracy beliefs of either the control group or one of the experimental conditions may already have been larger than the other groups before the intervention.

The second limitation we identified is the lack of cultural diversity within the samples tested. Ideally, studies that wish to identify conspiracy interventions that can be widely used should use cross-cultural studies to verify their effects are only observed a specific cultural context. The studies in this review were predominantly conducted with western samples, with the majority of studies coming from either North America or the United Kingdom. While some studies, such as Poon et al (Poon et al., 2020), were conducted outside of the West, many of these studies recruited American samples using Amazon's Mechanical Turk. It should also be noted that the 24 studies included in this review only represent 10 groups of authors. Furthermore, half of these studies originate from three research groups, which may suggest a possible source of bias.

Finally, many of the studies used novel, custom measures of conspiracy ideation to measure conspiracy belief. The lack of consistency in measurement makes it difficult to generalize positive findings of these studies. Many studies only examined belief in specific conspiracy theories. Such findings can be difficult to generalize, as an intervention that is successful in changing anti-vaccine conspiracy beliefs may not be particularly effective when used to confront 9/11 conspiracy beliefs. Furthermore, a number of the conspiracy belief measures that were used within these studies were also self-designed, even when general conspiracy belief was being measured. Finally, exclusively using measures of conspiracy belief makes it difficult to draw any conclusions regarding how these interventions affect the underlying cognitive process behind conspiracy beliefs. A focus on more consistent usage of valid and reliable conspiracy measures would be desirable in future studies.

2.5.2 Limitations of the current systematic review

There were also some possible limitations in the systematic review itself. The outcome of interest for this review was the Cohen's *d* effect size, meaning that this review focused

mainly on pairwise comparisons. While this allowed for consistency when comparisons were made between the interventions, it also came with some drawbacks. For example, a number of studies examined additional variables that may have mediated the changes induced by the interventions. However, as this review focused on pairwise comparisons, some of these details may not have been examined.

2.5.3 Implications for research / practice / design

This review identified a number of directions for future research. Firstly, we recommend that future studies on conspiracy belief interventions focus on using longitudinal designs. Understanding the effects of these interventions beyond their immediate changes is important in terms of identifying how effective they are at challenging conspiracy beliefs. Second, we recommend that future studies be conducted with non-Western populations, with more of a focus on cross-cultural sampling. Conspiracy beliefs may often be specific to cultural context, as such, we recommend that research be expanded beyond British and American samples.

In terms of practical implications for challenging conspiracy beliefs, we recommend that those with an interest in reducing the misinformation that conspiracy theories spread should do the following:

1. *Avoid appealing to emotions and affect:* Interventions that manipulated the emotional state of participants or appealed to feelings of empathy had small effects in terms of changing conspiracy beliefs.

2. *Counterarguments are not effective:* Counterarguments against specific conspiracy beliefs that are given after participants have been exposed to a conspiracy theory tend not to be particularly effective.
3. *Prevention is the best cure:* Interventions that provided counterarguments for conspiracy theories were most effective when the counterargument came before the participants were exposed to the particular conspiracy theories that the study focused on. The findings suggest it is more difficult to challenge conspiracy beliefs once participants have started to believe in them. If participants have been taught why certain conspiracy theories are implausible before they have been exposed to conspiratorial media they are much more resistant to conspiracy beliefs.
4. *An analytical mindset and critical thinking skills are the most effective means of challenging conspiracy beliefs:* Participants who were primed to have an analytical mindset were less likely to have conspiracy beliefs than controls. Furthermore, when interventions moved beyond putting participants in an analytical mindset, and actually explicitly taught them how to evaluate conspiracy beliefs using specific critical thinking skills, they were much less likely to have conspiracy beliefs.

2.6 Conclusion

In conclusion, this review sought to investigate how effective current interventions are at changing conspiracy beliefs. We found that overall, the majority of current conspiracy interventions are ineffective in terms of changing conspiracy beliefs. Despite this, we have identified several promising interventions that may be fruitful to pursue in future studies. We propose that a focus on inoculation-based and critical thinking interventions will bear more promising results for future research, though further efforts are needed to reduce participant burden and more easily implement these interventions in the real world.

Declaration of Conflicting Interests

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The data collected and analysed in this study can be found on the Open-Science Framework, [https://osf.io/6yjt3/?view_only=ccfb3cd302524505ab069bdbdb3d4b10].

Acknowledgements

We thank Dr. Rebecca Umbach (Google) for comments on the manuscript. We thank Dr. Joe Campbell (University of Nottingham) for providing his LaTeX script to help develop Fig 1. in this manuscript.

ORCID iD

Cian O'Mahony: <https://orcid.org/0000-0002-6388-6182>.

2.7 References

- Adam-Troian, J., Caroti, D., Arciszewski, T., & Ståhl, T. (2019). Unfounded beliefs among teachers: The interactive role of rationality priming and cognitive ability. *Applied Cognitive Psychology*, 33(4), 720–727. Scopus. <https://doi.org/10.1002/acp.3547>
- Allington, D., McAndrew, S., Moxham-Hall, V., & Duffy, B. (2021). Coronavirus conspiracy suspicions, general vaccine attitudes, trust and coronavirus information source as predictors of vaccine hesitancy among uk residents during the covid-19 pandemic. *Psychological Medicine*. psych. <https://doi.org/10.1017/S0033291721001434>
- Bago, B., Rand, D. G., & Pennycook, G. (2022). Does deliberation decrease belief in conspiracies? *Journal of Experimental Social Psychology*, 103, 104395. <https://doi.org/10.1016/j.jesp.2022.104395>
- Banas, J. A., & Miller, G. (2013). Inducing Resistance to Conspiracy Theory Propaganda: Testing Inoculation and 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>
- Bartlett, J., & Miller, C. (2010). *The power of unreason: Conspiracy theories, extremism and counter-terrorism*. Demos London.
- Bierwiazek, K., Gundersen, A. B., & Kunst, J. R. (2022). The role of conspiracy beliefs for COVID-19 health responses: A meta-analysis. *Current Opinion in Psychology*, 46, 101346. <https://doi.org/10.1016/j.copsyc.2022.101346>
- Bierwiazek, K., Kunst, J. R., & Pich, O. (2020). Belief in covid-19 conspiracy theories reduces social distancing over time. *Applied Psychology: Health and Well-Being*. psych. <https://doi.org/10.1111/aphw.12223>

- Bonetto, E., Troïan, J., Varet, F., Lo Monaco, G., & Girandola, F. (2018). Priming Resistance to Persuasion decreases adherence to Conspiracy Theories*. *Social Influence*, 13(3), 125–136. Scopus. <https://doi.org/10.1080/15534510.2018.1471415>
- Brotherton, R., & French, C. C. (2014). Belief in Conspiracy Theories and Susceptibility to the Conjunction Fallacy. *Applied Cognitive Psychology*, 28(2), 238–248. <https://doi.org/10.1002/acp.2995>
- 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>
- 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>
- 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. Scopus. <https://doi.org/10.1007/s11162-018-9513-3>
- Einstein, K. L., & Glick, D. M. (2015). Do I Think BLS Data are BS? The Consequences of Conspiracy Theories. *Political Behavior*, 37(3), 679–701. Scopus. <https://doi.org/10.1007/s11109-014-9287-z>
- Enders, A. M., Uscinski, J. E., Klofstad, C. A., Seelig, M. I., Wuchty, S., Murthi, M. N., Premaratne, K., & Funchion, J. R. (2021). Do conspiracy beliefs form a belief system? Examining the structure and organization of conspiracy beliefs. *Journal of Social and Political Psychology*, 9(1), 255–271. <https://doi.org/10.5964/jspp.5649>
- Forgas, J. P., & Baumeister, R. F. (Eds.). (2019). *The Social Psychology of Gullibility: Fake News, Conspiracy Theories, and Irrational Beliefs* (1st ed.). Routledge. <https://doi.org/10.4324/9780429203787>

- Goldacre, B. (2012). *Bad pharma: How drug companies mislead doctors and harm patients*. Fourth Estate.
- Grimes, D. R. (2016). On the viability of conspiratorial beliefs. *PLoS ONE*, 11(1). Scopus. <https://doi.org/10.1371/journal.pone.0147905>
- Imhoff, R., Bertlich, T., & Frenken, M. (2022). Tearing apart the “evil” twins: A general conspiracy mentality is not the same as specific conspiracy beliefs. *Current Opinion in Psychology*, 46, 101349. <https://doi.org/10.1016/j.copsyc.2022.101349>
- Imhoff, R., Dieterle, L., & Lamberty, P. (2021). Resolving the puzzle of conspiracy worldview and political activism: Belief in secret plots decreases normative but increases nonnormative political engagement. *Social Psychological and Personality Science*, 12(1), 71–79. <https://doi.org/10.1177/1948550619896491>
- 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., & Paterson, J. L. (2020). Pylons ablaze: Examining the role of 5G COVID-19 conspiracy beliefs and support for violence. *British Journal of Social Psychology*, 59(3), 628–640. Scopus. <https://doi.org/10.1111/bjso.12394>
- Klimiuk, K., Czoska, A., Biernacka, K., & Balwicki, Ł. (2021). Vaccine misinformation on social media—Topic-based content and sentiment analysis of Polish vaccine-deniers’ comments on Facebook. *Human Vaccines & Immunotherapeutics*, 1–10. <https://doi.org/10.1080/21645515.2020.1850072>

- Knobel, P., Zhao, X., & White, K. M. (2022). Do conspiracy theory and mistrust undermine people's intention to receive the COVID-19 vaccine in Austria? *Journal of Community Psychology*, 50(3), 1269–1281. <https://doi.org/10.1002/jcop.22714>
- Lazić, A., & Žeželj, I. (2021). A systematic review of narrative interventions: Lessons for countering anti-vaccination conspiracy theories and misinformation. *Public Understanding of Science*. Scopus. <https://doi.org/10.1177/09636625211011881>
- Lewandowsky, S., Oberauer, K., & Gignac, G. E. (2013). NASA faked the moon landing—Therefore, (climate) science is a hoax: An anatomy of the motivated rejection of science. *Psychological Science*, 24(5), 622–633. [psych. https://doi.org/10.1177/0956797612457686](https://doi.org/10.1177/0956797612457686)
- Lewandowsky, S., & van der Linden, S. (2021). Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology*. Scopus. <https://doi.org/10.1080/10463283.2021.1876983>
- Marshall, J., & Linehan, C. (2021). Are Exergames Exercise? A Scoping Review of the Short-Term Effects of Exertion Games. *IEEE Transactions on Games*, 13(2), 160–169. <https://doi.org/10.1109/TG.2020.2995370>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & for the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, 339(jul21 1), b2535–b2535. <https://doi.org/10.1136/bmj.b2535>
- Mulukom, V. van, Pummerer, L., Alper, S., Bai, H., Čavojová, V., Farias, J., Kay, C. S., Lazarević, L. B., Lobato, E. J. C., Marinthe, G., Banai, I. P., Šrol, J., & Žeželj, I. (2020). *Antecedents and consequences of COVID-19 conspiracy beliefs: A systematic review*. <https://doi.org/10.31234/osf.io/u8yah>
- Murphy, K. R., Myors, B., & Wolach, A. (2014). *Statistical Power Analysis* (0 ed.). Routledge. <https://doi.org/10.4324/9781315773155>

- 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(MAY). Scopus. <https://doi.org/10.3389/fpsyg.2018.00684>
- Nyhan, B., & Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine*, 33(3), 459–464. <https://doi.org/10.1016/j.vaccine.2014.11.017>
- Nyhan, B., Reifler, J., Richey, S., & Freed, G. L. (2014). Effective Messages in Vaccine Promotion: A Randomized Trial. *Pediatrics*, 133(4), e835–e842. <https://doi.org/10.1542/peds.2013-2365>
- Nyhan, B., Reifler, J., & Ubel, P. A. (2013). The Hazards of Correcting Myths About Health Care Reform. *Medical Care*, 51(2), 127–132. <https://doi.org/10.1097/MLR.0b013e318279486b>
- 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. psych. <https://doi.org/10.3389/fpsyg.2016.01525>
- Plait, P. C. (2002). *Bad astronomy: Misconceptions and misuses revealed, from astrology to the moon landing “hoax.”* Wiley.
- Poon, K.-T., Chen, Z., & Wong, W.-Y. (2020). Beliefs in conspiracy theories following ostracism. *Personality and Social Psychology Bulletin*, 46(8), 1234–1246. psych. <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. Scopus. <https://doi.org/10.3389/fpsyt.2020.568942>

- Richardson, W. S., Wilson, M. C., Nishikawa, J., & Hayward, R. S. (1995). The well-built clinical question: A key to evidence-based decisions. *ACP Journal Club*, 123(3), A12-13.
- 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>
- Scherer, L. D., Yates, J. F., Baker, S. G., & Valentine, K. D. (2017). The Influence of Effortful Thought and Cognitive Proficiencies on the Conjunction Fallacy: Implications for Dual-Process Theories of Reasoning and Judgment. *Personality and Social Psychology Bulletin*, 43(6), 874–887. <https://doi.org/10.1177/0146167217700607>
- Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., ... Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, l4898. <https://doi.org/10.1136/bmj.l4898>
- Stojanov, A. (2015). Reducing conspiracy theory beliefs. *Psihologija*, 48(3), 251–266. Scopus. <https://doi.org/10.2298/PSI1503251S>
- Stojanov, A., Bering, J. M., & Halberstadt, J. (2020). Does perceived lack of control lead to conspiracy theory beliefs? Findings from an online mturk sample. *PLoS ONE*, 15(8 August). Scopus. <https://doi.org/10.1371/journal.pone.0237771>
- Sutton, R. M., & Douglas, K. M. (2020). Conspiracy theories and the conspiracy mindset: Implications for political ideology. *Current Opinion in Behavioral Sciences*, 34, 118–122. Scopus. <https://doi.org/10.1016/j.cobeha.2020.02.015>

- Swami, V., Chamorro-Premuzic, T., & Furnham, A. (2010). Unanswered questions: A preliminary investigation of personality and individual difference predictors of 9/11 conspiracist beliefs. *Applied Cognitive Psychology*, 24(6), 749–761.
<https://doi.org/10.1002/acp.1583>
- Swami, V., Pietschnig, J., Tran, U. S., Nader, I. W., Stieger, S., & Voracek, M. (2013). Lunar Lies: The Impact of Informational Framing and Individual Differences in Shaping Conspiracist Beliefs About the Moon Landings: Conspiracy theories. *Applied Cognitive Psychology*, 27(1), 71–80. <https://doi.org/10.1002/acp.2873>
- Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition*, 133(3), 572–585. psych.
<https://doi.org/10.1016/j.cognition.2014.08.006>
- Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the Backfire Effect: Measurement and Design Considerations. *Journal of Applied Research in Memory and Cognition*, 9(3), 286–299. <https://doi.org/10.1016/j.jarmac.2020.06.006>
- van der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating Against Fake News About COVID-19. *Frontiers in Psychology*, 11. Scopus.
<https://doi.org/10.3389/fpsyg.2020.566790>
- van Prooijen, J.-W., Douglas, K. M., & De Inocencio, C. (2018). Connecting the dots: Illusory pattern perception predicts belief in conspiracies and the supernatural: Illusory pattern perception. *European Journal of Social Psychology*, 48(3), 320–335.
<https://doi.org/10.1002/ejsp.2331>
- Whitson, J. A., Kim, J., Wang, C. S., Menon, T., & Webster, B. D. (2019). Regulatory focus and conspiratorial perceptions: The importance of personal control. *Personality and Social Psychology Bulletin*, 45(1), 3–15. psych.
<https://doi.org/10.1177/0146167218775070>

Wood, M. J. (2016). Some dare call it conspiracy: Labeling something a conspiracy theory does not reduce belief in it. *Political Psychology*, 37(5), 695–705. psych.

<https://doi.org/10.1111/pops.12285>

Xu, Z. (2019). Personal stories matter: Topic evolution and popularity among pro- and anti-vaccine online articles. *Journal of Computational Social Science*, 2(2), 207–220.

<https://doi.org/10.1007/s42001-019-00044-w>

Chapter 3: The Critical Thinking about Conspiracies Assessment (CTAC): Development and Validation

This paper is currently under review in *Applied Cognitive Psychology*.

Study Motivation

The findings of the systematic review conducted in Chapter 2 revealed that current assessments of conspiracy thinking have several limitations, including a focus exclusively on rewarding participants for rejecting conspiracy theories, and an emphasis on measuring belief in conspiracy theories rather than focusing on the discernment. This study sought to develop a psychological assessment that would address these limitations by measuring the underlying cognitive processes underpinning conspiracy thinking, and by measuring how people assess both plausible and implausible conspiracy theories. The current paper consists of four studies that validate this assessment.

Author Contributions

COM, CL, and GM developed the design and concept of this research. Survey design, data collection, data analysis, data interpretation, and data visualisation were conducted by COM under the supervision of GM. MM assisted in data interpretation and analysis for the CTAC Long-Form. COM wrote the manuscript and revised it in response to GM and CL's feedback.

3.1 Abstract

How should we measure people's vulnerability to unreasonable conspiracy theories?

This is a key question for those designing interventions that aim to limit the spread of

inaccurate beliefs. We argue that existing approaches are either too specific

(measuring belief in a known conspiracy theory) or too general (measuring broad

beliefs about the world). We developed the Critical Thinking about Conspiracies

assessment (CTAC), where participants critically assess fictional conspiracies. In

Study 1 (n = 3341) we conducted several iterative pilot studies to develop the items. In

Study 2 (n = 151) we conducted an exploratory factor analysis, reporting a two-factor

solution, where the ability to correctly appraise implausible and plausible conspiracy

theories are distinct but interrelated components. In Study 3 (n = 106) we conducted a

confirmatory factor analysis to assess the goodness of fit and demonstrated

convergent validity. In Study 4 (n = 225), we validated a long-form 15-item version

of the CTAC to improve the internal reliability. The short form CTAC has 8 items and

the long form CTAC has 15 items, which can both be easily administered online and

takes approximately 5-10 minutes to complete.

3.2 Introduction

Conspiracy theories are arguments that suggest that secretive groups are responsible for orchestrating and manipulating important events (Douglas et al., 2017). For example, some groups believe that President John F. Kennedy was assassinated by the CIA, or that the Covid-19 pandemic was fabricated by world governments. Many such widely believed conspiracy theories have been proven to lack evidence and seem unlikely to be true (Grimes, 2016, Flynn et al., 2017).

Participating in conspiracy ideation has been associated with a number of considerable negative consequences, such as vaccine hesitancy, social anxiety, isolation, rejection of science, and vandalism (Douglas et al., 2017; Imhoff et al., 2021; Jolley et al., 2022; Soveri et al., 2021). It is important to note however, that these negative outcomes specifically result from belief in *unviable*, *untrue*, or *misleading* conspiracy theories. Real conspiracies can and do occur. For example, Richard Nixon demonstrably used his power as president to conceal his role in the Watergate burglaries. Researchers are therefore primarily interested in belief in *unviable* conspiracy theories and many interventions attempt to lower susceptibility to these beliefs (for a review see O'Mahony et al., 2023). The majority of assessments that have been designed to measure conspiracy beliefs fall into one of two categories: specific conspiracy beliefs measures and general conspiracy belief measures.

Specific conspiracy belief assessments measure how much a person endorses beliefs related to a specific conspiracy theory. Much of the early research conducted on conspiracy belief took this approach, by measuring the extent that people believed in conspiracies related to specific topics such as vaccines (Jolley & Douglas, 2017), Princess Diana's death (Douglas & Sutton, 2008), 9/11 attacks (Swami et al., 2010), and the spread of HIV/AIDS (Bogart et al., 2010). In a measure like this, participants are asked whether they endorse a range of very specific statements such as "If aircraft struck the Pentagon and the World Trade Centres, they

were probably controlled by the US military” and “The fact that neither George W. Bush nor his security personnel responded to the 9/11 attacks in a manner that indicated the President might be in danger is proof that they had privileged access to the planning and execution of the attacks”.

Specific conspiracy belief assessments have a number of limitations. Firstly, the popularity of conspiracy theories can vary considerably across time and culture. As Brotherton et al (2013) notes, conspiracies about the 2005 London bombings may be common within the UK, but not so much in other parts of the world. These assessments make assumptions about participants’ familiarity with certain conspiracy theories. Secondly, even within a given conspiracy, the specific content may change as new evidence emerges or claims are refuted. As noted by Wood (2017), while the claim that ‘no Jews died on 9/11’ was popular in 2005, such arguments have disappeared from current conspiracies regarding 9/11. The constant shape-shifting nature of conspiracy theories, particularly in a digital age, means that if we want to measure beliefs (or develop effective interventions), we would need to constantly develop and/or update our assessment tools. Importantly too, while belief in one conspiracy does indeed predict belief in other conspiracy theories, there is little evidence that an individual who scores highly on a measure specific to a single conspiracy has similar conspiracy beliefs regarding other topics (Wood et al., 2012). There have been no studies to date that have investigated whether interventions that target specific beliefs (e.g., belief in 9/11 conspiracy beliefs) are also effective in challenging other beliefs such as anti-vaccination conspiracy beliefs.

One way to strengthen scales that are designed to measure belief in specific conspiracies is the inclusion of a mix of items related to a number of different conspiracies (e.g. conspiracies about the death of Princess Diana and UFOs) (Darwin et al., 2011; Douglas & Sutton, 2011). However, as Brotherton et al.,(2013) notes, these assessments tend to be

used only by the researchers who create them and are rarely adopted by other researchers, perhaps reflecting how difficult it is to generalize lists of relevant conspiracy theories across time and cultures. In some cases, these assessments have not been published or psychometrically validated (Darwin et al., 2011). Lack of standardized measures of conspiracy belief restricts the generalizability and reproducibility of conspiracy research.

There are also assessments that aim to measure a person's general tendency to believe in conspiracy theories. Previous research has indicated that conspiracy beliefs are part of a larger worldview, where an individual displays a general tendency to believe in conspiracy theories often referred to as *conspiracy ideation* (Williams et al., 2022). In this paper we distinguish between conspiracy beliefs (meaning a belief or set of beliefs that suggests that small groups manipulate important events, often in secret) and conspiracy ideation (meaning a person's general tendency to engage with conspiracy theories). Wood et al., (2012) found that belief in specific conspiracy theories about Osama Bin Laden could be predicted by belief in more broad and abstract conspiracy beliefs about the US government hiding information about his death. Some researchers have taken the approach of measuring broader conspiracy beliefs with the intention of covering the wider spectrum of conspiracy beliefs. For example, the Generic Conspiracy Beliefs scale (Brotherton et al., 2013) and the Conspiracy Mentality Questionnaire (Bruder et al., 2013) are widely used measures of an individual's tendency to endorse conspiracy beliefs.

General conspiracy measures also have their limitations, however. First, the items are purposefully extremely vague, as the aim is to avoid being bound by time or culture. This vagueness, however, leads to significant issues of interpretation. One item in the Generic Conspiracy Beliefs Scale is: "A lot of important information is deliberately concealed from the public out of self-interest" (Brotherton et al., 2013). Neither the "important information" the statement is referring to, nor who is concealing this information is made clear here,

leaving a lot up to individual interpretation – many participants might reasonably say they agree that surely *some* entities conceal *some* information, but does that really reflect an unreasonable or conspiratorial mindset? Another issue is that these measures contain items that may vary in cultural relevance. Another item in the Generic Conspiracy Beliefs Scale is: “The government is involved in the murder of innocent citizens and/or well-known public figures and keeps this a secret.” (Brotherton et al., 2013). While such statements might seem unreasonable in some countries, this statement would be factually true for those who lived under the regime of Bashar al-Assad or Ayatollah Khomeini (Chen, 2022; Amnesty International, 2018). While general conspiracy measures offer a more standardised measure of conspiracy beliefs and are more widely used across various studies by a number of different researchers (Bonetto et al., 2018; Poon et al., 2020; Stojanov, 2015; Swami et al., 2014), they have inherent flaws, highlighting the need for a new measure of conspiratorial thinking.

3.2.1 Proposing a New Approach to Measuring Conspiratorial Thinking

Due to the limitations of existing approaches to measuring conspiracy belief, a new approach is needed to address these concerns. This measure should not be bound by time or culture, nor should it make any assumptions about a participant’s knowledge of popular conspiracy theories nor their media consumption. Here, we argue that the best way to measure vulnerability to conspiracy theories is to focus on the *cognitive process* underlying conspiracy beliefs and conspiracy ideation, not the *outcome*. Focusing on the cognitive processes underlying conspiracy beliefs is separate from measuring conspiracy belief and conspiracy ideation. We will collectively refer to these processes as *conspiracy thinking*,

which we define as a set of cognitive and social processes associated with conspiracy ideation and both forming and endorsing conspiracy beliefs.

One reason why we suggest focusing on the cognitive processes underlying conspiracy beliefs rather than the outcome is the body of research that indicates a multitude of social processes that are associated with conspiracy belief and conspiracy ideation (Koroma et al., 2022). Social exclusion (Poon et al., 2020), in-group narcissism (Golec De Zavala & Lantos, 2020) and populist attitudes (Van Prooijen et al., 2015) have been shown to be associated with conspiracy beliefs. Researchers have suggested that belief in conspiracy theories serves to address social identity-related goals, such as reducing feelings of blame to their perceived social group (Bezalel, 2021). The possibility of socially-motivated beliefs may act as a confounding variable in any scale that measures belief in real-world conspiracy theories. The social processes underlying conspiracy belief and ideation may also differ depending on culture and time, as they are related to one's perceived social identity, religious beliefs, and group relations (Koroma et al., 2022). Measuring cognitive processes allows us to examine how participants evaluate information, without depending on information that may vary between different cultures.

Many previous studies have established that critical thinking plays a key role in conspiracy beliefs. The “gullible conspiracy hypothesis” has suggested that conspiracy beliefs are the product of errors in critical thinking and reduced tendency to rely on rational thinking (Van Prooijen, 2019). Cognitive styles such as analytical thinking and critical thinking have been shown to be associated with lower conspiracy beliefs (Alper et al., 2021; Georgiou et al., 2019; Gligorić et al., 2021; Lantian et al., 2021; Pennycook et al., 2020; Ståhl & Van Prooijen, 2018; Swami et al., 2014; Van Der Wal et al., 2018; J.-W. Van Prooijen, 2019). Conspiracy beliefs have been associated with lower cognitive ability (Adam-Troian et al., 2019; Rizeq et al., 2021). Furthermore, a greater susceptibility to certain logical fallacies

and biases such as the conjunction fallacy (Brotherton & French, 2014), jumping to conclusions and intuitive thinking (Pytlik et al., 2020), illusory pattern perception (Van Prooijen et al., 2018), and the illusion of explanatory depth (Vitriol & Marsh, 2018) are associated with a greater tendency towards conspiracy beliefs. Finally, research has indicated that experimentally promoting analytical thinking and deliberation has been associated with fewer conspiracy beliefs (Bago et al., 2022; Swami et al., 2014)

When we look at what works in terms of reducing belief in unviable conspiracy theories, it is also evident that critical thinking is an effective target for researchers. Indeed, in a recent systematic review, explicitly teaching critical thinking skills was found to be one of the most effective interventions (O'Mahony et al., 2023). Participants who enrolled in a pseudoscience class presented on a number of topics from UFOs, paranormal phenomena, and conspiracy theories and made arguments for and against their existence (Dyer & Hall, 2019). The study found that when compared to students who took a conventional research method class, the students in the pseudoscience class believed in significantly fewer conspiracy theories (Dyer & Hall, 2019). One of the advantages of this kind of intervention is that it focuses primarily on the *process* of falling for conspiracy belief by tackling the critical thinking skills of the participants, as opposed to focusing on the *outcome* (e.g. debunking specific conspiracy theories). Developing a measure that assesses the cognitive processes underlying conspiracy ideation is thus not only more likely to be effective but will also be more scalable as it will not be bound by time and culture in the way that existing measures are. While fundamental conspiracy beliefs may be difficult to change within the timeframe of a short intervention, individuals may be more likely to change how they approach a conspiracy and change how they critically evaluate novel conspiracies.

Our proposed measure of critical thinking about conspiracies would not be the first domain-specific measure of critical thinking. Several other researchers have tailored general

measures of critical thinking to subjects such as chemistry, physics, medicine, and psychology (Angel et al., 2000; Lawson et al., 2015; Nursa'adah et al., 2021; Tiruneh et al., 2017). Often these assessments present participants with hypothetical vignettes related to a core problem related to the domain of interest. For instance, the Psychological Critical Thinking Exam (PCTE; Lawson et al., 2015) measures how psychology students assess typical problems in psychology, such as identifying that an experiment is missing a control group. Much in the same way Lawson tailored the PCTE to issues that were specific to psychological research, our proposed assessment should primarily focus on measuring critical thinking in the specific context of hypothetical conspiracy theories.

Finally, some have recently argued that interventions in the general field of misinformation should use assessments that focus on *discernment* as opposed to measuring belief. Guay et al., (2022) argues that interventions that only measure belief in false statements cannot assess whether these interventions may be - unwittingly and undesirably - making people sceptical of content that may be true. A recent review found that less than half of existing misinformation studies measured both true and false statements (Murphy et al., 2023). Furthermore, of the studies that did include true statements, only 15% reported any measure of discernment. To really understand the effectiveness of a misinformation intervention, it is essential to measure how the intervention effects belief in both fake news *and* true news – especially where the goal is improved discernment rather than just increasing blind scepticism. Here, we argue that this question of discernment is equally important for the measurement of conspiracy beliefs. It is not sufficient to simply measure whether individuals can correctly reject unfounded beliefs (e.g. to give participants a list of claims about the possibility of a staged moon landing and measure whether they can reject them) (Swami et al., 2013). We must also assess how effectively they can identify a conspiracy that may be true and measure how effectively they can discern the plausible from the implausible. By

pairing discernment with a measurement of the underlying cognitive processes of conspiracy belief, we also sidestep a pertinent issue. While these unfounded conspiracy theories are often used synonymously with harmful conspiracy theories such as anti-vaccine narratives, unviable conspiracy theories are not necessarily the same as “harmful” conspiracy theories. One can believe in unviable conspiracy theories, such as the idea that the moon is actually made of cheese. While this conspiracy theory is demonstrably false, believing in such a theory does not necessarily impact one’s day-to-day decisions in the same way as believing in anti-vaccine conspiracy theories would. By focusing on the underlying cognitive processes, we are not focusing on whether participants believe in dangerous information, or even necessarily identifying “true” and “false” information. Instead, by focusing on the underlying cognitive processes, we can examine whether a person is applying a set of skills across a range of contexts and can correctly identify when strong or poor arguments are being used to support a conspiracy theory, rather than making value judgments on how harmful a conspiracy theory is. Only by measuring discernment can we accurately assess how participants are weighing up potentially conspiratorial information (and crucially, whether that ability is improving after any given intervention).

3.2.2 Present Study

We propose a domain-specific assessment of critical thinking that presents participants with a number of hypothetical conspiracy theories, both implausible and plausible. We believe this will address the shortcomings of current conspiracy belief measures by measuring the cognitive processes underpinning conspiracy beliefs, as well as measuring participants’ ability to discern between plausible and implausible conspiracy theories.

This article will outline our development of the Critical Thinking about Conspiracies assessment (CTAC). We conducted three studies to design this assessment. In Study 1 we

conducted a series of iterative pilot studies to design and develop the format and content of the assessment. In Study 2 we conducted an exploratory factor analysis to detect the underlying factors and latent variables in our assessment. Study 3 used a confirmatory factor analysis to assess the fit of our model, along with cross validating our assessment with related measures.

3.3 Study 1: Scale Development

As a critical thinking-based measurement of conspiratorial thinking, we decided to ground the development of our measure in existing measurements of critical thinking (CT). Specifically, we drew on the Halpern Critical Thinking Assessment (HCTA, Halpern, 2012). The HCTA focuses on five main CT domains; *reasoning, argument analysis, hypothesis testing, likelihood and uncertainty analysis, and decision-making and problem-solving* (Halpern, 2012). We chose the HCTA in particular as it has been suggested to be based on CT skills that are commonly mentioned across various definitions of CT (Tiruneh et al., 2017). Furthermore, many of the domains measured in the HCTA are associated with conspiracy beliefs. For instance, those who are inclined to believe in conspiracy theories are more likely to engage in jumping-to-conclusions bias, and the conjunction fallacy, both of which are CT errors that correspond to the *likelihood and uncertainty analysis* domain (Brotherton & French, 2014; Pytlik et al., 2020). In developing the CTAC, we created vignettes that contained a logical error linked to one of these CT domains.

3.3.1 Initial Scale format

We developed vignettes that propose hypothetical conspiratorial scenarios. We opted to use the vignette format used in the PCTA, as the item length was enough to provide participants with sufficient information to make an informed judgement (Lawson et al., 2015). Each of these vignettes were approximately 80-100 words each, to avoid participant

fatigue. Participants could choose from a number of explanations that they feel best describe how reasonable the claims in the vignettes are.

While some CT measures (such as the HCTA) use both open-ended responses and forced-choice items, feasibility checks resulted in the decision to use only forced-choice items in the CTAC. In particular, this choice was informed by difficulties in reliably and objectively scoring open-ended responses to complex conspiratorial questions. Instead, by providing a number of discrete options and asking participants to select the most reasonable conclusion, the scoring is more standardized, and the measure is much simpler to administer at scale. Items were based on a range of conspiracy topics such as sports, finance, academic research, industry, government activities and political scandals. We decided to cover a range of conspiracy topics to avoid fixating on one particular type of conspiracy theory, as previous research has indicated that familiarity with conspiracy theories can influence how participants think about conspiracy theories (Bost et al., 2010). The scope of these conspiracy theories also varied from scenarios related to town-level money laundering to government-level cover ups. We purposefully developed a very broad base of vignettes to ensure we weren't only describing stereotypical conspiracy theories that participants might reflexively reject due to not wanting to be labelled as a conspiracist (Lantian et al., 2018) .

We sought to include vignettes that describe both plausible and implausible conspiracy theories, to measure true discernment (Guay et al., 2022) and avoid rewarding participants for blind skepticism. We concluded that those who possess high CT ability in evaluating conspiracy theories should also be able to discern plausible conspiracies from implausible conspiracies.

3.3.2 Pilot Studies

Several iterative pilot studies were conducted to design the format and content of the CTAC (N=3341). First, we drafted 70 vignettes, each corresponding to one of the five critical thinking domains in the HCTA as a starting point. Using a think-aloud protocol with two participants, we identified which vignettes were the most readable and appropriate to include in our initial pool for testing.

After the think aloud protocol, we had an item pool of 35 vignettes. At this point, the items had five response options; one correct response (correctly endorsing or rejecting, for the right reason), one partially correct response (correctly endorsing or rejecting, for the wrong reason), two red-herring responses (incorrectly endorsing or rejecting), and finally the option to indicate that you did not know the answer. An example vignette can be seen below:

Dr. James Greenhouse is a former climate scientist from the University of Suoh who has recently featured on a number of podcasts discussing the science surrounding climate change. He became famous for protesting against the existence of climate change. Other scientists have publicly challenged him, stating that he was fired by the University of Suoh for his controversial views. Dr. Greenhouse says he was fired because multiple figures of authority feared that he would expose their wrongdoings. He added: “Many reputable scholars, such as Thomas Szatz, were kicked out of their universities because they dared to speak the truth!”

What do you make of the situation?

- a) Dr. Greenhouse should not be listened to. He is a singular academic and his viewpoints cannot be taken as scientific consensus. (4)
- b) The fact that people are willing the silence Dr. Greenhouse's viewpoints is evidence that his claims are true. Other academics would simply ignore him if his claims had no merit. (0)

- c) Dr. Greenhouse should not be listened to. He is popular due to controversy as opposed to his skill as a researcher. (1)
- d) Dr. Greenhouse should be trusted. He has too much to risk by discussing these controversial issues, and as such would not have any beneficial reason to lie. (0)
- e) I am not sure how to answer this question. (0)

In our initial piloting stages, participants were awarded 4-points for correctly identifying the most appropriate answer and 1-point for selecting answers that indicated some level of critical thinking, but missing the primary issue presented in the vignette. We recruited 1041 participants to take our first survey-based pilot study, with the goal of identifying which items in the 35-item pool showed the best item discrimination and to examine the frequency of correct responses among the sample. Items were broken into five categories based on the critical thinking domains outlined in the HCTA. Due to the considerable length of the 35-item version of the CTAC, participants were randomly assigned to 1 item per critical thinking category, resulting 5 items in total for each participant. In analysing this data, we assessed the item discrimination indices and item difficulty indices of each item. We then identified 10 candidate vignettes that had good-to-excellent item discrimination, and ideal item difficulty.

We then proceeded to examine the reliability of these items and identify whether there were clear underling factors in this 10-item version of the CTAC. We used an even split of implausible and plausible items. We conducted another pilot study with 200 participants to test which factors affected the reliability of this assessment. We investigated the reliability of a number of combinations of these CTAC items consistently produced weak reliability scores ($\alpha \sim .4$). We found that a large proportion of the sample had picked the partially correct answer and considered that perhaps the scoring difference (1, 4) between correct and partially correct was too large. We found that if we rescored the partially correct answers (3, 4), the

reliability was much stronger ($\alpha \sim .7$). We also found that several of the vignettes were somewhat verbose and identified a number of items that had responses that were too similar to be distinct responses and thus were causing issues with our reliability. These items and responses were revised for the next pilot study.

Additionally, we found that 15% of participants chose the “I am not sure how to answer this question” response. On reflection, we decided this response option was problematic. While critical thinking measures often provide participants with the choice to specify they do not know the answer (to avoid guessing), we noted that conspiracy theories are often ambiguous by nature (Wood, 2017). As such, after reading our brief vignettes, participants may (perhaps correctly) feel like they do not have sufficient information to make an informed decision and would default to not picking any of the possible explanations. Furthermore, previous research has indicated that endorsing conspiracy beliefs is associated with social stigma and the fear of social exclusion (Lantian et al., 2018). As such, participants may opt out of choosing any conclusion, defaulting to the (very reasonable) answer that we simply hadn’t given them enough information to go on. We decided forcing participants to select the most reasonable response would better assess their critical thinking skills and tested this in the next pilot study.

We ran another pilot test with 599 participants. Using a revised 11-item version of the CTAC, we found that the reliability was marginally better than the last pilot test ($\alpha \sim .5$). We identified that the issue resided in the partial scoring, as partially correct answers were difficult to keep consistent across all items. The disparity between the partially correct and correct items may not be consistent between items. As a result, we found that after removing the partially scored items, the reliability improved from poor to excellent ($\alpha = .75$). We decided ultimately to reshape the vignettes to have only one correct answer and no partially correct answers to avoid inconsistencies in scoring. We then tested how this impacted the

reliability of a revised 12-item version of the CTAC in our next pilot study. Items were retained based on how they affected the reliability, and new vignettes were drafted to have an even number of plausible to implausible items.

We recruited 675 participants to test the reliability and underlying factors of a 12-item version of the CTAC. We found that the 12-item version had good reliability ($\alpha \approx .6$), but there was no discernible pattern among the factors identified in our preliminary factor analysis. We also found that plausible and implausible items were loading negatively to one another. As such, we aimed to identify whether validating both the plausible and implausible subscales as their own independent scales could help increase their reliability.

We ran two separate surveys for both the implausible ($n=137$) and plausible scales ($n=193$). We found that the versions of the implausible scale with a length ranging from 9-11 items showed excellent reliability ($\alpha = .73 - .75$), while the plausible scale with a range of 13-14 items had unsatisfactory reliability ($\alpha = .4 - .5$). We posited that participants might have been suspicious of consecutively endorsing too many conspiracy items in the plausible scale and were changing their answers based on this expectation. As such, we recombined the implausible and plausible subscales. We found that an 18-item version of the CTAC had an excellent overall reliability ($\alpha = .87$) and excellent reliability for both the implausible ($\alpha = .73$) and plausible ($\alpha = .76$) subscales.

While the 18-item assessment had strong reliability, there were concerns about the length of the assessment, and preliminary factor analysis identified a number of items with poor factor loadings and cross-loading that warranted removing. We prioritised the CTAC having a strong underlying factor structure over high reliability so that the assessment would be theoretically practical. The CTAC was also designed as a measure to be used in intervention studies. A large item list would make the CTAC impractical for experimental use due to the time needed to complete the assessment. Interventions that use longitudinal designs

may need to present the CTAC more than once to measure how participants' appraisal of conspiracy theories change over time. As such, the number of items were reduced to discern the underlying factor structure and to create a more practical assessment. By flagging items that had loadings under .4, and items that were cross-loading to two or more factors, we excluded a total of 10 items, bringing us to a total of 8 items.

In our final pilot study, we recruited 150 participants and identified two clear factors using an 8-item version of the CTAC: one for the implausible items, and the second for the plausible items. We then used this final item list for our exploratory factor analysis.

3.3.3 Final CTAC Format

The final version of the CTAC consisted of an 8-item assessment. Preliminary factor analysis indicated a split of 5 implausible items to 3 plausible items. We felt that this imbalance was appropriate as it is more likely for conspiracies that people encounter to be implausible rather than plausible (Grimes, 2016). Every vignette is followed by four response options, all providing some explanation to address the events that occurred in the vignette. For every item, two response options endorse the conspiracy, while the other two response options reject the conspiracy. As we were interested in the reasoning rather than just the outcome, a participant could correctly reject a conspiracy but may miss the core critical thinking skill that the item was measuring and would thus be scored as zero. At the beginning of the scale participants are told that they will be asked to assess several hypothetical scenarios in the form of vignettes and to “choose the answer that best reflects what you think”. An example

item is:

There is a claim that a British rock band named the ‘Sliding Pebbles’ released an album that contained secret messages meant to influence listeners. Several listeners claimed that if you play track 6 backwards, you can make out the words ‘join the army’. As such, there have been claims of government involvement with the production of the album. What do you make of this claim?

- a) This claim is unreasonable. It makes far too many assumptions (i.e. the band are involved with the government, humans can process sound backwards).
- b) This claim is reasonable. If listeners can clearly make out the words ‘join the army’, then it is obviously intentional, it is unreasonable to believe it is just an accident.
- c) This claim is unreasonable. If the government could control the public with music, they would influence people to do something more useful than just joining the army.
- d) This claim is reasonable. Similar occurrences have happened in the past.

This example draws on the argument analysis domain of the HCTA. This item tests whether participants will identify the simplest solution for the problem without making unnecessary assumptions. Answer a) adequately addresses this vignette, by identifying that the conspiracy is making too many unwarranted assumptions (i.e., assuming humans can process sounds backwards, that the band is colluding with the government). Answer b) jumps to the conclusion that there is foul play solely on the basis that people recognise a pattern.

Answer c) makes a number of unwarranted assumptions, one being that the government would use this technology for grander purposes than making people join the army. Answer d) draws their conclusions from inference without considering any other explanations and uses evidence that is not within the vignette itself. Furthermore, there are a myriad of reasons a paper could be redacted, it is not always a result of foul play.

3.3.4 Participant Feedback

In our first pilot study, we included a measure of generic conspiracy ideation, the Generic Conspiracy Beliefs Scale (GCBS, (Brotherton et al., 2013) to compare to the CTAC. When asked for feedback on the survey (“Is there anything you'd like to tell us about your experience of the survey?”), we found that several participants took the opportunity to criticise the GCBS, mainly suggesting that the assessment was vague, constrained by time and culture, and was not specific about the scope of conspiracies it was referring to. To be clear, participants were blind to the goals of our study and from the comments, most seemed to assume that we were the research team behind the GCBS. These comments echoed our original concerns that motivated us to develop the CTAC and some of the most common feedback is detailed below.

Many comments related to the GCBS being vague in nature and therefore difficult to answer (“*It's hard to know what I'm to answer as it is so vague*”, “*The second set of questions [referring to the GCBS] are very vague- There's very little that you can say 100% is or isn't true without evidence*”). Some participants provided explicit examples, with one highlighting the item “Technology with mind-control capacities is used on the people without their knowledge”. The participant critiqued this, saying “*This 100% happened, the most obvious recent example being Facebook testing to see if they could manipulate users'*

emotions via algorithms". The participant is likely referring to an experiment conducted by Facebook which sought to alter users' emotions by manipulating their feed (Kramer et al., 2014). Other participants made similar remarks, though some even pointed out that they suspected what it was that the scale was trying to measure "*...the Cambridge Analytica story, once it broke, demonstrated that yes, indeed new technology does get tested on people without their informed consent. I have the feeling the question is, however, targeted more towards people who believe in concepts such as "there is a microchip in my vaccine"*".

Participants also indicated that often the scope of the conspiracy statements in the GCBS was not clear. When answering items related to belief in small groups controlling the world, many participants made remarks like "*I don't believe in Illuminati type conspiracies, but the fact that wealthy private interests groups have strong influences on any government certainly factored into my responses for those parts of the survey*". Similarly, another participant said: "*The 'small groups' with influence over governments do exist, although it is less furtive than the questions might suggest. Civil servants often have more power than figureheads of state, with the ability to drive policy and government in particular ways. This is just how government works, rather than an indicator of some underground, nefarious group that is controlling world politics.*" One participant asked, "*Is this more in relation to conspiracy theories of Illuminati or, the more realistic, billionaire donors swaying decisions?*" It is clear that participants have a sense of what the scale is trying to measure (conspiracy beliefs), and they are making their own judgments about what counts as a stereotypical conspiracy belief (the Illuminati, microchips in vaccines) and what does not (shady business deals or political lobbying). This lack of specificity is problematic for researchers measuring conspiracy beliefs, as so much is left to participants' interpretations.

In a similar vein, many comments referred to the constraints of the assessment based on time and culture, specifically in relation to the use of the term “government” in three of the fifteen items in the GCBS. Participants suggested that the likelihood of these conspiracy theories occurring depends greatly on which “government” these items are referring to (*“For example, do the government secretly have people killed is a definitely not in Ireland but a definite yes in Russia and other countries. I answered any questions of this type as being about Ireland”, “If you are talking about the Irish Government it’s probably not true they are involved in routine experiments on the public, murders, cover ups, false flag operations, whereas the US government has demonstrably been involved in all of those things in the past.”*). Another common critique made by participants was that items made reference to the likelihood of conspiracies occurring but made no mention regarding their frequency (*“While on a routine basis none of these conspiracies are true, the few times they’ve occurred would indicate they “Definitely” happened”*).

Though fewer comments were made regarding the CTAC, most of them were positive and suggested some enjoyment of the quiz-like structure: *“The vignettes and questions were really interesting and thought-provoking”, “The questions about how we would assess certain situations were super interesting”*. However, some participants critiqued the CTAC, suggesting that there wasn’t sufficient information in some vignettes to draw accurate conclusions: *“The questions that I answered “Not sure how to answer this question” had insufficient information to evaluate the claim, and none of the other answers fit (they all committed to some position, while I had no position).”* Overall, participants seemed to enjoy the CTAC, with one referring to it as “the most interesting I’ve tasted [sic] in a real long time!”. Other comments said: *“Quick and easy, and thought provoking. Enjoyed taking this survey”, “It was fun ☺ [in the sense that it was unusual and different from common surveys].”*

3.4 Study 2: Exploratory Factor Analysis

3.4.1 Methods

3.4.1.1 Participant

A total of 151 participants were recruited from Prolific. This number was based on the guideline that a confirmatory factor analysis ought to have at least 10 participants per item (Nunnally, 1978). Participants were between 19-80 years of age ($M = 41.72$, $SD = 14.93$). Participants reported their sex as female ($n = 94$) and male ($n = 56$). The majority of participants were from the United Kingdom (91.33%), and the remaining participants were from Ireland (3.33%), the United States (4%), and Australia (1.33%). No participants were excluded from the analysis as none had failed two attention checks.

3.4.1.2 Materials

The 8-item version of the CTAC was administered through a Qualtrics survey (see Table 3.1. for the full instrument).

3.4.1.3 Procedure

Participants were recruited to take part in the experiment through the Prolific online recruitment platform. The study was advertised as an experiment that was “studying how people use critical thinking to evaluate online information”. Participants were directed to the survey hosted on Qualtrics. Once informed consent was obtained, participants were shown the 18-item version of the CTAC. The items and responses were both randomized. After completing the CTAC participants were then redirected to the debriefing page. The debrief page used a strategy similar to Greene & Murphy (2023) to ensure participants that the conspiracy material used in the survey was entirely fictional, and should not be taken into consideration. Participants were then redirected to Prolific to confirm that they had finished the survey.

3.4.2 Results

An exploratory factor analysis was performed using principal components analysis and varimax rotation. A two-factor solution was identified through a combination of examining the scree plot, and removing items with strong cross-loading, or low item loadings. As can be seen in Table 3.1, the analysis revealed that all implausible items strongly coded to the first factor, and all plausible items coded to the second factor. The 8-item scale showed good reliability, Cronbach's $\alpha = .62$. There was evidence of a negative skew, with a mean score of 5.28 above the mid-point of the scale (4).

Table 3.1: Factor loadings for 8-item version of CTAC

	Implausible	Plausible
Imp1	.608	.311
Imp2	.643	-.112
Imp3	.473	.256
Imp4	.692	.057
Imp5	.770	-.115
Pla1	-.027	.608
Pla2	.181	.694
Pla3	-.023	.749

Category	Item	Responses
Implausible	The public has long suspected that the United States government has been covering up the existence of aliens. Recently, the Pentagon released a statement accompanied by some images that report that an 'Unidentified Flying Object' (UFO) has been spotted. What do you think is the correct interpretation of this report?	oThe Pentagon has just confirmed that aliens do in fact exist. The report clearly supports that they have seen a UFO. (0) oThe Pentagon confirmed they have spotted something in the sky they cannot identify and do not recognise. This, however, does not automatically suggest the existence of aliens. (1) oThe UFO sightings are simply a result of advanced military technology or experimental aircraft. The government is probably using this to test new equipment and made a pretence to conceal the project from the public. (0) oThe UFO sightings are likely just a distraction from something more significant that the government doesn't want us to know. They're probably using this to divert our attention from more pressing issues. (0)
Implausible	There is a claim that a British rock band named the 'Sliding Pebbles' released an album that contained secret messages meant to influence listeners. Several listeners claimed that if you play track 6 backwards, you can make out the words 'join the army'. As such, there have been claims of government involvement with the production of the album. What do you make of this claim?	oThis claim is unreasonable. It makes far too many assumptions (i.e. the band are involved with the government, humans can process sound backwards). (1) oThis claim is reasonable. If listeners can clearly make out the words 'join the army', then it is obviously intentional, it is unreasonable to believe it is just an accident. (0) oThis claim is unreasonable. If the government could control the public with music, they would influence people to do something more useful than just joining the army. (0) oThis claim is reasonable. Similar occurrences have happened in the past. (0)
Implausible	Researchers in a prominent university have been investigating means of creating clean energy that will be 5 times more effective than current fossil fuel energy sources. Last week, a section of the university caught fire. Among the regions effected 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?	oThis reasoning is flawed. An action being useful to someone does not automatically imply they committed it. (1) oThis reasoning is flawed. Oil companies wouldn't want to draw bad press on themselves (0) oThis 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. (0) oThis reasoning is valid. We know Big Oil can be ruthless in silencing opposition, so it is likely this is what happened here. (0)
Implausible	In 1992, Dr. John M. Vincent apparently made a remarkable discovery regarding the use of antineoplastons in treating deadly forms of cancer. While the original findings were originally published in the medical journal 'Directions in Current Oncology', the journal later redacted the paper. Dr. Vincent claims that his research threatened those who made a profit from treating cancer patients for life through means such as chemotherapy. What do you make of this situation?	oit is more plausible that the journal redacted the findings of the paper for legitimate reasons, as opposed to silencing the author. (1) oThe fact that the journal redacted the paper is clear proof that Dr. Vincent was silenced. (0) oSimilar cases have happened before, therefore it is not impossible that Dr. Vincent's claims are true. (0) oDr. Vincent's claims have no merit. If he was silenced it was likely for good reason. (0)
Implausible	Dr. James Greenhouse is a former climate scientist from the University of Suoh who has recently featured on a number of podcasts discussing the science surrounding climate change. He became famous for protesting against the existence of climate change. Other scientists have publicly challenged him, stating that he was fired by the University of Suoh for his poor research practices. Dr. Greenhouse says he was fired because multiple figures of authority feared that he would expose their wrongdoings. He added: "Many reputable scholars, such as Thomas Szatz, were kicked out of their universities because they dared to speak the truth!" What do you make of the situation?	oDr. Greenhouse should not be listened to. He is a singular academic and his viewpoints cannot be taken as scientific consensus. (1) oThe fact that people are willing to silence Dr. Greenhouses viewpoints is evidence that his claims are true. Other academics would simply ignore him if his claims had no merit. (0) oDr. Greenhouse should not be listened too. Anyone who appears in the media should not be trusted (0) oDr. Greenhouse should be trusted, scientific breakthroughs often come from those questioning the status quo (0)
Plausible	There are rumors on the internet that striker Antonio Vincenzi threw the semi-finals game against Barcelona. It was noted that while Vincenzi has had a track record of top performance, he made a number of ridiculous decisions during the game such as accidentally passing to his opponents and committing reckless fouls.	oThis claim is unreasonable. Although Vincenzi could have been tempted by money, he would care too much about his fans to betray their trust. (0) oThis claim is unreasonable. Even the best of footballers has bad days where they make stupid decisions on the pitch. It was just a coincidence he had been photographed with the ringleader of the gambling syndacite and there is nothing here that warrants suspicion. (0) oThis claim is reasonable. Footballers only care about money, so he would throw the game if offered enough. Further investigation is needed. (0) oThis claim is reasonable. Vincenzi's performance tied with the picture with a known betting agent is too much of a coincidence to ignore. Further investigation is needed. (1)
Plausible	People are proposing that he may have been paid off to throw the game. Soon after, a picture was found of Vincenzi in a hot tub with the ringleader of a famous gambling syndacite. What do you make of this claim?	
Plausible	Two employees at Arcadia Financial Holdings bought shares in companies that were soon acquired, and then sold those shares for a quick profit once the news of the acquisition became public. The employees have had success with this method on eight consecutive occasions. While many are praising them as hot-shot investors, there are rumours that the investors recruited an employee at Lennon Brothers to tip them off about upcoming mergers, insinuating they were engaging in illegal insider trading. One particular Lennon Brothers worker was involved in all eight of the investments in question. What do you think is most likely?	oLennon brothers is a huge company and handle thousands of mergers, this is likely just be a coincidence (0) oThey likely engaged in insider trading. Predicting acquired companies eight consecutive times, along with the Lennon worker connection suggests insider knowledge. It is too much of a coincidence (1) oThe Arcadia employees may just be proficient at spotting opportunities, whether it's luck or skill, this does happen sometimes (0) oOf course they engaged in insider trading. Even without any evidence, it's well known most bankers do (0)
Plausible	Star NFL player Chad Oatway recently had his contract terminated by his club. In the weeks prior to this decision, Oatway had become vocal about being a Mormon, and would constantly broadcast his beliefs during victory speeches and interviews. While his team state that he was cut from his team for poor performance on the field, his fans cite that he had an excellent track record, as evident from his recent player statistics. It has been suggested that he was removed from the team due to his beliefs. What do you believe happened?	o This suggestion is unreasonable. His fans are biased and the timing with him becoming vocal as a Mormon is probably just a coincidence. (0) oThis suggestion is unreasonable. The club would get in a lot of trouble for firing someone for their religious beliefs so would never do it. (0) oThis suggestion is reasonable. The fact that the team's excuses for firing him contradict his player statistics and performance, and the close timing with Oatway's openness about his beliefs is too much of a coincidence and warrants further investigation. (1) oThis claim is reasonable. Alcoholic drink sponsors had a motive to cut a Mormon, as Mormons do not condone the use of alcohol (0)

Figure 3.1: Item list for final 8-item CTAC

3.5 Study 3: Confirmatory Factor Analysis

Having identified a possible two-factor solution for the CTAC, with latent constructs related to implausible and plausible items, we conducted a confirmatory factor analysis to test the fit of this proposed two-factor model. We also tested the convergent validity of the CTAC by testing its relationship with adjacent measures: conspiracy ideation and cognitive reflection.

3.5.1 Methods

3.5.1.1 Participants

106 participants were recruited through the Prolific online recruitment platform. This number was based on the guideline that a confirmatory factor analysis ought to have at least 10 participants per item (Nunnally, 1978). Participants were between 18-78 years of age ($M = 42.76$, $SD = 14.90$). Participants reported their sex as female ($n = 72$), male ($n = 33$), and prefer not to say ($n = 1$). The majority of participants were from the United Kingdom (87%), and the remain participants were from Ireland (8.5%), the United States (2%), and Australia (2%). No participants failed two attention checks, meaning none were excluded from the analysis.

3.5.1.2 Materials

Critical Thinking about Conspiracies Assessment: The 8-item version of the CTAC developed through the exploratory factor analysis was administered in random order to the participants.

Generic Conspiracist Beliefs Scale (GBCS): 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.”). Responses range from 1 (definitely not true) to 5 (definitely true). Higher scores are indicative of higher conspiracy belief. We expected that as critical appraisal of conspiracy theories increased as reflected by scores on the CTAC, that it would be associated with lower general conspiracy beliefs. Scores ranged from a possible 0 to 75. The mean score was 12.03, suggesting low levels of belief in these statements overall. The internal reliability was excellent ($\alpha=.93$).

Cognitive Reflection Test (CRT): The Cognitive Reflection Test (Frederick, 2005) is a 3-item reasoning test that assesses a participant’s ability to engage in reflective thinking as opposed to responding with an intuitive answer. We selected this item as an indication of analytical thinking, as we expected that higher scores on the CTAC would reflect a higher ability to apply analytical thinking. The test presents participants with numerical reasoning scenarios and is scored dichotomously. Participants receive one point for each correct reflective response. Scores ranged from a possible 0 to 3. The mean score was 1.19. We used an adapted version of the CRT that used four response options (based on (Sirota & Juanchich, 2018)) as opposed to two to reduce the likelihood of participants guessing the correct response. Each item had one correct response, and three incorrect responses. The internal reliability was excellent ($\alpha=.74$). An example item was:

If it takes 6 hackers 6 minutes to crack 6 passwords, how long would it take 90 hackers to crack 90 passwords?

a) 6 minutes

- b) 90 minutes
- c) 30 minutes
- d) 600 minutes

3.5.1.3 Procedure

Participants were recruited through the Prolific online recruitment platform and redirected to the Qualtrics survey. Once informed consent was obtained participants were presented with the 8-item version of the CTAC. Items and responses were randomised. Participants also completed two attention checks that were included within the CTAC and presented in a random order amongst the real items. The attention checks consisted of questions where participants were told exactly how to respond and had to select that answer option to be marked as correct. Any participant who failed both these checks would have been excluded from the analysis (though none did). Participants then completed the GCBS and the CRT (with items presented randomly within each scale). Finally, participants were debriefed and informed about that the study was examining conspiracy beliefs. The same debriefing strategy from Study 1 was used to mitigate the possibility that participants believed any of the misinformation items presented within the study.

3.5.2 Results

The findings of the EFA in Study 2 suggested a two-factor solution. For the purpose of comparison, we conducted confirmatory factor analysis comparing both a single factor and two-factor solution to the CTAC to assess their respective fit. While the two-factor solution proposed two distinct latent constructs corresponding to the ability to analyse implausible and plausible conspiracy theories, the single factor solution proposed that all eight items in the CTAC loaded to a single latent variable (Critical Appraisal). Standardised coefficients were

used with a maximum likelihood estimation model. The scale was shown to have good reliability ($\alpha=.62$)

3.5.2.1 One-factor solution

The single-factor solution showed poor fit to the data. The chi-squared goodness-of-fit was significant ($\chi^2 = 34.74$, $df = 20$, $p = .021$), suggesting a significant difference between the proposed model and the observed data. Furthermore, the Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR) all indicated inadequate fit (CFI = .81 RMSEA = .09, SRMR = .09).

3.5.2.2 Two-factor solution

As seen in Figure 3.2, the two-factor solution showed a greater fit to the observed data in comparison to the single-factor model. The chi-squared test was not statistically significant ($\chi^2 = 17.43$, $df = 19$, $p = .56$). The CFI exceeded the recommended threshold of 0.95, and the RMSEA and SRMR fell within the acceptable ranges, indicating excellent fit (CFI = 1.00, RMSEA < 0.01, SRMR = 0.06). The two-factor solution suggested a better fit for the model, with two latent factors for detecting implausible and plausible conspiracy theories.

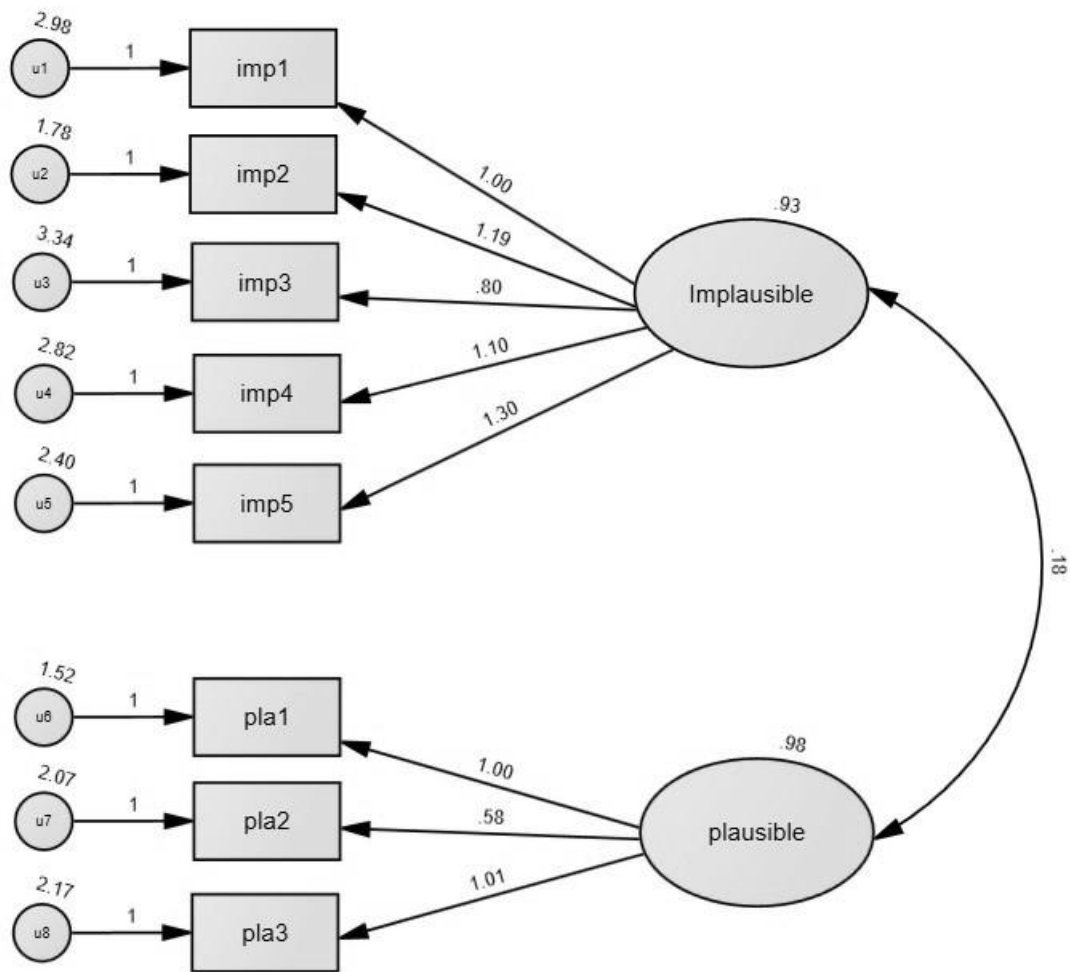


Figure 3.2: Two-factor solution of the CTAC

3.5.2.3 Cross Validation

Scores on the CTAC were cross-validated with two associated measures: generic conspiracy belief and cognitive reflection. We expected that a higher ability to critically appraise conspiracy theories would be associated with fewer conspiracy beliefs, and a higher ability to engage in cognitive reflection. As such, we hypothesized that higher scores on the CTAC would be correlated negatively with scores on the Generic Conspiracy Beliefs Scales (GCBS) and would be positively correlated with scores on the Cognitive Reflection Test

(CRT). We found that scores CTAC showed a medium negative correlation with the GCBS ($r = -.58, p < .001$), and there was a small positive correlation with scores on the CRT ($r = .31, p < .001$). These results suggest that higher scores on the CTAC suggest a reduced tendency to endorse conspiracy theories and reflect a higher tendency to engage in cognitive reflection. Additionally, we found that scores on the CRT were positively correlated with both the Implausible ($r = .23, p = .020$) and Plausible ($r = .27, p = .006$) subscales of the CTAC.

3.6 Study 4: CTAC Long Form

While our findings from Study 1 and 2 identified a short eight-item assessment with a clear two-factor solution, there were concerns over the low reliability of the CTAC. As our pilot studies had identified longer versions of the CTAC with stronger reliability, we opted to validate a long-form version of the CTAC that would have more robust reliability than its short-form counterpart.

3.6.1 Methods

3.6.1.1 Procedure

All procedures and design were identical to the one used in Study 2.

3.6.1.2 Materials

Critical Thinking about Conspiracies Assessment: An 18-item version of the CTAC was used for this analysis. Scores ranged from a possible 0 – 18.

3.6.2 Exploratory Factor Analysis

An exploratory factor analysis was performed using principal components analysis and varimax rotation. A two-factor solution was identified through a combination of examining the scree plot, and removing items with strong cross-loading, or low item loadings. The analysis revealed that all implausible items strongly coded to the first factor, and all plausible items coded to the second factor. We dropped three items that were cross-

loading with both factors, leaving us with a 15-item version of the scale. There were eight implausible items and seven plausible items. The factor loadings for each item can be seen in Table 3.2. The 15-item scale showed good reliability, Cronbach's $\alpha = .73$. There was evidence of a negative skew, with a mean score of 10.07 above the mid-point of the scale (8).

Table 3.2: Factor loadings for 15-item version of CTAC

	Implausible	Plausible
Imp1	.409	.161
Imp2	.650	.072
Imp3	.599	.130
Imp4	.540	.302
Imp5	.565	.145
Imp6	.661	.014
Imp7	.613	-.103
Imp8	.304	.307
Pla1	.075	.483
Pla2	-.145	.457
Pla3	.054	.575
Pla4	.183	.450
Pla5	.142	.653
Pla6	.111	.627
Pla7	.217	.482

3.6.3 Confirmatory Factor Analysis

3.6.3.1 Participants

154 participants were recruited through the Prolific online recruitment platform. This number was based on the guideline that a confirmatory factor analysis ought to have at least 10 participants per item (Nunnally, 1978). Participants were between 19-75 years of age ($M = 43.57$, $SD = 14.47$). Participants reported their sex as female ($n = 109$), male ($n = 43$), and prefer not to say ($n = 1$). The majority of participants were from the United Kingdom (98%),

and the remaining participants were from Ireland (2%). No participants failed two attention checks, meaning none were excluded from the analysis.

3.6.3.2 Materials

Critical Thinking about Conspiracies Assessment: The 15-item of the CTAC was used for this analysis. Scores ranged from a possible 0 – 15. The mean score was 8.96 (SD = 2.64) which was slightly above the midpoint of the assessment. The internal reliability was good ($\alpha = .71$)

Generic Conspiracist Beliefs Scale (GCBS): The 15-item GCBS was used to measure participants' general conspiracy ideation. We used this as a cross-validation measure to see whether higher critical appraisal of conspiracy theories (measured on the CTAC) were linked to lower conspiracy ideation. Scores ranged from a possible 0 to 75. The mean score was 38.48 (SD = 12.97), around the midpoint of the assessment. The internal reliability was excellent ($\alpha = .94$).

Cognitive Reflection Test (CRT): Once again, we used the 3-item Cognitive Reflection Test (Frederick, 2005) to assesses a participant's ability to engage in reflective thinking as opposed to responding with an intuitive answer. We used this as a cross-validation measure to see whether scores on the CTAC were linked to higher ability to engage in cognitive reflection. Scores ranged from a possible 0 to 3. The mean score was 1.29 (SD = 1.19). The internal reliability was good ($\alpha = .73$)

3.6.4 Results

3.6.4.1 Two-Factor Solution

Following the structure of the 8-item CTAC, along with the results of the exploratory factor analysis from the EFA of the 15-item CTAC, we tested a two-factor solution for the 15-item Long Form CTAC. The two-factor solution proposed two distinct latent constructs corresponding to the ability to analyse implausible and plausible conspiracy theories. Standardised coefficients were used with a maximum likelihood estimation model. The chi-squared test was not statistically significant ($\chi^2 = 97.26$, $df = 89$, $p = .26$). The CFI exceeded the recommended threshold of 0.95, and the RMSEA and SRMR fell within the acceptable ranges, indicating excellent fit (CFI = 0.96, RMSEA = 0.025, SRMR = 0.06).

3.6.4.2 Cross-validation

Scores on the CTAC were cross-validated with two associated measures: generic conspiracy belief and cognitive reflection. We expected that a higher ability to critically appraise conspiracy theories would be associated with fewer conspiracy beliefs, and a higher ability to engage in cognitive reflection. As such, we hypothesized that higher scores on the CTAC would be correlated negatively with scores on the Generic Conspiracist Beliefs Scales (GCBS) and would be positively correlated with scores on the Cognitive Reflection Test (CRT). We found that scores CTAC showed a medium negative correlation with the GCBS ($r = -.36$, $p < .001$), and there was a small positive correlation with scores on the CRT ($r = .30$, $p < .001$). These results suggest that higher scores on the CTAC suggest a reduced tendency to endorse conspiracy theories and reflect a higher tendency to engage in cognitive reflection. Additionally, we found that scores on the CRT were positively correlated with both the Implausible ($r = .26$, $p = .002$) and Plausible ($r = .24$, $p = .003$) subscales of the CTAC.

3.7 Discussion

In this paper we present a novel assessment of ability to critically assess conspiracy theories, using a number of fictional vignettes. The final 8-item measure is relatively brief –

participants in this study took approximately 5 minutes to complete the CTAC – and does not require coding of open-ended responses, making it attractive for large online studies. Additionally, the 15-item long-form version of the CTAC is offers a more reliable and robust version of the assessment. This is the first measure to assess critical thinking applied to conspiracy theories.

One important manner in which the long and short CTAC differ from existing measures of conspiracy belief is that it assesses belief in both plausible and implausible conspiracies. Both the exploratory and confirmatory factor analysis indicated a two-factor solution for the long and short-form version of the CTAC, with two distinct skills pertaining to critically appraising implausible and plausible conspiracy theories. The two-factor solution indicates that the ability to critically appraise plausible and implausible conspiracy theories are two distinct but interrelated skills. These findings are consistent with previous research that indicates that there are different factors associated with implausible and plausible conspiracy beliefs (Frenken et al., 2024). Our results also indicated that cognitive reflection was associated with critically appraising both implausible and plausible conspiracy theories, which is inconsistent with previous findings that suggest the reasoning is associated with endorsing implausible conspiracy theories but not plausible ones (Hattersley et al., 2022).

Unlike other domain-specific critical thinking assessments which have identified distinct factors related to specific critical thinking skills such as reasoning or argument analysis (Nursa'adah et al., 2021), the underlying factors of the long and short form versions of the CTAC reflect a distinction between appraising implausible and plausible conspiracy theories rather than specific critical thinking skills. We suspect that a lack of distinct latent variables reflecting specific critical thinking skills within the CTAC may be due to a significant overlap between individual critical thinking skills. Despite many of the early items in our pilot version of the CTAC focusing on a single specific skill related to those outlined

by the HCTA (HCTA: Halpern, 2012), the later, more realistic items often consisted of multiple interrelated critical thinking skills. This likely reflects the ambiguous nature of conspiracy theories (Wood, 2017). Since conspiracy theories may have no definitive explanation, participants completing the CTAC are required to evaluate the likelihood of different explanations for the scenarios they read and must analyse different arguments to determine which are the most likely. Thus, nearly all CTAC items require proficiency in HCTA skills related to *likelihood and uncertainty analysis* along with *argument analysis*, despite what the original item intended to measure. These findings were reflected in our pilot testing studies, where many items cross-loaded to multiple factors.

The dual component structure of the CTAC allows the measurement of discernment rather than blink scepticism. Stojanov & Halberstadt (2019) highlighted the importance of healthy scepticism in the context of conspiracy beliefs, as while conspiracy theories are rare, they do occasionally occur. As such, both the short and long form versions of the CTAC were designed so participants were not encouraged to always reject the conspiracy, but to consider the most plausible outcome with the information that was made available. This differs from measures such as the GCBS where participants are scored on the degree in which they reject all items in the assessment (Brotherton & French, 2014). The CTAC allows researchers to measure ability to discern between conspiracy theories that are plausible or implausible. We argue that this provides researchers with a more useful understanding of how people approach conspiracy theories, in comparison to the degree in which participants believe certain statements.

Another advantage of the CTAC is that it is not bound by cultural relevance. A common critique of standardised measures of conspiratorial thinking is that the content of items may be plausible in one country, but not in others. This critique was common among the feedback we received from participants, and was consistent with critiques previously

made by other researchers towards standardised measures of conspiracy ideation (Bruder et al., 2013). Bruder and colleagues previously critiqued the GCBS, remarking that endorsing the statements in that scale may be a good indicator of conspiracy ideation in a democratic country, but they may in fact be indicators of healthy scepticism in an autocratic country (2013). We argue that the CTAC sidesteps these constraints by using fictional scenarios, providing enough information to avoid ambiguity about the scope of the conspiracy (another common critique among participants) and preventing participants from having to interpret broad and vague conspiracy statements.

The CTAC is not without concerns or limitations. A possible limitation of the CTAC is the length of the items. While most assessments such as the Conspiracy Mentality Scale (Bruder et al., 2013) and Generic Conspiracy Beliefs Scale (Brotherton et al., 2013) use single-sentence items, the CTAC items typically range between 80-100 word paragraphs. While the length of these items was necessary to include enough information to allow participants to make an informed decision, there is a concern that the length of the assessment may fatigue or overload participants, particularly those with poor literacy (Stanton et al., 2002). In addition to the wordcount per item, the complexity of the items in the CTAC may also expect a considerable amount of attention and working memory to complete, as these items require more deliberate thinking than simply rating how much someone agrees with particular conspiracy theories. Another limitation of the CTAC is the concern of how the various subject matter in the assessment may bias participants' answers. Items within the CTAC cover a number of topics such as sports conspiracies, financial misconduct, and academia-related scandals. It is possible that those more familiar with some of these topics would have more familiarity with these types of scenarios and would be better equipped to critically appraise these items more than items related to topics less familiar to them. However, any personal familiarity with a single item would be counterbalanced by the other

topics in the assessment. Finally, due to the small number of items within the CTAC we found that while the implausible factor showed good internal reliability, the plausible factor showed a lower internal reliability. This was a deliberate trade off, as we wanted to prevent cognitive fatigue in participants that would have reduced the efficacy of this assessment (Stanton et al., 2002).

In summary, the CTAC is an assessment that assesses the process of endorsing or rejecting conspiracy beliefs, rather than simply measuring the belief itself. The studies in this paper have demonstrated the reliability and utility of this assessment as a novel means of measuring conspiratorial thinking which we propose will aid future conspiracy interventions by introducing a more nuanced and sensitive measure of how individuals approach conspiracy theories.

3.8 References

- Adam-Troian, J., Caroti, D., Arciszewski, T., & Ståhl, T. (2019). Unfounded beliefs among teachers: The interactive role of rationality priming and cognitive ability. *Applied Cognitive Psychology*, 33(4), 720–727. <https://doi.org/10.1002/acp.3547>
- Alper, S., Bayrak, F., & Yilmaz, O. (2021). Psychological correlates of COVID-19 conspiracy beliefs and preventive measures: Evidence from Turkey. *Current Psychology*, 40(11), 5708–5717. <https://doi.org/10.1007/s12144-020-00903-0>
- Bago, B., Rand, D. G., & Pennycook, G. (2022). Does deliberation decrease belief in conspiracies? *Journal of Experimental Social Psychology*, 103, 104395. <https://doi.org/10.1016/j.jesp.2022.104395>
- Bezalel, G. Y. (2021). Conspiracy Theories and Religion: Reframing Conspiracy Theories as *Bliks*. *Episteme*, 18(4), 674–692. <https://doi.org/10.1017/epi.2019.46>

- Bogart, L. M., Wagner, G., Galvan, F. H., & Banks, D. (2010). Conspiracy Beliefs About HIV Are Related to Antiretroviral Treatment Nonadherence Among African American Men With HIV. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 53(5), 648–655. <https://doi.org/10.1097/QAI.0b013e3181c57dbc>
- Bonetto, E., Troïan, 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>
- Bost, P. R., Prunier, S. G., & Piper, A. J. (2010). Relations of Familiarity with Reasoning Strategies in Conspiracy Beliefs. *Psychological Reports*, 107(2), 593–602. <https://doi.org/10.2466/07.09.17.PR0.107.5.593-602>
- Brotherton, R., & French, C. C. (2014). Belief in Conspiracy Theories and Susceptibility to the Conjunction Fallacy. *Applied Cognitive Psychology*, 28(2), 238–248. <https://doi.org/10.1002/acp.2995>
- 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>
- Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring Individual Differences in Generic Beliefs in Conspiracy Theories Across Cultures: Conspiracy Mentality Questionnaire. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00225>
- Chen, H., Akbarzai, S., & Khalidi, A. (2022, July 1). *More than 300,000 killed in decade of Syrian conflict, Un says*. CNN. <https://edition.cnn.com/2022/07/01/world/syria-war-united-nations-report-intl-hnk/index.html>

- Darwin, H., Neave, N., & Holmes, J. (2011). Belief in conspiracy theories. The role of paranormal belief, paranoid ideation and schizotypy. *Personality and Individual Differences*, 50(8), 1289–1293. <https://doi.org/10.1016/j.paid.2011.02.027>
- Douglas, K. M., & Sutton, R. M. (2008). The Hidden Impact of Conspiracy Theories: Perceived and Actual Influence of Theories Surrounding the Death of Princess Diana. *The Journal of Social Psychology*, 148(2), 210–222. <https://doi.org/10.3200/SOCP.148.2.210-222>
- Douglas, K. M., & Sutton, R. M. (2011). Does it take one to know one? Endorsement of conspiracy theories is influenced by personal willingness to conspire. *British Journal of Social Psychology*, 50(3), 544–552. <https://doi.org/10.1111/j.2044-8309.2010.02018.x>
- 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>
- 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/s11162-018-9513-3>
- Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4), 25–42. <https://doi.org/10.1257/089533005775196732>
- 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>

- Georgiou, N., Delfabbro, P., & Balzan, R. (2019). Conspiracy beliefs in the general population: The importance of psychopathology, cognitive style and educational attainment. *Personality and Individual Differences*, 151, 109521. <https://doi.org/10.1016/j.paid.2019.109521>
- Gligorić, V., Da Silva, M. M., Eker, S., Van Hoek, N., Nieuwenhuijzen, E., Popova, U., & Zeighami, G. (2021). The usual suspects: How psychological motives and thinking styles predict the endorsement of well-known and COVID -19 conspiracy beliefs. *Applied Cognitive Psychology*, 35(5), 1171–1181. <https://doi.org/10.1002/acp.3844>
- Golec De Zavala, A., & Lantos, D. (2020). Collective Narcissism and Its Social Consequences: The Bad and the Ugly. *Current Directions in Psychological Science*, 29(3), 273–278. <https://doi.org/10.1177/0963721420917703>
- Greene, C. M., & Murphy, G. (2023). Debriefing works: Successful retraction of misinformation following a fake news study. *PLOS ONE*, 18(1), e0280295. <https://doi.org/10.1371/journal.pone.0280295>
- Grimes, D. R. (2016). On the Viability of Conspiratorial Beliefs. *PloS One*, 11(1), e0147905. <https://doi.org/10.1371/journal.pone.0147905>
- 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>
- Halpern, D. (2012). *HCTA Halpern Critical Thinking Assessment* [dataset]. <https://doi.org/10.1037/t10940-000>
- 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>

- Imhoff, R., Dieterle, L., & Lamberty, P. (2021). Resolving the Puzzle of Conspiracy Worldview and Political Activism: Belief in Secret Plots Decreases Normative but Increases Nonnormative Political Engagement. *Social Psychological and Personality Science*, 12(1), 71–79. <https://doi.org/10.1177/1948550619896491>
- Iran: *Blood-soaked secrets: Why Iran's 1988 prison massacres are ongoing crimes against humanity*. Amnesty International. (2021, December 9). <https://www.amnesty.org/en/documents/mde13/9421/2018/en/>
- 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>
- Koroma, D., Pestalozzi, M. I., & Znoj, H. (2022). How Social Exclusion, Embitterment, and Conspiracy Beliefs Mediate Individual's Intention to Vaccination against COVID-19: Results from a Moderated Serial Mediation Analysis. *Psychopathology*, 55(2), 93–103. <https://doi.org/10.1159/000521016>
- Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790. <https://doi.org/10.1073/pnas.1320040111>
- Lantian, A., Bagneux, V., Delouvée, S., & Gauvrit, N. (2021). Maybe a free thinker but not a critical one: High conspiracy belief is associated with low critical thinking ability. *Applied Cognitive Psychology*, 35(3), 674–684. <https://doi.org/10.1002/acp.3790>

- Lantian, A., Muller, D., Nurra, C., Klein, O., Berjot, S., & Pantazi, M. (2018). Stigmatized beliefs: Conspiracy theories, anticipated negative evaluation of the self, and fear of social exclusion. *European Journal of Social Psychology*, 48(7), 939–954.
<https://doi.org/10.1002/ejsp.2498>
- 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>
- 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>
- Nunnally, J. C. (1978). *Psychometric theory* (Second). McGraw-Hill.
<https://go.exlibris.link/76m5ls8j>
- Nursa'adah, E., Sari, D., & Affifah, I. (2021). Domain-specific critical thinking skills: Designing and analysis the instrument test on thermochemistry. *Journal of Physics: Conference Series*, 1731(1), 12025. <https://doi.org/10.1088/1742-6596/1731/1/012025>
- 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>
- Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2020). On the belief that beliefs should change according to evidence: Implications for conspiratorial, moral, paranormal, political, religious, and science beliefs. *Judgment and Decision Making*, 15(4), 476–498. <https://doi.org/10.1017/S1930297500007439>

- 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/fpsyt.2020.568942>
- Rizeq, J., Flora, D. B., & Toplak, M. E. (2021). An examination of the underlying dimensional structure of three domains of contaminated mindware: Paranormal beliefs, conspiracy beliefs, and anti-science attitudes. *Thinking & Reasoning*, 27(2), 187–211. <https://doi.org/10.1080/13546783.2020.1759688>
- Sirota, M., & Juanchich, M. (2018). Effect of response format on cognitive reflection: Validating a two- and four-option multiple choice question version of the Cognitive Reflection Test. *Behavior Research Methods*, 50(6), 2511–2522.
<https://doi.org/10.3758/s13428-018-1029-4>
- Soveri, A., Karlsson, L. C., Antfolk, J., Lindfelt, M., & Lewandowsky, S. (2021). Unwillingness to engage in behaviors that protect against COVID-19: The role of conspiracy beliefs, trust, and endorsement of complementary and alternative medicine. *BMC Public Health*, 21(1), 684. <https://doi.org/10.1186/s12889-021-10643-w>
- Ståhl, T., & Van Prooijen, J.-W. (2018). Epistemic rationality: Skepticism toward unfounded beliefs requires sufficient cognitive ability and motivation to be rational. *Personality and Individual Differences*, 122, 155–163. <https://doi.org/10.1016/j.paid.2017.10.026>
- Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). ISSUES AND STRATEGIES FOR REDUCING THE LENGTH OF SELF-REPORT SCALES.

- Personnel Psychology*, 55(1), 167–194. <https://doi.org/10.1111/j.1744-6570.2002.tb00108.x>
- Stojanov, A. (2015). Reducing conspiracy theory beliefs. *Psihologija*, 48(3), 251–266. <https://doi.org/10.2298/PSI1503251S>
- Stojanov, A., & Halberstadt, J. (2019). The Conspiracy Mentality Scale: Distinguishing Between Irrational and Rational Suspicion. *Social Psychology*, 50(4), 215–232. <https://doi.org/10.1027/1864-9335/a000381>
- Swami, V., Chamorro-Premuzic, T., & Furnham, A. (2010). Unanswered questions: A preliminary investigation of personality and individual difference predictors of 9/11 conspiracist beliefs. *Applied Cognitive Psychology*, 24(6), 749–761. <https://doi.org/10.1002/acp.1583>
- Swami, V., Pietschnig, J., Tran, U. S., Nader, I. W., Stieger, S., & Voracek, M. (2013). Lunar Lies: The Impact of Informational Framing and Individual Differences in Shaping Conspiracist Beliefs About the Moon Landings. *Applied Cognitive Psychology*, 27(1), 71–80. <https://doi.org/10.1002/acp.2873>
- 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>
- Tiruneh, D. T., De Cock, M., Weldelessie, A. G., Elen, J., & Janssen, R. (2017). Measuring Critical Thinking in Physics: Development and Validation of a Critical Thinking Test in Electricity and Magnetism. *International Journal of Science and Mathematics Education*, 15(4), 663–682. <https://doi.org/10.1007/s10763-016-9723-0>
- Van Der Wal, R. C., Sutton, R. M., Lange, J., & Braga, J. P. N. (2018). Suspicious binds: Conspiracy thinking and tenuous perceptions of causal connections between co-

- occurring and spuriously correlated events. *European Journal of Social Psychology*, 48(7), 970–989. <https://doi.org/10.1002/ejsp.2507>
- Van Prooijen, J., Douglas, K. M., & De Inocencio, C. (2018). Connecting the dots: Illusory pattern perception predicts belief in conspiracies and the supernatural. *European Journal of Social Psychology*, 48(3), 320–335. <https://doi.org/10.1002/ejsp.2331>
- Van Prooijen, J.-W. (2019). Belief in Conspiracy Theories. In J. P. Forgas & R. F. Baumeister (Eds.), *The Social Psychology of Gullibility* (1st ed., pp. 319–332). Routledge. <https://doi.org/10.4324/9780429203787-17>
- 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>
- Vitriol, J. A., & Marsh, J. K. (2018). The illusion of explanatory depth and endorsement of conspiracy beliefs. *European Journal of Social Psychology*, 48(7), 955–969. <https://doi.org/10.1002/ejsp.2504>
- Williams, M. N., Marques, M. D., Hill, S. R., Kerr, J. R., & Ling, M. (2022). Why are beliefs in different conspiracy theories positively correlated across individuals? Testing monological network versus unidimensional factor model explanations. *British Journal of Social Psychology*, 61(3), 1011–1031. <https://doi.org/10.1111/bjso.12518>
- 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>

Chapter 4: True discernment or blind scepticism? Comparing the effectiveness of four conspiracy belief interventions

This paper has been published in *Advances in Psychology*.

O'Mahony, C., Murphy, G., & Linehan, C. (2024). True discernment or blind scepticism?

Comparing the effectiveness of four conspiracy belief interventions.

advances.in/psychology, 2, e215691. <https://doi.org/10.56296/aip00030>

Study Motivation

In Chapter 2, my systematic review identified several promising interventions for challenging unfounded conspiracy beliefs. However, these different forms of conspiracy interventions

have never been directly compared to one another in an experimental study. In addition, building on Chapter 3, we noted that most experimental conspiracy intervention studies only measure endorsement of unfounded conspiracy theories as an outcome measure rather than measuring discernment between plausible and implausible conspiracy theories aimed to address this limitation use the Critical Thinking about Conspiracies assessment (CTAC) in an experimental study as more appropriate measure of the underlying cognitive processes of conspiracy thinking by measuring how participants assess plausible and implausible conspiracy theories. The current research takes four different conspiracy interventions (Priming, Inoculation, Active Inoculation, and Discernment) and compares their effectiveness in terms of increasing critical appraisal of conspiracy theories.

Author Contributions

COM, CL, and GM developed the design and concept of this research. Survey design, data collection, data analysis, data interpretation, and data visualisation were conducted by COM under the supervision of GM. COM wrote the manuscript and revised it in response to GM and CL's feedback.

4.1 Abstract

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 $N = 1766$), 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 significantly 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 ($d = 0.14 - 0.72$). These findings highlight both the importance of teaching discernment and measuring discernment as an outcome.

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 cooperation 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.

4.1.1 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; Većkalov 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).

4.1.2 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.

4.1.3 Overview of Present Study

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.

4.2 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, Priming, Inoculation, Active Inoculation), we expected that critical appraisal of conspiracy theories will increase.

4.2.1 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/?view_only=73313f722b8546ad991b83fccd86c97e

4.2.1.1 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.

4.2.1.2 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 preregistered 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; Mean age = 41.68, $SD = 13.86$). 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%).

4.2.1.3 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 who 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 4.1). 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 (Banas & Miller, 2013; Stall & Petrocelli, 2023), (iii) identifying logical leaps (Pytlik 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.

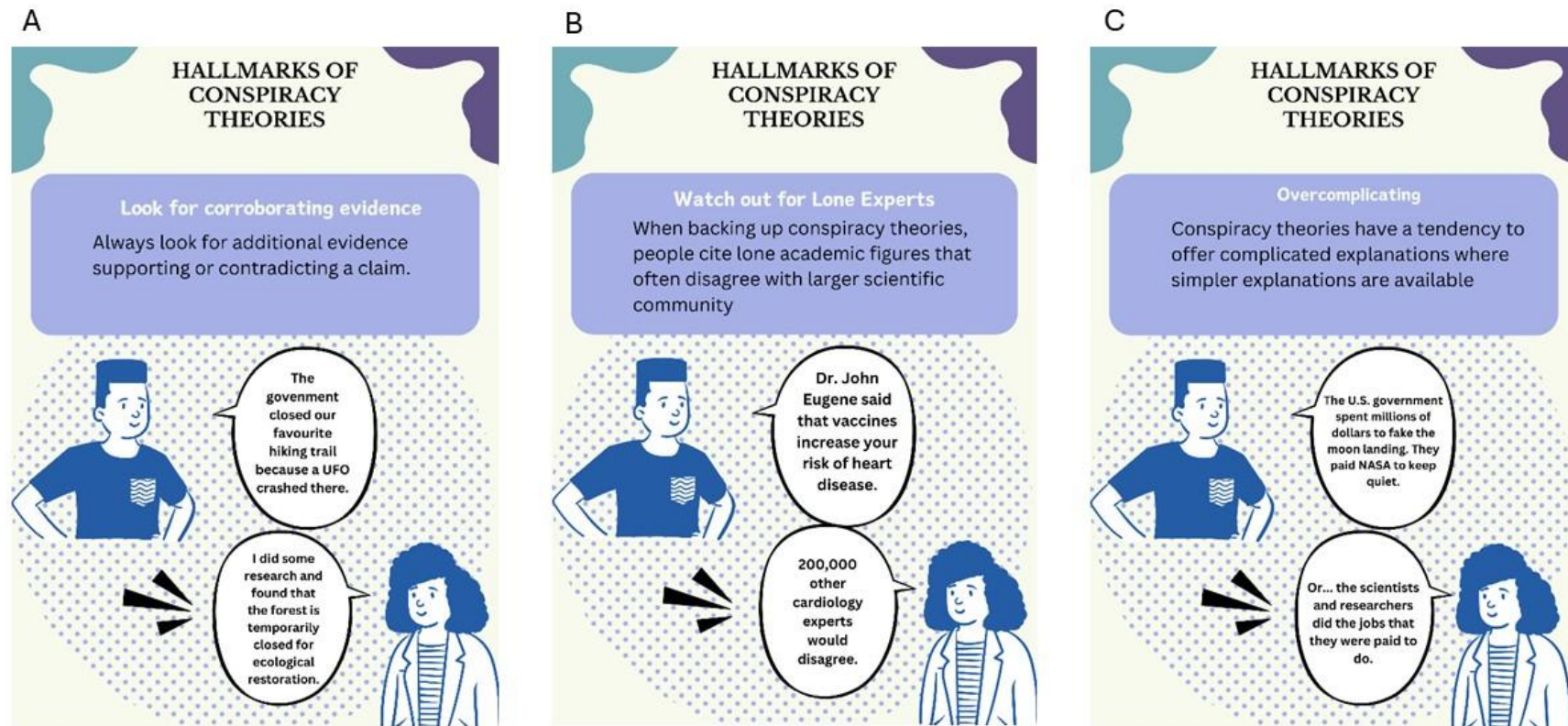


Figure 4.1: Three of the five inoculation infographics used in the Inoculation and Active conditions, relating to (a) corroborating evidence, (b) lone experts, and (c) overcomplication. The two other infographics related to identifying the “who benefits” fallacy and logical leaps can be seen in our Supplementary Online Materials.

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 $\alpha = .88$.

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.

4.2.1.4 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). 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:

1. Post-Introduction: Immediately following the introduction phase of the Qualtrics survey, participants in the Control and Priming conditions received a nonsensical attention check where participants were asked to agree with the following statement: “*I swim across the Atlantic Ocean to get to work every day*”.

2. Infographics: During the infographic section in the Inoculation and Active conditions, participants were presented with two comprehension checks related to two of the five infographics to ensure that they had read the infographics properly (e.g. asking participants to describe what the “who benefits fallacy” was, immediately after they had seen the related infographic). The Control and Priming conditions did not receive this attention check as they did not see any infographics.
3. CTAC: Every participant completed an attention check during the CTAC section, which was presented in a random order amongst the CTAC items. We wanted to ensure that participants were properly reading the vignettes, as due to the length and cognitive demand of the items we predicted that some participants may become fatigued and skim these items.

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 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.

4.2.1.5 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.

4.2.2 Results

4.2.2.1 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. In summary, 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).

4.2.2.2 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.

4.2.3 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 Supplementary Materials:

https://osf.io/suw57/?view_only=73313f722b8546ad991b83fccd86c97e) 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.

4.3 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.

4.3.1 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, Mean age = 38.05, $SD = 13.02$) 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%).

4.3.2 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 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 ($\alpha = .85$).

4.3.3 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”.

4.3.3.1 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 instead were asked the following after reading the Discernment message:

What is the key point of the paragraph you just read?

- *Conspiracy theories should always be dismissed (0)*
- *Critically thinking about conspiracy theories is not the same as blind scepticism (1)*
- *Most conspiracy theories turn out to be true, so you should generally believe them (0)*
- *Only trust conspiracy theories if they come from mainstream media outlets (0)*

This comprehension check was used to ensure participants had read and understood the Discernment message. Finally, like all other conditions, participants in the Discernment conditions received an attention check during the CTAC segment of the survey. Similar to Study 1a, participants were excluded if they answered at least one infographic attention check incorrectly, or if they failed at least two or more of any attention check.

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

4.3.4 Results

4.3.4.1 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.

4.3.4.2 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 4.2, 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. 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

(https://osf.io/suw57/?view_only=73313f722b8546ad991b83fccd86c97e).

Additionally, 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 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 ($F(4, 1170) = 2.28, p = .059, \eta^2 < 0.01$).

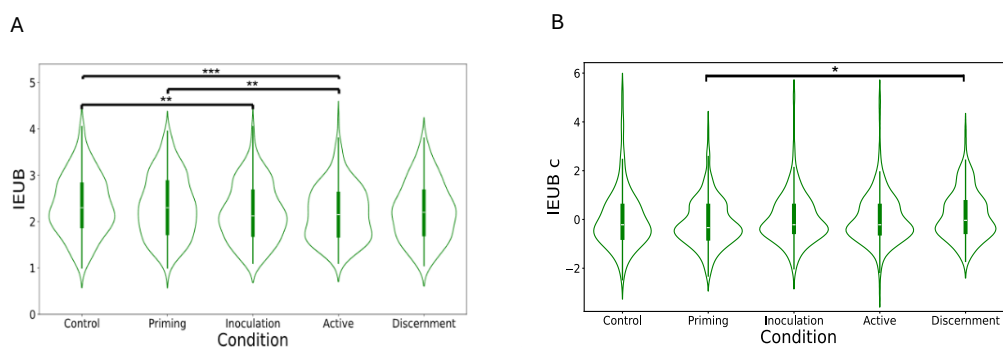


Figure 4.2: Violin plots showing the kernel densities of A) epistemically unwarranted beliefs, and B) c scores for epistemically unwarranted beliefs to estimate response bias

4.3.5 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.

4.4 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. In addition, we included two additional measures of conspiracy thinking to more broadly assess the effectiveness of the interventions. 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.

4.4.1 Methods

4.4.1.1 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, mean age=41.26, $SD = 13.31$). 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.

4.4.1.2 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 Appendix). 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 very good ($\alpha = .78$). 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”* and 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.

4.4.1.3 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.

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 number of attention checks were responsible for this slight deviation between groups.

4.4.2 Results

As shown in Table 4.1, 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 judgements of plausible and implausible conspiracy theories (JICT). More accurate likelihood judgements (JICT) were also associated with lower levels of general conspiracy ideation (GCBS).

Table 4.1 *Correlations for Dependent Variables*

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

Note. ** $p < .001$

4.4.2.1 Did the interventions improve critical appraisal of conspiracy theories?

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 4.3, 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 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 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.

4.4.2.2 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 4.4c, 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.

4.4.2.3 Did the interventions improve likelihood ratings of plausible and implausible conspiracy theories?

As seen in Figure 4.4a, 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 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. 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 ($F(4, 542) = 3.00, p = .018, \eta^2 = 0.02$). Those in the Inoculation condition were significantly less inclined to rate plausible conspiracy theories as being likely than those in the Priming ($p = .049, d = -0.37$) 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.

As seen in Figure 4.4b we calculated d' for the JICT to calculate whether participants could discern 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

(https://osf.io/suw57/?view_only=73313f722b8546ad991b83fccd86c97e).

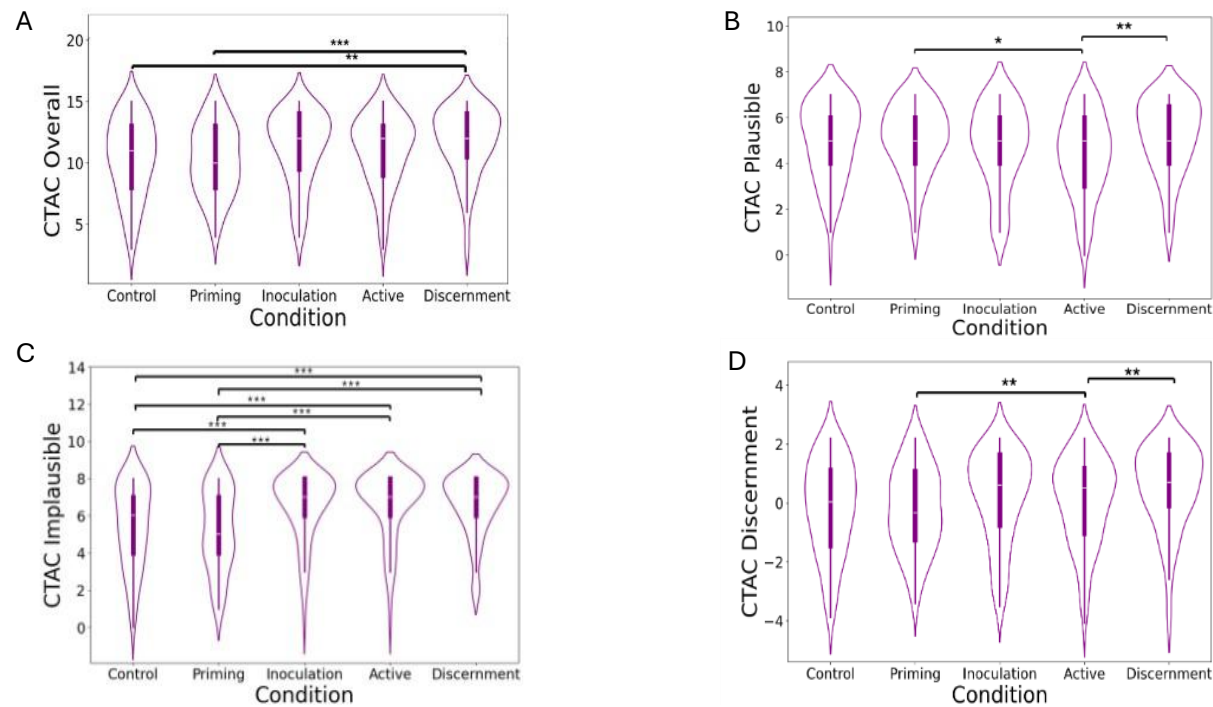


Figure 4.3: Violin plots showing the Study 2 kernel density distribution of A) overall critical appraisal of conspiracy theories (CTAC Long Form Total), B) critical appraisal of plausible conspiracy theories (CTAC Long Form Plausible), C) critical appraisal of implausible conspiracy theories (CTAC Long Form Implausible), and D) critical discernment between plausible and implausible conspiracy theories (CTAC Long Form d')

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

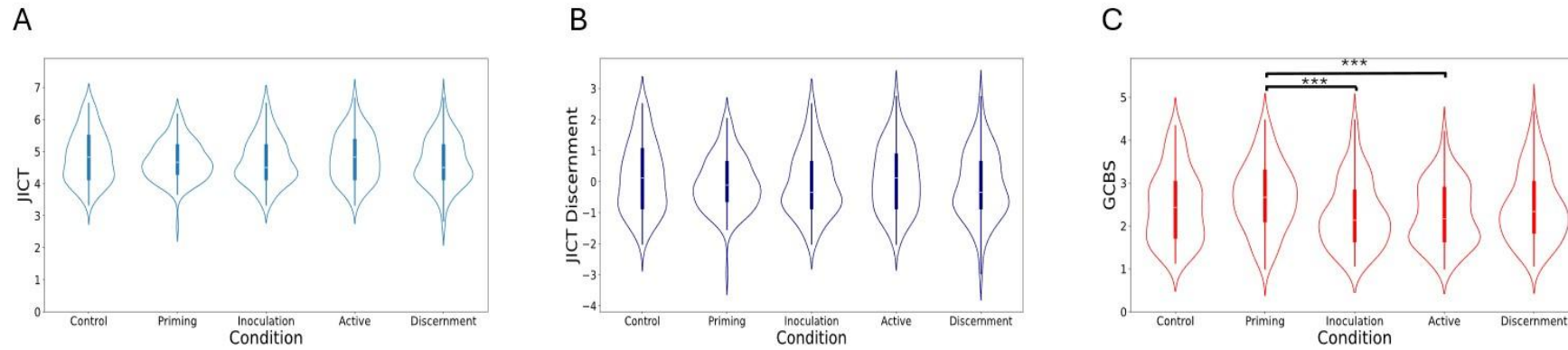


Figure 4.4: Violin plots showing the kernel density distribution of A) conspiracy judgements of the likelihood of plausible and implausible conspiracy theories (JICT), B) discernment between plausible and implausible conspiracy theories (JICT d'), and C) general conspiracy ideation (GCBS).

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

4.5 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 and 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). 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; Većkalov 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, in contrast to 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). This 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). This is not to say that these interventions may result in a *backfire effect*, where participants become more emboldened in their beliefs after they have been challenged (Nyhan and Reifler, 2013). Instead, we suggest that these interventions are reducing a response bias that encourages participants to identify when something is a conspiracy theory and reject it solely on that basis. 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 "overcorrection" 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. We also note that the main effects of intervention types on all outcome measures were small. 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. Our 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 (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. It may be that, therefore, 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 study. 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.

4.6 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., Troïan, 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/s11162-018-9513-3>
- 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>
- Förster, 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://ucc.idm.oclc.org/login?URL=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2021-55332-001&site=ehost-live>

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>
- 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., Bőthe, 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>
- Pytlík, 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/fpsy.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>

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>

Chapter 5: Designing a serious game to increase critical appraisal of conspiracy theories

Authors: Cian James O'Mahony, Aaron Bolger, Conor Linehan, Gillian Murphy

Study Motivation

In Chapter 4 we identified a novel intervention that could improve participants ability to critically appraise both plausible and implausible conspiracy theories. In this paper, we sought to take the intervention techniques identified in Chapter 4, and package them in a game that would offer a more engaging experience to participants while retaining the educational benefits of successful discernment intervention. We developed a critical thinking game title *Nashville PI* and compared it to a control game to test how effective it was at increasing participants critical appraisal of conspiracy theories.

Author Contributions

COM, CL, and GM developed the design and concept for this research. COM was responsible for creating the paper prototypes, interactive digital prototypes, concept art, playtesting, and game mechanic design, with feedback and supervision from CL and GM. COM also wrote the script, narrative, and concept for Nashville PI, and integrated the writing, dialogue, and game variables into the JSON format. AB handled the graphic design, coding, and game hosting, while also providing feedback to COM on the game's design and mechanics. Survey design, data collection, analysis, and interpretation were carried out by COM, under GM's supervision. Finally, COM wrote and revised the manuscript based on feedback from GM and CL.

5.1 Abstract

Building upon the findings in the previous chapter, we aimed to package the novel discernment intervention we developed into an enjoyable interactive game. In this Chapter, we designed *Nashville PI*, an innovative online critical-thinking game inspired by research on serious games for combating misinformation. In the game, players assume the role of a PR assistant managing the social media account of a country music star attempting a career comeback - who also tends to share conspiratorial posts. Players learn to critically evaluate and prevent the posting of unfounded conspiracy theories. To assess the game's effectiveness, we measured participants' critical appraisal skills using the Critical Thinking about Conspiracies Assessment (CTAC). Our evaluation found that while players enjoyed the narrative and concept of the game, several mechanics limitations prevented the game from being effective as an intervention to increase critical appraisal of conspiracy theories. Our empirical findings suggested that those who played *Nashville PI* were significantly worse at appraising conspiracy theories than those who had played a control game (*Tetris*). As the concept of the game shows promise, we examine the factors that may have limited the game's efficacy as an intervention and outline future amendments that will address these shortcomings.

5.2 Introduction

In the previous chapter, I demonstrated how a brief 10-minute intervention could effectively increase critical discernment between plausible and implausible conspiracy theories (O'Mahony et al., 2024). The interventions in the previous chapter indicated that pre-emptive inoculation interventions are not just successful in terms of reducing general conspiracy ideation, but that they can be used to teach critical thinking skills that participants can use to evaluate novel conspiracy theories and to improve their ability to distinguish between probable and improbable conspiracy narratives. The results of that study were promising, but one may ask why anyone would choose to participate in these interventions in the first place. As I note in Chapter 2 and 4, understanding and supporting people's motivation to take part in an intervention is essential for ensuring effectiveness. Belief in conspiracy theories is associated with distrust of large institutions, including governments, universities, and higher education institutions (Van Prooijen et al., 2022; Vranic et al., 2022). As those who believe in more unfounded conspiracy theories (i.e., those who would be in need of these interventions the most) would arguably be the least likely to take part in interventions designed by academic researchers voluntarily. Some researchers have noted the possibility of including such interventions as part of existing educational programmes (Kiili et al., 2024). I am interested in designing interventions that people want to engage with, rather than interventions that can be experienced only through some form of coercion (i.e., participating in a study or through formal education). Recent research has shown the promise of using video games as a way of delivering interventions for various forms of misinformation (Basol et al., 2021; Cook et al., 2023; Hu et al., 2023; Roozenbeek & Van Der Linden, 2019, 2020).

5.2.1 Why use games to challenge unfounded conspiracy beliefs?

A recent systematic review examining 15 papers found that various games were predominantly effective as a means of making players less susceptible to misinformation such as fake news, general misinformation, and political radicalism (Kiili et al., 2024). Games provide a promising avenue for interventions for several reasons. The first advantage of using games as an intervention to challenge unfounded beliefs is that games are a means of facilitating intrinsic motivation among players (Linehan et al., 2011). Games have been used as educational materials for some time due to their inherent advantage of using goal-driven behaviour to increase players' intrinsic motivation (Rodriguez-Calzada et al., 2024). As such, games are an effective way to disseminate an intervention where players may be more motivated to participate in these interventions in comparison to standard academic inoculation interventions. Misinformation games, therefore, offer a promising avenue to develop scalable interventions that are inherently more appealing to participants than standard conspiracy interventions.

The second reason for using games as an intervention is that they require players to actively engage with the intervention material in a deliberate manner, in contrast to other styles of interventions where participants passively consume the content (Linehan et al., 2011). Previous research has shown that active game-based interventions are more effective than their passive counterpart in decreasing participants' susceptibility to misinformation (Hu et al., 2023). One explanation as to why game-based interventions are effective is that they facilitate deliberate learning, where players are forced to take theoretical knowledge they have learned and put it into practice (Linehan et al., 2011). Previous research has indicated that active deliberate practice improves intellectual skills far better than passive consumption of information (Ericsson, 2016). For instance, research examining training radiologists found that students and trainees excelled during deliberate practice sessions with viewing and

interpreting scans where they were given explicit feedback on their performance (Pusic et al., 2011). As I discussed in Chapter 2, similar observations have been made with regard to training students to evaluate conspiracy theories, using a three-month university module (Dyer & Hall, 2019). However, unlike a time-consuming university module that would require a considerable amount of commitment from participants, games can offer a cost-effective and time-efficient means of facilitating deliberate practice. While Chapter 4 found that there was no significant difference between active and passive interventions, these findings were in contrast with a larger body of work indicating their efficacy (Banas & Rains, 2010; Cook et al., 2023; Ecker et al., 2022). I also note the possibility that active interventions may have a sleeper effect, which may explain why some active interventions are only seen to be more effective after a delay (McGuire, 1964).

5.2.2 Why design a novel game?

Nearly all of the games previously used in misinformation research are designed as inoculation-based interventions, where the game serves to expose players to a weakened form of the sorts of misinformation they may encounter in real-world scenarios. Misinformation games such as *GoViral!* or *Bad News* achieve this by exposing players to common rhetoric techniques that are used to perpetuate misinformation such as fake news (Basol et al., 2021; Roozenbeek & Van Der Linden, 2019). Unlike topic-specific interventions, misinformation games such as *Bad News* teach players about broad concepts related to misinformation in order to confer a *blanket-of-protection* against various forms of fake news (Roozenbeek & Van Der Linden, 2019). This approach to challenging misinformation is quite similar to the approach we took in Chapter 4 when designing the Inoculation, Active, and Discernment interventions. Games such as the *Bad News* game have also shown success in reducing participants scores on measures of conspiracy beliefs (Roozenbeek & Van Der Linden, 2019). Results of that study are promising and demonstrate the potential for misinformation games

to challenge unfounded conspiracy theories, as well as fake news and other forms of misinformation.

As games such as *HarmonySquare* (Roozenbeek & Van Der Linden, 2020) and *Bad News* (Roozenbeek & Van Der Linden, 2019) have shown success in reducing conspiracy beliefs, this begs the question; why not use these games to challenge conspiracy beliefs? There are several reasons why existing misinformation games are not the ideal method for interventions that challenge conspiracy beliefs. Firstly, misinformation games typically have players deal with information that has been deliberately fabricated rather than misinformation arising from genuine misconceptions or sincerely held beliefs, such as conspiracy theories.

The *Bad News* game focuses on teaching players about the intentional dissemination of inaccurate news, or simply *disinformation* (Roozenbeek & Van Der Linden, 2019). The player learns how to deliberately use manipulation techniques to mislead people online. Similarly, *HarmonySquare* focuses on teaching players how bot accounts, fake likes, and trolling function as a means of amplifying the reach of the player's social media accounts (Roozenbeek & Van Der Linden, 2020). The aim of games such as *HarmonySquare*, *Bad News*, and *GoViral!* is to intentionally increase the player's reach online (Basol et al., 2021; Roozenbeek & Van Der Linden, 2019, 2020). As such, conspiracy theories are presented in these games as narrative tools that have been developed to intentionally mislead people. Indeed, individuals like Donald Trump have used conspiracy theories as political weapons, such as Trump framing accusations against him as a conspiracy orchestrated by Hillary Clinton to tarnish his reputation (Sarlin, 2016). However, this approach to conspiracy theories overlooks the crucial factor that the majority of conspiracy theories are sincere narratives that have been made by people to explain the world in times of uncertainty and ambiguity (Douglas et al., 2017; Stasielowicz, 2022; Wood, 2017). Predominantly the person creating conspiracy narratives and the person believing these narratives are the same person. Many

conspiracy theories have not been deliberately constructed to mislead other people but are conclusions that have been drawn as the results of lapses in critical thinking (). Therefore, educating participants about manipulation techniques may do little to address the common logical errors associated with people constructing their own conspiratorial explanations. As I argue in Chapter 4, it is potentially much more advantageous to tackle the underlying cognitive processes of conspiracy thinking and to teach participants the skills to distinguish plausible and implausible conspiracy theories.

The second limitation of such games is that they don't measure conspiracy ideation appropriately. They predominantly teach players to reject all misinformation rather than teach discernment between information that is likely to be trustworthy and that which is unlikely to be trustworthy. As I noted in Chapter 3, many interventions rely on the generalist assumption that all conspiracy theories are unfounded, and thus teaching players to reject all conspiracy theories is a desirable outcome. *Bad News* uses a single item measure of conspiracy belief as an outcome measure: "*The Bitcoin exchange rate is being manipulated by a small group of rich bankers. #InvestigateNow*" (Roozenbeek & Van Der Linden, 2019). Including a single item measurement is understandable, as the main purpose of *Bad News* was to target fake news as opposed to primarily focusing on conspiracy theories. However, for the purposes of investigating whether games can effectively increase how participants critically evaluate conspiracy theories, more steps are needed.

In this chapter, I propose a serious game that teaches participants how to critically distinguish between plausible and implausible conspiracy theories. The following study consists of a piloting phase, where we used several interactive paper and digital-based interactive prototypes to workshop several concepts and mechanics for this game. We then conducted an empirical evaluation to test how effective this novel game was at increasing participants critical appraisal of plausible and implausible conspiracy theories.

5.2.3 How do you design a useful, successful educational game?

In order to design a game that can successfully teach participants the desired outcomes, we needed to utilise a strong design process. We opted to initially use several low-fidelity prototypes such as paper prototyping and minimal interactive digital prototypes to test the main mechanics of our games. We utilised this approach to developing the early concepts of our games as low-fidelity prototypes serve as a fast and cost-effective way of developing and testing ideas, allow researchers the freedom to quickly amend or change aspects of gameplay that are not producing the desired outcomes among participants (Linehan et al., 2009; Ollila et al., 2008; Snyder, 2003).

We also drew on psychological literature to inform how we designed various mechanics to effectively teach participants the desirable learning outcomes. Previous research has identified several elements that contribute to a game's success as an educational tool. The first is that games need to utilise *intrinsic learning*, where learning outcomes are embedded within the mechanics of the game (Linehan et al., 2011). Many educational games tend to neglect this factor. For instance, the educational game series *MathBlaster* will have players engage in an action-oriented game that is periodically interrupted by math quizzes (Davidson & Associates, 1983). The math problems in *MathBlaster* are removed from the otherwise action-based gameplay, reducing the players sense of intrinsic motivation. In contrast, games that effectively implement intrinsic motivation have been shown to be more motivationally and educationally effective as learning tools. The second important factor in terms of designing a successful educational game is implementing specific and immediate feedback to the player. Research has shown that future performance in a skill is best predicted by how specific feedback is and how soon this feedback is delivered. Finally, an effective educational game should take long and complex skills and break them up into simpler components. With

these factors in mind, we began developing several low fidelity prototypes to test potential concepts for our game

5.3 Game Design

5.3.1 Initial concept

Drawing inspiration from the infographic interventions created in Chapter 4, along with the critical thinking skills identified in Chapter 3 we outlined several main skills we wanted to teach participants: *validating sources* (Banas & Miller, 2013; Cook et al., 2017), *identifying the most probable among many explanations for an event* (Banas & Miller, 2013; Stall & Petrocelli, 2023), *identifying corroborating or conflicting evidence for a conspiracy theory* (Stall & Petrocelli, 2023), and *training participants not to jump to conclusions before exploring alternative explanation* (Pytlik et al., 2020; Roozenbeek et al., 2022). With this initial list of desirable skills in mind, we designed several concepts for a game that would adequately teach these skills, using paper prototypes. We decided to draw inspiration from detective games, as we felt that this genre of game was best suited to situate the skills we were trying to teach with our game. Researcher Clara Fernandez-Vara (2013), who has written extensively on the topic of detective games, says that detective games such as those in the *Sherlock Holmes* series are characterised by *examining and interpreting evidence*, and *modelling insight thinking*. We felt that these two components of detective games coaligned with the main components we identified in Chapter 3 as critical thinking skills associated with discerning plausible and implausible conspiracy theories.

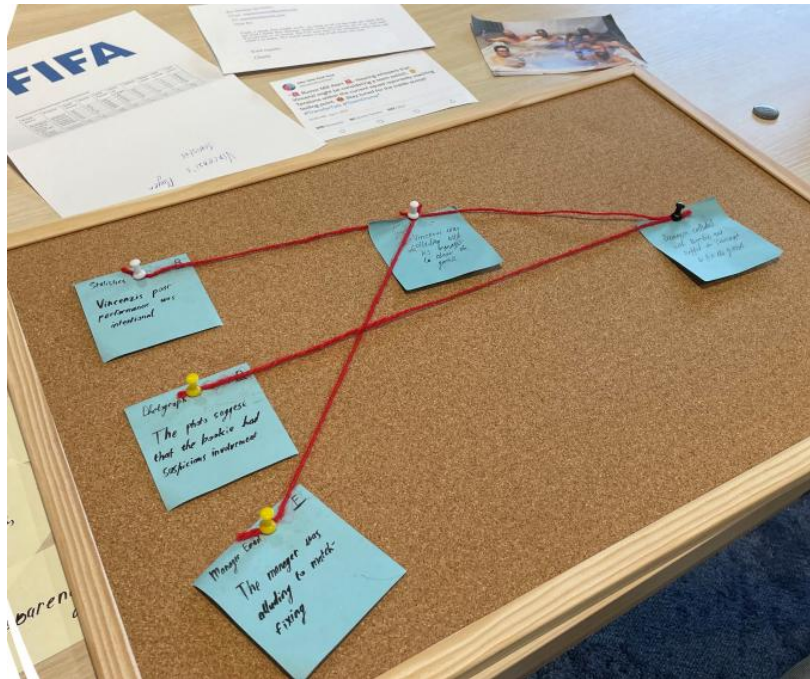


Figure 5.1: Paper prototype of the deduction-based concept of the game. The player could interpret clues in one of two ways and use strings to combine clues to make deductions

We prototyped several concepts for the game, including a deduction-based game akin to the *Sherlock Holmes* series, where players take the role of an internet detective, combining separate clues to come to a conclusion about various conspiracy theories (See Figure 5.1). Another concept we workshopped was a game where players take the role of an investigative journalist who must use their time as a finite resource to verify and check their sources, in order to beat their rivals and publish a sensational but robust and well-researched story. Finally, we designed a contradiction-based game, drawing inspiration from games such as *Ace Attorney* (Capcom, 2001), where players must tease out “fact and relevant data from the mass of contradictory and misleading evidence”. The game involved the player arguing with a family member over the Christmas Holidays about whether a casualty during a stage production was due to foul play or merely an accident. This contradiction-based game revolved around the simple mechanic of the player being confronted by their relative with a

statement, and the player must find evidence that either supports or contradicts this statement. Playtesting found that the contradiction-style gameplay was the most enjoyable among participants, and we found that the mechanics of the contradiction game were much more intuitive to players than the journalist and deduction-based games we had prototyped.

5.3.1.1 Arguing with your Uncle Online

After several rounds of paper prototyping, we decided to move forward with the ‘Arguing with your relatives’ concept. In the game, players interact with their fictional uncle online, who has a penchant for publicly sharing his conspiratorial beliefs with the internet. The player’s family is concerned about the uncle’s image, as they fear it may impact the success of the uncle’s business. As a result, the player is tasked with “doing damage control” and trying to persuade their uncle not to post unfounded conspiracy theories online.

The fundamental mechanics of this game are simple. The player is confronted with an argument or statement from the *interlocutor* (the uncle in this version) through a social media post, and the player is tasked with searching for evidence that either contradicts or supports the interlocutor’s claims. The player must select which part of the interlocutor’s argument they disagree with (*observation*), and once the player feels they have collected enough evidence, they then construct an answer. Players pick what type of argument they wish to present to the interlocutor (Do they think there is a conflict of interest? Do they think the information is from an unreliable source?). As seen in Figure 5.2, the player then responds to the interlocutor by replying to their social media post. The interlocutor then replies, indicating whether the player has successfully convinced him (correctly applied critical thinking skills) or has been unconvincing (incorrectly applied critical thinking skills). Each argument that the player engages with in a level is treated as one unit, or a “block”. Levels consisted of typically three to four blocks before the level concludes.

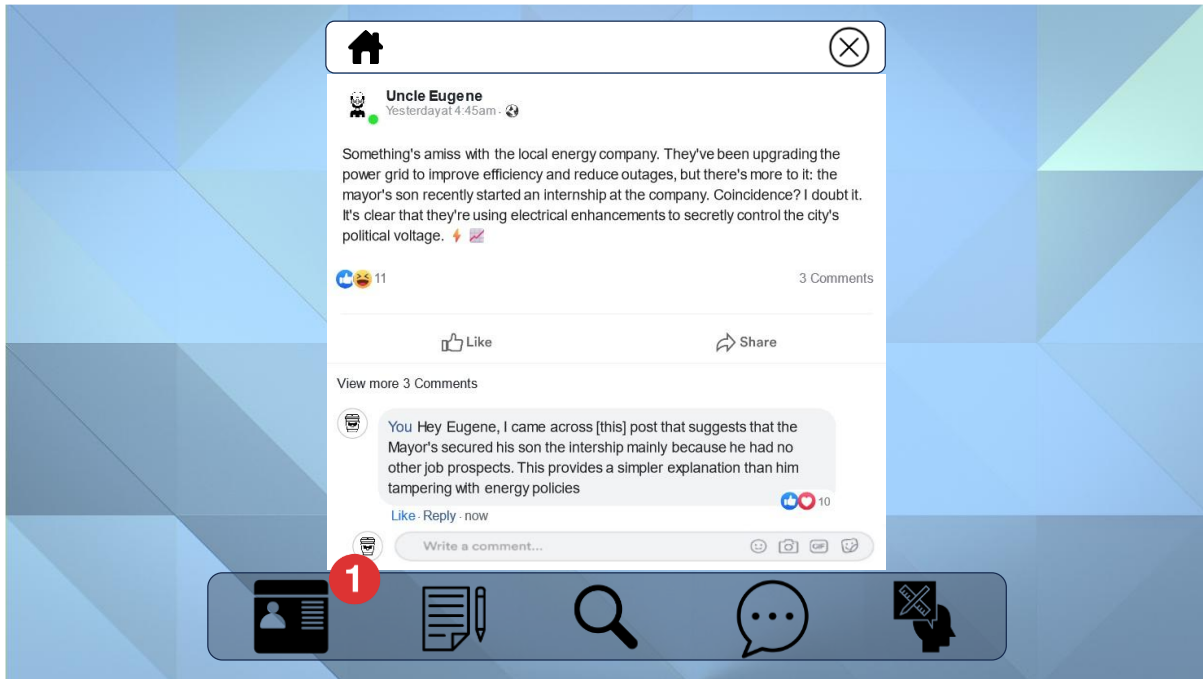


Figure 5.2: Screenshot of interactive PowerPoint prototype for 'Arguing with Your Uncle Online'. The game interface resembles a desktop environment. Different apps such as a browser and notepad (toolbar at bottom) can be used in their investigations. The player can be seen interacting with their fictional uncle through a social media application.

Previous research on educational game design shows that the behaviours a game wants to encourage should match the ones players actually use in the game, not similar substitute behaviours (Linehan et al., 2011). As such, we intended for the game to mimic the real-world skills we were aiming to teach as much as possible. We were aiming to improve player's ability to critically appraise conspiracy theories online. As seen in Figure 5.3, we designed the game's interface to mimic a desktop environment, which applications that mimic real-world computer functions such as a browser, notes app, and a social media/messaging system to interact with the interlocutor.

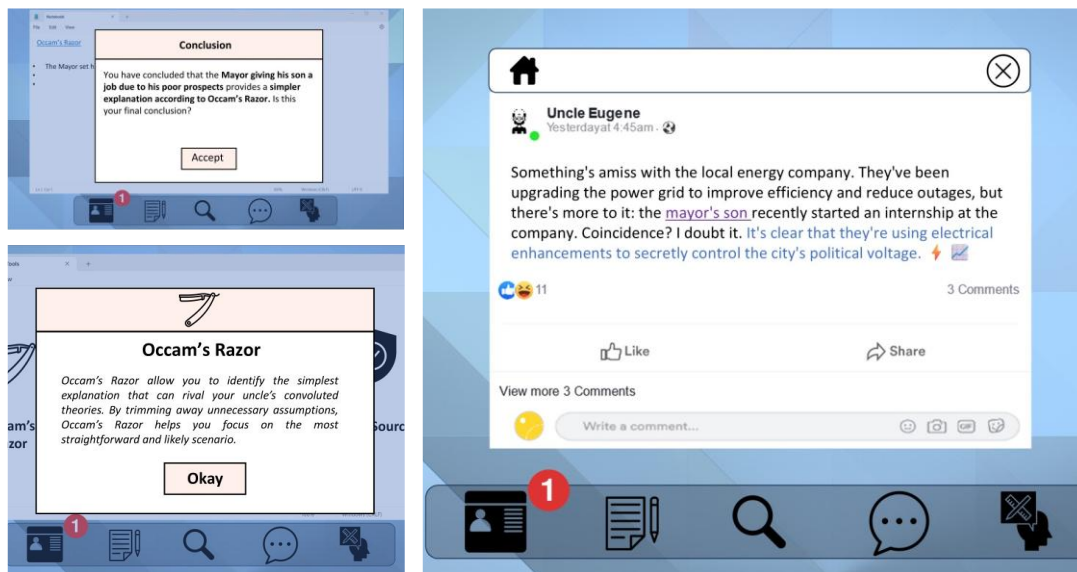


Figure 5.3: Interactive PowerPoint prototype of Arguing with Your Uncle Online. Players can interact with Uncle Eugene using the social media app (right). Clicking purple text will allow players to search for clues in the game's browser application. Clicking on blue text adds the text as a note in their notes app. Players can read descriptions of the various critical thinking skills the game teaches (bottom left). Players are shown a preview of their final arguments before responding (top left)

We then needed to decide how the player would interact with the game. As mentioned previously we intended to design the game environment so that it would mimic the real-world behaviours we intended to teach as closely as possible (Linehan et al, 2011). The game would need some way for the players to interact with clues and evidence and would also need a way to create their counterargument to the interlocutor. We decided that using text parsing (i.e., having the player type in their response, and the game interpreting the text) would not be scalable and would be prone to player typo errors. Instead, we examined how previous games have handled players collecting clues in a desktop environment. We ultimately drew inspiration from a digital investigation section from the game *Life is Strange: True Colors*

(Deck Nine, 2021), which can be seen in Figure 5.4. The player is shown a desktop environment with two applications: an email application, and a notes app. Certain pieces of text are highlighted so the player can click on them and add them to their notes app to answer questions they have during their investigation. We decided this was a way of making clues clear to players and also made the handling of clues and evidence more scalable.

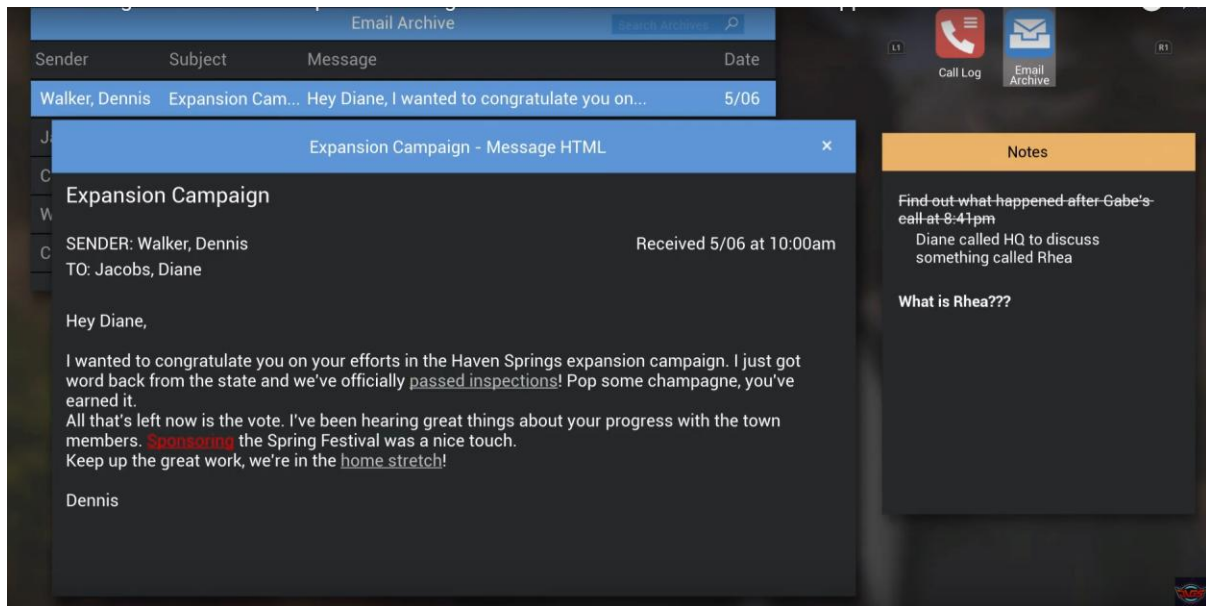


Figure 5.4: UI concept art from *Life is Strange: True Colors* (designed by Jan-Wah Li). Players search for clues by clicking on highlighted pieces of text that are then added to your notes app on the right. We drew inspiration from this design to determine how we handled visually representing clues and notes in the game.

I took advice from COL and AB on developing low-fidelity interactive prototypes, I developed several interactive prototypes using PowerPoint with interactive hyperlinks to simulate the functions of the game. Once we were satisfied with the overall presentation of the game and the overall gameplay loop, we built a working prototype of a single level using the Godot game engine. Screenshots of the Godot prototype can be seen in Figures 5.5 and 5.6. We conducted several pilot playtesting sessions with *Arguing with Your Uncle Online*. We found that the general mechanics of the game were well received among players, such as

looking for evidence, investigating, and constructing arguments. Feedback did suggest that the overall theme and framing of the game may be polarising. One participant noted that the concept of arguing with family members (even in a comical sense) may not be appealing to players and may be off-putting to some. This participant noted that a game that centred on arguments may appeal to male players, but they mentioned that they nor their female friends would be interested in playing such a game. Upon reflection, this stage of game development occurred when my primary supervisor (a woman) was on maternity leave, thus the entire research team was comprised of young men. The main focus of the game had become centred around conflict and inconsequential arguments. Player feedback suggested that some participants found that this concept was not entirely compelling (particularly to female players), so we decided to steer the framing of the conflict away from the concept of familiar arguments. We also considered the idea that a conspiracy-obsessed uncle may play into the typical stereotype of a conspiracy theorist, which I highlight as a problematic issue in the Introduction. Finally, around the same time it had come to our attention that there had been another game developed by Cook et al (2023) titled *Cranky Uncle* with a somewhat similar premise. This was another motivating factor to change the theme of the game.

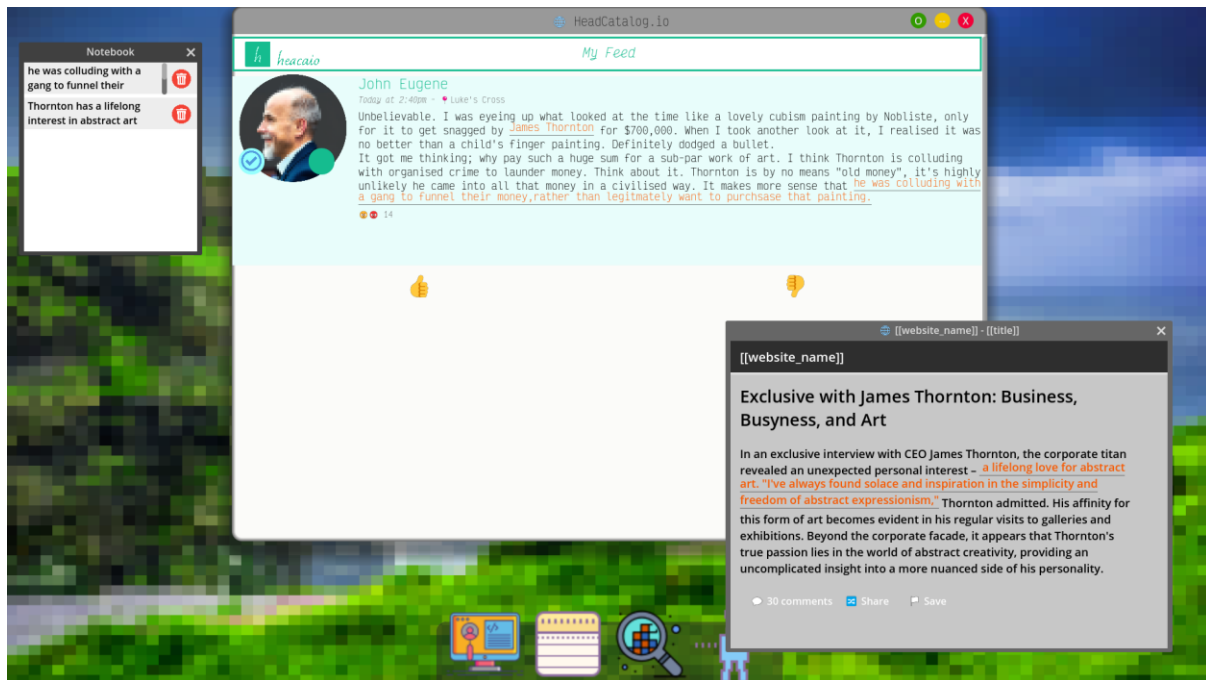


Figure 5.5: Godot prototype of Arguing with Your Uncle Online. Players can search for clues in their browser app (bottom right) and add evidence to their notebook (top left)

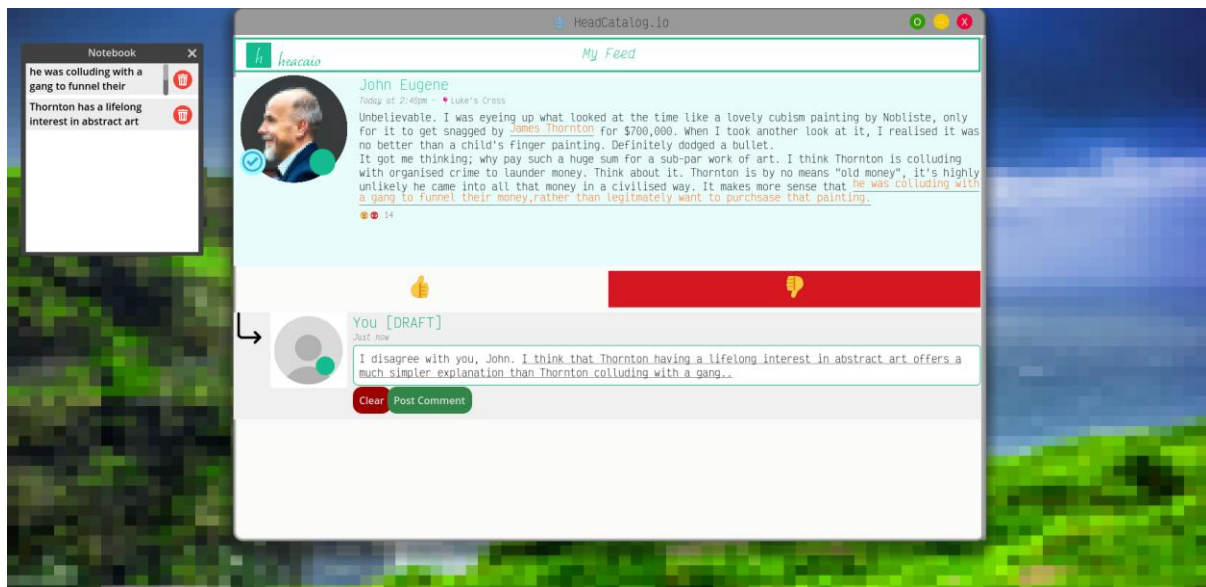


Figure 5.6: Godot prototype of Arguing with Your Uncle Online showing how players can construct responses. Players can click the like button to agree, or the dislike button to disagree. They then can take evidence from their notebook to populate the blank spaces in the response template

As playtesting indicated that the game's core mechanics were well received, we opted to solely change the framing of the game. Developing a new framing for the game, we drew on participant feedback and decided to move from a framework of *arguing* to a framework of *playing detective*. Instead of the player arguing with a family member, they would act as the PR assistant to a famous but conspiratorial musician and try to apply critical thinking skills to prevent them from posting any unfounded claims online. Changing the focus of the game to managing the public image of a musician rather than having them engage in arguments with a family member achieved two things. First, players now had a clear tangible goal to work towards; to preserve the public image of their client. The second is that this steered the focus of the game away from seemingly pointless arguments, which previous play testers found to be off-putting, and framed it in a more interesting and purposeful context. Pilot feedback found that players were more receptive to the idea of acting as a PR agent to a musician. To decide which sort of musician the player would be working for, we workshopped several ideas with volunteers such as rap, classic rock, folk, and country music. Volunteers found country music to be the most appealing, as there was a common sentiment that country music was associated with several classic tropes and an inherent sense of humour.

5.3.1.2 Nashville PI

5.2.1.2.1 Narrative Revisions

Due to the positive reception associated with country music, we conceptualised the game *Nashville PI*. In *Nashville PI*, players take on the role of a freelance social media manager who has been hired by the PR team of an aging musician named Bud Bateman. In the game, Bud Bateman has been enjoying a successful comeback and hopes that his new album will win him a Grammy. However, his PR team notes that he has a penchant for posting conspiracy-riddled social media posts and they worry that this may compromise his

comeback. The player is tasked with helping Bud to separate fact from fiction, by applying critical thinking skills to convince him of whether his claims are founded or unfounded.

The core mechanics of *Nashville PI* are almost identical to *Arguing with your Uncle Online*. In *Nashville PI*, Bud Bateman acts as the interlocutor, who comes to the player at the beginning of each level with a conspiracy theory he is convinced is true. The player uses phrases that Bud has mentioned as a cue to search for relevant evidence. The player then constructs an argument where they either agree or disagree with Bud's claims. If the player correctly solves that particular block, they proceed to the next block in the level. If the player does not correctly address Bud's argument, Bud will respond to the player, hinting that they have missed something, and asks them to go back and find the correct answer. This block will then loop until the player has correctly addressed Bud's argument. At the end of the level, Bud will typically declare that he is about to make an irreversible and drastic decision (e.g., publicly accusing his sponsors of criminal activity, accusing the FBI of covering up murders on his social media page). If the player answers this block incorrectly, the block does not loop, and Bud continues with his planned stunt. In the next level players receive an email from their boss Julia where she informs the player of whether their actions successfully helped Bud or placed the PR team in a predicament. She then provides the player with context for story of the next level and the game loop repeats.

While the game still requires the player to argue with the interlocutor, the framing of these interactions is less confrontational than in *Arguing with Your Uncle Online*. Bateman treats the player as an ally who can help him uncover real conspiracies and keep him in check when he is jumping to conclusions, often referring to the player endearingly as 'partner'. In contrast, the titular uncle in *Arguing with Your Uncle Online* viewed the player as a nuisance and a rival. We believed that the collaborative type of interaction between the player and the

interlocutor in Nashville PI would reduce the confrontational and argumentative tone of *Arguing with your Uncle Online* that early play testers found off-putting.

5.2.1.2.2 Mechanical Revisions

Based on player feedback from *Arguing with your Uncle Online*, we decided to simplify several of the existing mechanics in *Nashville PI*. For instance, some players reported identifying the correct information to successfully complete a block but used the wrong argument type to compose their response (i.e., saying that a piece of evidence offers a simpler explanation rather than directly contradicts a statement). We found that many of the argument types in the game overlapped considerably, which caused some confusion among players. Importantly, sometimes players demonstrated applying the appropriate critical thinking skills, but ultimately failed a block due to a technicality. To amend this confusion, we simplified the argument mechanic by removing the *conflict-of-interest*, *post-hoc fallacy*, and *validate sources* argument templates, focusing primarily on using contradiction/consistent argument templates instead. We drew inspiration from puzzle games such as *The Return of Obra Dinn*. *The Return of Obra Dinn* is a puzzle game where players take on the role of an insurance inspector investigating a ghost ship, who is tasked with identifying all deceased crew members on board and how they died (Pope, 2018). Developer Lucas Pope opted to have players make logical deductions within their heads rather than explicitly “showing their work” through the actual mechanics of the game (Pope, 2018). Pope achieved this by only requiring players to name the identity of a body onboard and noting their cause of death. Players have numerous ways to identify a crew member's identity, and players can make as many logical deductions in their own minds as the game does not require them to “show their work” through the game’s mechanics. Pope’s approach contrasts games such as *Sherlock*

Holmes: Crime & Punishments (Frogwares, 2014), where the game has players walk through the logical steps they are taking to come to their conclusions, combining clues to make new observations (See Figure 5.7). One problem with the mechanic in *Sherlock Holmes: Crime & Punishments* is that allowing the game to make deductions on behalf of the player takes control and autonomy away from the player. Whereas in *Obra Dinn* players have numerous ways to figure out how to find the identity of a victim and their cause of death.

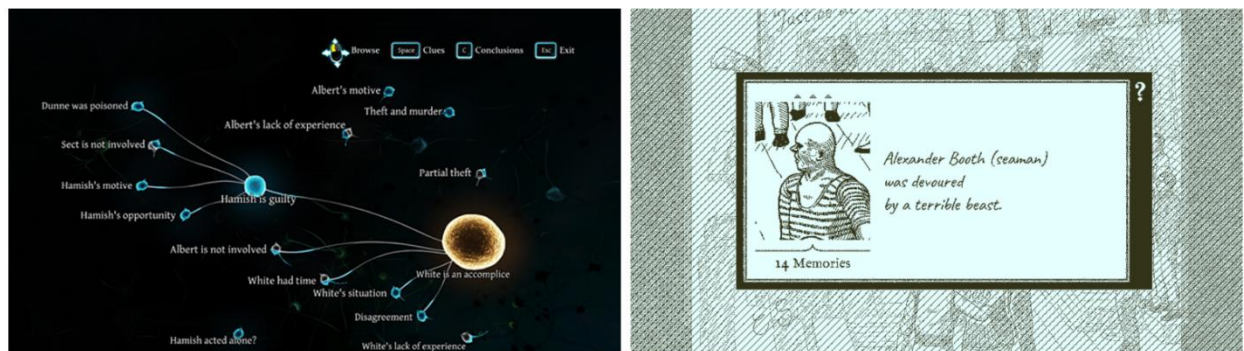


Figure 5.7: Illustration of deduction mechanic in 'Sherlock Holmes: Crime and Punishment'(Left). The players combine clues to come to a conclusion, showing their thinking process through the game. 'The Return of Obra Dinn' (Right) only asks players to name a victim and their cause of death. Players have more flexibility to come to their own conclusions as they are not required to show their thinking process through the game's interface

We decided to draw inspiration from Pope's approach to handling logical thinking in games. For example, in *Arguing with Your Uncle Online*, players encountered information that comes from a reliable source. Players were required to explicitly tell the interlocutor that the source is unreliable:

[This] article suggests that Daily Fib has a track record of fabricating stories.

This suggests that your source is unreliable.

The problem with this approach is that players argued that the *conflict-of-interest* argument also would have been a valid response here, so the overlap between the various argument templates ultimately confused players. In contrast, *Nashville PI* expects participants to identify unreliable sources too but does not expect players to explicitly state this. For instance, in one level of *Nashville PI*, Bud Bateman asks the player to find possible ways a murder could be covered up. The player encounters an internet user on a question forum suggesting that the plant oleander can be used as an undetectable poison. The oleander clue is merely a red herring. If players read another article within this level, they will find a testimony from a forensic scientist that says that the assumption that oleander is an untraceable poison is a myth. In this instance, the player still identifies whether a source is reliable or not, but they come to that conclusion using lateral reading, rather than explicitly interacting with the game and telling Bud that this source cannot be trusted. The player still shows their reasoning for why they disagree with Bud by providing counterevidence, but players are not expected to explicitly address every aspect of the information they encounter.

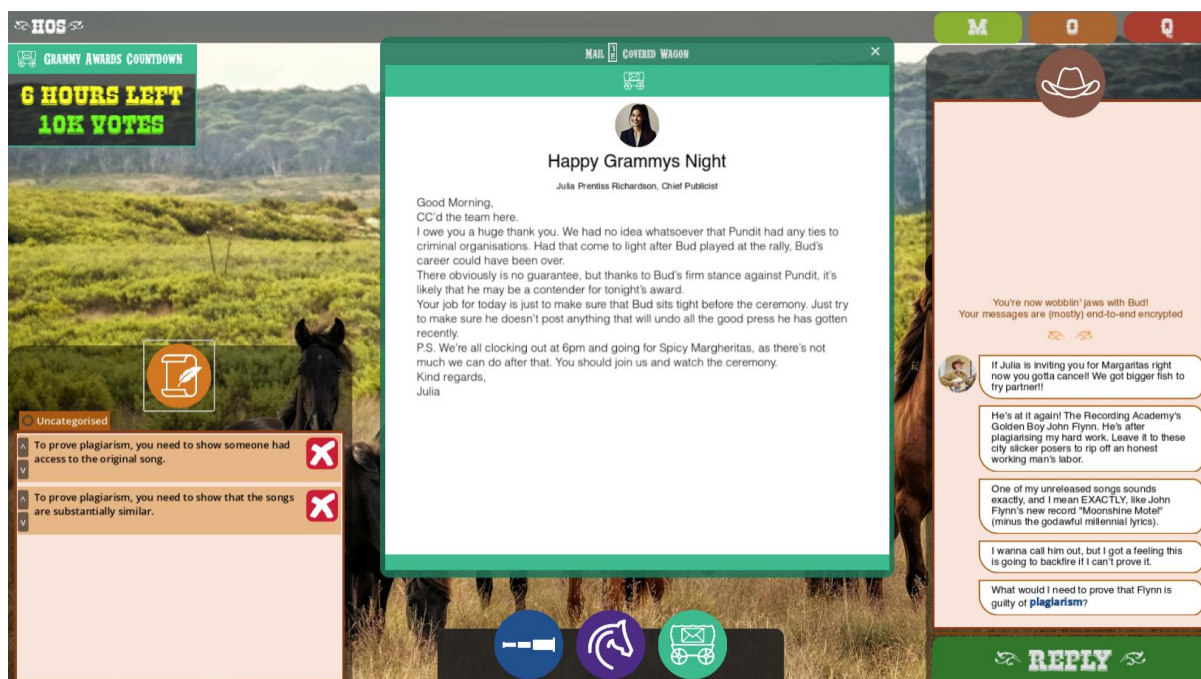


Figure 5.8: Screenshot of Nashville PI. Players receive an email (green application) at the beginning of each level which provides them with feedback on how they performed in the previous level. The emails also provide context about the current level. Players interact with Bud Bateman with the Howdy messaging app (right). The HUD (top left) tells players how many Grammy votes they have secured, and a countdown telling them how close they are to the Grammys (end of the game).

5.2.1.2.3 List of final learning outcomes

As we finalised the design of *Nashville PI* (see Figure 5.8), we settled on the final critical thinking skills we wanted to teach participants throughout this game. Many of these skills involved condensing and combining skills identified in Chapter 4 and embedding them within the gameplay loop of *Nashville PI*.

Skill 1: Corroborating Evidence

A core skill taught in this game is the ability to search for and identify corroborating evidence (O'Mahony et al., 2024; Stall & Petrocelli, 2023). Most of the skills taught in this game stem from the foundational ability to identify and use corroborating information. Players are

encouraged to search for evidence that either supports or contradicts Bud's statements, enabling them to make informed decisions about whether Bud's comments have merit. Players encounter various sources and pieces of information, and it is up to them to determine from context which clues are more trustworthy or adequately explain the conspiracy Bud is proposing.

Players manually search for information outside the messaging app they use to converse with Bud, simulating the real-world behaviour of researching evidence. This not only fosters the player's skill in searching for information but also incentivises them to read all available material before drawing conclusions. Players may initially find clues that seemingly support or contradict Bud's claims. However, dedicating more time to a thorough search may uncover more comprehensive or accurate information that better contradicts or supports Bud's arguments. This gameplay mechanic discourages players from engaging in "jumping-to-conclusions" biases, which are common in conspiracy thinking (Pytlik et al., 2020; Roozenbeek et al., 2022). Furthermore, incentivising players to evaluate all available information helps combat confirmation bias (Wagner-Egger, 2022).

Skill 2: Lateral Reading

Building on the Corroborating Evidence skill, lateral reading is a key component of this game. To determine the reliability of information, players must cross-reference what they read with other sources. Lateral reading has been identified as an important skill for fostering critical thinking and evaluating misinformation (Breakstone et al., 2021).

For example, in Level 2, players encounter an article suggesting that a chemical listed as an ingredient in Buckshot Bourbon is dangerous to humans. A player might initially take this evidence at face value. However, reading further reveals that the chemical is only toxic in

high concentrations, and the concentration in bourbon is dramatically lower and safe to consume. Similarly, in one level, the player investigates possible cause of the death of fictional country singer Kelly Kentucky. In an online discussion, they come across the claim that the poisonous plant Oleander can kill people without leaving a trace. However, a seemingly unrelated article reveals this to be a myth, as Oleander is detectable in an autopsy. By discouraging the “jumping to conclusions” bias (Pytlik et al., 2020), lateral reading encourages players to consider evidence in a broader context and avoid taking individual pieces of information at face value (O’Mahony et al., 2024a, b).

Skill 3: Identifying Simpler Explanations

In many levels, players uncover evidence that explicitly contradicts Bud’s claims. For instance, Bud may believe a chemical is dangerous, but the player finds evidence that it is safe. However, conclusive evidence is not always available. In such cases, players learn to identify simpler or more probable explanations for events in the absence of indisputable evidence.

While this distinction may seem subtle, scenarios where neither players nor Bud know all the facts and must weigh up the probabilities of multiple plausible explanations, mimic the ambiguous and uncertain nature of real-world conspiracy theories (Wood, 2017). Players are encouraged to spot explanations requiring fewer unwarranted assumptions and to question those that are overly complicated or rely on numerous unsupported claims (Banas & Miller, 2013; Stall & Petrocelli, 2023).

Skill 4: Behind the Curtain

Many misinformation games adopt the approach of showing players what happens behind the

curtain. In the *Bad News* game, players explore how misinformation is created and the persuasive techniques used to mislead the public. Similarly, this game includes a mechanic where players imagine what they would hypothetically need to do to conceal a conspiracy themselves.

For example, in Level 3, Bud asks the player to humour him by outlining the steps a secret organisation would need to take to cover up the alleged murder of his friend, Kelly Kentucky. Players identify the various people and groups that would need silencing, as well as the logistical challenges that could expose the conspiracy (e.g., bribing the coroner, tampering with backup samples, erasing access logs for the forensics lab). This mechanic aims to teach players the principle that the complexity of a conspiracy - and the number of people and organisations involved - makes it increasingly untenable (Grimes, 2016; Stall & Petrocelli, 2023). Players are thus forced to confront the implausibility of ensuring all loose ends are tied up.

Skill 5: Resist Blind Scepticism

As mentioned in Chapter 4, many interventions teach participants to only dismiss conspiracy theories. Our empirical findings suggested that it is important to discourage blind scepticism when teaching participants how to critically evaluate conspiracy theories (O'Mahony et al, 2024). Based on these findings, we decided to include levels where the conspiracy theories that Bud Bateman proposes are actually true. We strategically placed these at the middle and end of the game, as players would be familiar with the basic mechanics at that stage of the game and would be ready to try to discern conspiracies that are plausible from those that are implausible. Including both plausible and implausible conspiracy theories in this game teaches players that they need to evaluate evidence and arguments as opposed to dismissing anything that sounds somewhat conspiratorial.

5.4 Game Evaluation

With the final design of the game set, we sought to evaluate how effective it was as a psychological intervention to increase critical discernment of conspiracy theories. We then conducted an experimental study, where we compared the critical thinking skills of those who played a control game, and those who played *Nashville PI*.

5.4.1 Methods

The current study was preregistered (available here: <https://aspredicted.org/xx7q-7phh.pd>).

All data and materials for this study are available

here: https://osf.io/v56xg/?view_only=b0618b11b6114492ac4d856d8c460166. *Nashville PI*

was hosted on the game-sharing platform Itch, and can be accessed through this link:

<https://diobhal.itch.io/nashville-pi>

5.4.1.1 Participants

Power analysis using GPower version 3.1.9.7 indicated that 506 participants were required to detect a Cohen's d effect size of 0.3 with 80% power, with an independent t-test. We recruited 521 participants from the participant recruitment website Prolific (231 women, mean age = 36.36, SD = 12.38). Most participants were from the United Kingdom (47%), and the remaining participants were from the United States (43%), Canada (8%), Ireland (1%), Australia (1%), and New Zealand (<1%). Due to many participants not sending their completion codes, Prolific oversampled and we had a total of 521 participants in our final sample.

5.4.1.2 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). 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 8.37 ($SD = 3.70$) out of a possible 15. The scale showed good reliability in the current study (Cronbach's $\alpha = .79$). The implausible subscale showed good reliability ($\alpha = .74$), 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"). 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 15 to 75 (min = 15.00, max = 75.00). The mean score was 42.94 ($SD = 13.38$). The internal reliability was excellent ($\alpha = .94$).

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 the same adapted version of the IEUB that consisted of 20 statements that was used in Chapter 4. An overall mean score for participants was calculated for the purpose of

analysis with a range of 20 to 100 (min = 20.00, max = 89.00). The overall mean score was 52.73 ($SD = 14.37$). The scale showed strong reliability, Cronbach's $\alpha = .89$.

5.4.1.3 Procedure

The study was advertised on Prolific with the following banner “*Help a Country Star Go for Gold: Play a Game to Help Study Critical Thinking!*”. After reading the information sheet and providing informed consent, participants were randomly assigned to one of two conditions: Control (where participants played Tetris for two minutes), and Nashville (where participants played Nashville PI, our critical thinking game). After completing the game they were assigned to, participants then completed the CTAC, followed by the IEUB, and finally the GCBS. Participants were then asked to answer some final questions on what they thought of the game, including how much they liked the survey, how likely they were to recommend the survey, whether they thought they had gotten better at spotting plausible/implausible conspiracy theories, and finally, whether they had any comments about the game they would like to share. Finally, participants were debriefed and were redirected to Prolific.

Participants received a handful of attention checks during the survey. Participants in the Control and Nashville conditions both received a nonsensical attention check where participants were asked to agree with the following statement: “*I swim across the Atlantic Ocean to get to work every day*”. Every participant completed an attention check during the CTAC section, which was presented in a random order amongst the CTAC items. We wanted to ensure that participants were properly reading the vignettes, as due to the length and cognitive demand of the items we predicted that some participants may become fatigued and skim these items. Finally, participants in the Nashville group were asked a simple question about the game's story to see if they were paying attention during the game.

5.4.1.4 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.

5.4.2 Results

5.4.2.1 Did the intervention improve critical appraisal of conspiracy theories?

We ran an independent t-test to examine the effectiveness of Nashville PI on participants' ability to critically appraise conspiracy theories overall measured on the CTAC total. The analysis revealed there was a significant difference in overall CTAC scores between those who played Nashville PI ($M = 7.93$, $SD = 3.79$) and those in the Control group ($M = 8.86$, $SD = 3.54$), $t(517) = 2.87$, $p = .004$, $d = .25$), but contrary to our expectations, our game resulted in *worse* critical thinking about conspiracies than the simple control. We then compared the intervention groups on their ability to critically appraise implausible and plausible conspiracy theories. For CTAC Implausible scores, an independent t-test revealed no significant difference between those who played Nashville PI ($M = 3.95$, $SD = 2.37$) and those in the Control group ($M = 4.34$, $SD = 2.37$), $t(517) = 1.84$, $p = .067$, though note this trended towards better scores in the Control group. Finally, an independent t-test found that there was a significant difference in CTAC Plausible scores between those who played Nashville PI ($M = 3.98$, $SD = 1.92$) and those in the Control group ($M = 4.53$, $SD = 1.78$), $t(517) = 1.84$, $p < .001$, $d = .29$), again with better scores in the Control group.

5.4.2.2 Did the intervention decrease general conspiracy ideation?

To test the effect that our critical thinking game had on participants' general level of conspiracy ideation, we ran an independent t-test comparing GCBS scores between those who played Nashville PI and those who played the Control game. The t-test revealed that there was no significant difference in GCBS scores between those who played Nashville PI ($M = 42.67$, $SD = 13.34$) and those in the Control group ($M = 43.25$, $SD = 13.45$), $t(517) = 0.48$, $p = .629$. These results suggest that those who played Nashville PI did not differ in their overall tendency to endorse conspiracy theories from those in the Control group.

5.4.2.3 Did the intervention decrease epistemically unwarranted beliefs?

To test the effect that our critical thinking game had on participants' tendency to endorse epistemically unwarranted beliefs, we ran an independent t-test comparing IEUB scores between those who played Nashville PI and those who played the Control game. The t-test revealed that there was a significant difference in IEUB scores, with those who played Nashville PI ($M = 53.94$, $SD = 14.78$) endorsing *more* epistemically unwarranted beliefs and those in the Control group ($M = 51.37$, $SD = 13.80$), $t(517) = 0.28$, $p = .042$, $d = -.18$. These results suggest that those who played Nashville PI were more likely to endorse epistemically unwarranted beliefs than those in the Control group.

5.4.3 Narrative Analysis

The findings of our empirical analysis suggested that our critical thinking game Nashville PI *decreased* participants' ability to critically appraise conspiracy theories. To understand what might be occurring, we examined the feedback we received from participants who had played Nashville PI. We asked participants for their feedback on the survey with three questions (“What did you think about the game in general? Is there anything you think that could be improved?”, “Was there a part of the game you particularly liked?”, “Did you get stuck at any

particular point in the game?”). Below we discuss some patterns and themes that emerged among the feedback we received from these participants.

5.4.3.1 Did participants enjoy the game?

The responses we received from participants who had played Nashville PI indicated that they generally enjoyed the game (*“I was carried away by the game because it was interesting”, “It was interesting and engaging. I can't think of anything that I would improve”, “flows well and mind blowing”, “The entire game was very interesting”*). The theme and concept were well received among most players, saying that they *“liked the theme”* and thought *“the tone of the game was very fun”*. Participants were particularly drawn to the detective and investigation aspect. When asked *“Was there a part of the game you particularly liked?”*, participants frequently cited the detective elements of the game (*“Quite liked the detective type of game”, “I found it enjoyable trying to piece the information together and it highlighted how easily you can be swayed based on what information you stumble across first”, “solving mysteries”*). One particular player said that they like the *“[e]vidence presentation like Pheonix Wright game”*, referencing the popular detective video game series *Ace Attorney*.

Particular praise was given to the game’s theme, writing, and dialogue, in addition to Bud Bateman as a character (*“I enjoyed the story and writing; it was a well-written game”, “I liked Bud's outlandish personality and random commentary. It reminded me of my grandfather”, “I liked Bud's personality and dialogue”, “it was fun watching bud freak out”, “The humour in the texts”*). Many participants cited the narrative when asked if there was a part of the game they particularly liked (*“the story overall”*). This feedback suggested that the overall story, dialogue, and setting of the game was well received among participants

5.4.3.2 Did players have any issues with the game?

Despite the positive feedback concerning the game's theme, narrative, and dialogue, participants had several critiques concerning the mechanics of the game. The most common complaint was the length of the game. The length of the game had been determined to give players sufficient time to practice the skills taught in the game. However, many players felt that the game was too long and that this hindered their enjoyment (*"It was ok a little long."*, *"absolutely too long!"*, *"too long"*), though these participants were being paid for their time on Prolific and thus mentioned in their comment that they were concerned about the length relative to the amount they were getting paid, so it's not clear if they felt it was too long in general or just as a task on Prolific. The length of the game was exacerbated by another concern of players, which was that at various points of the game, it was unclear what the correct answer was or how the player should proceed (*"I think at many points in the game it was hard to progress and sometimes felt like I didn't know what to do"*, *"I got stuck in several points in the game, as there was far too much "evidence" provided and sometimes the Bud Bateman character would repeat the same query (which obviously indicated that it was unhappy with my choices for response)"*). One player summarised the overlap between problems with the length of the game with a lack of clarity on how to proceed in this comment: *"I think it was nice to have such an interactive game but I did find it to be difficult at times and got boring when I kept getting the same responses and couldn't easily crack what my response should be. This made the game take longer than I thought which was a bit frustrating"*. Getting "stuck in a loop" was quite common among participants, where they would provide the incorrect answer, and Bud's feedback would not give them a clear idea of how to correct their mistake: *"The game was great! I have no idea how I would've won though. It genuinely did seem that John Flynn had plagiarised the song, and that the evidence was pointing to this"*.

Participants also found that some elements of the gameplay were either cumbersome or unnecessary. Players felt that it was unnecessary to have to manually add Bud's observations into the response template and that it made more difficult than necessary to construct replies to Bud: *"The game was amusing, and I was legitimately trying to win. Pointless to have to paste in his viewpoint each time. Only the evidence responses were necessary"*. As mentioned previously, this is something we had already considered and was something identified in the final round of playtesting. Unfortunately, amending this element before evaluating the game was not feasible.

The feedback we received from participants indicated that the overall concept, setting, and narrative were well-received among the sample but the length of the game, a lack of clear feedback when players answered incorrectly, and the cumbersome nature of constructing answers were barriers to the players overall enjoyment and possibly hindered the player's ability to learn the critical thinking skills we sought to teach.

5.5 General Discussion

Serious games have previously shown success in challenging unwarranted beliefs such as fake news and misinformation (Basol et al., 2021; Cook et al., 2023; Roozenbeek & Van Der Linden, 2019, 2020). However, none have been designed to directly challenge the thinking underlying conspiracy beliefs and emphasise detecting false information as opposed to teaching players to discern between plausible and implausible information (Kiili et al., 2024). We developed a novel game *Nashville PI* that sought to teach players how to apply critical thinking skills to critically discern between plausible and implausible conspiracy theories. In complete opposition to our expectations, our results indicated that participants who played *Nashville PI* were *worse* at critically evaluating conspiracy theories than those who received Tetris in the control condition. These results are in contrast to previous research

that has suggested that video games can reduce players' tendency to endorse unfounded conspiracy theories (Basol et al., 2021; Roozenbeek & Van Der Linden, 2019, 2020). While the concept, narrative, and dialogue of *Nashville PI* were well-received among players, the length of the game and several limitations with regard to the mechanics and player experience hindered overall enjoyment and possibly reduced the efficacy of the game as an intervention.

One explanation for why *Nashville PI* was ineffective in terms of increasing critical appraisal of conspiracy theories was due to the length of the game. Participants' playtime ranged approximately between 35-40 minutes in addition to completing the Qualtrics questionnaire. The length of the game contrasts the comparatively short playtime of previous inoculation games such as *Bad News* and *Harmony Square* which take roughly 10 minutes to complete (Roozenbeek & Van Der Linden, 2019, 2020). After 40 minutes of playtime, participants then had to complete the also cognitively demanding CTAC, which they may have found difficult. This explanation might go some way to explaining why control participants outperformed the intervention group on the CTAC but does not explain why the intervention group also reported increased epistemically unwarranted beliefs.

In addition to the length of the game being a barrier to the game's effectiveness as an intervention, the lack of clear specific feedback given to players when they answered incorrectly is a possible reason why the game was ineffective in increasing participants' critical appraisal of conspiracy theories (Linehan et al., 2011). Participants commonly cited a lack of clear feedback and an uncertainty of which actions to take next as a limiting factor in their ability to play the game, which is a crucial element necessary for successful educational games (Linehan et al., 2011). While a certain level of feedback was given to players when they answered a question incorrectly, it was not always consistent. For instance, in the fifth level of the game, players are asked to see if there is a link between presidential candidate Joe Pundit and Saul Mulligan, the leader of an infamous crime syndicate. If they player does not

find this link, Bud says the following: “*Now that I think about it, I’m sure I remember reading something about Joe bailing Saul out of jail or something. You think you could find something proving that?*”. In this instance, Bud hints at the relevant information that the player has missed (i.e. the fact that Pundit once bailed Mulligan out of prison). However, in other levels the feedback Bud provides is less clear in what the player needs to do next: “*I dunno kid, I’m not really catchin’ what yer throwin’ ...*”, “*Is that all you’d need to do to cover up a crime like that?*”. As such, many players got caught in a loop and were not sure where they were making a mistake. Importantly, when the player answers correctly they do not receive explicit feedback on what they did well – this differs from the approach we took in our successful experimental intervention reported in Chapter 4, where participants were always shown what the correct answer was and crucially, *why*, whether they got a question right or wrong (O’Mahony et al., 2024).

By looking at previous examples of games that challenge unfounded beliefs, we can see how a lack of clear feedback may hinder a player's ability to learn the skills the game is trying to teach. For instance, in the *Bad News* game, players are asked at one point to post an article to get viewer engagement, if the player picks a boring topic that does not utilise the manipulation tactics they have been taught, they receive the following message: “*Yikes, that didn't work. Your article didn't use any manipulation techniques, so it didn't go very far. Let's try again*”. The *Bad News* game explicitly tells the player that their article did not use any manipulation techniques, and that the next action is clear to the player (i.e., post an article with more manipulative messaging). Similarly, previous games have highlighted the importance of also providing specific feedback when the player gives the correct answer (Linehan et al., 2011).

For instance, in *Harmony Square* players are asked to try to generate a controversy online (Roozenbeek & Van Der Linden, 2020). When players successfully attempt to polarise

a trivial issue in the game they receive this feedback: “*Smart. By pretending to take sides, we can turn this complete non-issue into a “debate” on Harmony Square’s social media page*”.

The feedback in these games not only corrects players when they are making mistakes but also explicitly tells them when they are succeeding in the game. This helps the player reflect on their actions and allows them to effectively learn the skills the game is trying to teach them. Returning to an example in *Nashville PI*, a player commented that they were sure they had the correct answer in level 6 when they accused Bud’s rival John Flynn of plagiarising his music and were surprised when they were told that they were wrong: “*It genuinely did seem that John Flynn had plagiarised the song and that the evidence was pointing to this*”. In this instance, players were attempting to figure out who leaked Bud’s music to John, but if they accused the wrong member of Bud’s team, they failed the level. The player was correct in their assertion that Flynn was guilty of plagiarism but had ultimately accused the wrong member of Bud’s team. However, due to a lack of clear feedback, the player did not know where they had gone wrong and assumed that the reason they failed the level was because Flynn was not guilty of plagiarism. Due to the player’s difficulty understanding what skills the game was trying to teach them, along with the uncertainty of why their answers were correct or incorrect, we propose that the game may not have been ineffective in teaching critical discernment of conspiracy theories due to some flaws in gameplay mechanics, and not as of a result of the teaching outcomes themselves. We propose that future work conducting playtesting analysis on the game will help further highlight and address the shortcomings of the game we have listed earlier and may possibly improve the game’s ability to effectively teach players how to apply critical thinking skills to conspiracy theories by giving them more explicit feedback on their gameplay choices and decreasing the length of the game. It should be noted that as was the case in Chapters 3 and 4, a methodological constraint of this study was that it mainly examined participants from English-speaking countries. Furthermore,

while our findings did show that the game initially decreased participants' critical appraisal of conspiracy theories, the overall effects were quite small.

Our findings have several theoretical implications. The most significant implication mirrors our findings in Chapter 4 where we found that conspiracy interventions can potentially cause harm. We found similar implications with *Nashville PI*, which was not only ineffective as a conspiracy intervention but decreased participants' ability to critically appraise both plausible and implausible conspiracy theories. These findings serve as a reminder of the possible risks of developing well-intentioned conspiracy interventions, in that they can both improve and diminish critical appraisal of conspiracy theories. Another implication for our findings is that they have demonstrated the necessity for measuring the underlying cognitive processes of conspiracy beliefs, along with using assessments that include both plausible and implausible conspiracy items. By using the CTAC, we were able to detect that *Nashville PI* had a negative effect on participants' ability to appraise both plausible and implausible conspiracy theories. These results suggested that participants were not merely suspicious or accepting of conspiracy theories in general but were overall worse at applying critical thinking skills to evaluate conspiracy theories regardless of their plausibility. Had we used an outcome measure that only measured belief in unfounded conspiracy theories, we would not have been able to make this observation.

In summary, while our study found that a novel critical thinking game was ineffective at teaching participants how to critically appraise conspiracy theories, we posit that a handful of flaws in the user experience and feedback mechanics of the game may be holding back the game's potential to be used as a successful intervention rather than the skills the game was teaching being inadequate. We believe that future work employing design research to identify areas of difficulty and uncertainty among players will help address these issues and will potentially improve this game's utility as an intervention to help players improve their ability

to critically discern between plausible and implausible conspiracy theories.

5.6 References

- 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>
- Breakstone, J., Smith, M., Connors, P., Ortega, T., Kerr, D., & Wineburg, S. (2021). Lateral reading: College students learn to critically evaluate internet sources in an online course. *Harvard Kennedy School Misinformation Review*.
<https://doi.org/10.37016/mr-2020-56>
- 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>
- Capcom. (2001). *Phoenix Wright: Ace Attorney*. Capcom Co., Ltd.
- Cook, J., Ecker, U. K. H., Trecek-King, M., Schade, G., Jeffers-Tracy, K., Fessmann, J., Kim, S. C., Kinhead, D., Orr, M., Vraga, E., Roberts, K., & McDowell, J. (2023). The cranky uncle game—Combining humor and gamification to build student resilience against climate misinformation. *Environmental Education Research*, 29(4), 607–623.
<https://doi.org/10.1080/13504622.2022.2085671>

- 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>
- Davidson & Associates. (1983). *Math Blaster*. Davidson & Associates, Inc.
- Deck Nine. (2021). *Life is Strange: True Colors*. Square Enix.
- 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>
- 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/s11162-018-9513-3>
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. <https://doi.org/10.1038/s44159-021-00006-y>
- Ericsson, K. A. (2016). *Peak: Secrets from the new science of expertise*. Houghton Mifflin Harcourt.
- Fernández-Vara, C. (2013). The Game's Afoot: Designing Sherlock Holmes. *DiGRA Conference*. <https://api.semanticscholar.org/CorpusID:12919180>
- Frogwares. (2014). *Sherlock Holmes: Crimes & Punishments*. Focus Home Interactive.
- Grimes, D. R. (2016). On the Viability of Conspiratorial Beliefs. *PloS One*, 11(1), e0147905. <https://doi.org/10.1371/journal.pone.0147905>
- Hu, B., Ju, X.-D., Liu, H.-H., Wu, H.-Q., Bi, C., & Lu, C. (2023). Game-based inoculation versus graphic-based inoculation to combat misinformation: A randomized controlled

- trial. *Cognitive Research: Principles and Implications*, 8(1), 49.
<https://doi.org/10.1186/s41235-023-00505-x>
- 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>
- Linehan, C., Kirman, B., Lawson, S., & Chan, G. (2011). Practical, appropriate, empirically-validated guidelines for designing educational games. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1979–1988.
<https://doi.org/10.1145/1978942.1979229>
- Linehan, C., Lawson, S., & Doughty, M. (2009). Tabletop Prototyping of Serious Games for “Soft Skills” Training. *2009 Conference in Games and Virtual Worlds for Serious Applications*, 182–185. <https://doi.org/10.1109/VS-GAMES.2009.9>
- 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)
- Ollila, E. M. I., Suomela, R., & Holopainen, J. (2008). Using prototypes in early pervasive game development. *Computers in Entertainment*, 6(2), 1–17.
<https://doi.org/10.1145/1371216.1371220>
- O’Mahony, C., Murphy, G., & Linehan, C. (2024). True discernment or blind scepticism? Comparing the effectiveness of four conspiracy belief interventions. *Advances in Psychology*, 2(1). <https://doi.org/10.56296/aip00030>
- Pope, L. (2018). *Return of the Obra Dinn*. 3909 LLC.
- Pusic, M., Pecaric, M., & Boutis, K. (2011). How Much Practice Is Enough? Using Learning Curves to Assess the Deliberate Practice of Radiograph Interpretation: *Academic Medicine*, 86(6), 731–736. <https://doi.org/10.1097/ACM.0b013e3182178c3c>

- 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/fpsy.2020.568942>
- Rodriguez-Calzada, L., Paredes-Velasco, M., & Urquiza-Fuentes, J. (2024). The educational impact of a comprehensive serious game within the university setting: Improving learning and fostering motivation. *Heliyon, 10*(16), e35608.
<https://doi.org/10.1016/j.heliyon.2024.e35608>
- 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. (2020). Breaking Harmony Square: A game that “inoculates” against political misinformation. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-47>
- 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>
- Snyder, C. (2003). *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces* (1st ed). Elsevier Science & Technology.
- 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>
- Stasielowicz, L. (2022). Who believes in conspiracy theories? A meta-analysis on personality correlates. *Journal of Research in Personality, 98*, 104229.
<https://doi.org/10.1016/j.jrp.2022.104229>

- Van Prooijen, J.-W., Spadaro, G., & Wang, H. (2022). Suspicion of institutions: How distrust and conspiracy theories deteriorate social relationships. *Current Opinion in Psychology*, 43, 65–69. <https://doi.org/10.1016/j.copsyc.2021.06.013>
- Vranic, A., Hromatko, I., & Tonković, M. (2022). “I Did My Own Research”: Overconfidence, (Dis)trust in Science, and Endorsement of Conspiracy Theories. *Frontiers in Psychology*, 13, 931865. <https://doi.org/10.3389/fpsyg.2022.931865>
- Wagner-Egger, P. (2022). *The Noises of Conspiracy: Psychology of Beliefs in Conspiracy Theories*. <https://doi.org/10.31234/osf.io/fv52e>
- 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>

Chapter 6: Discussion

This thesis aims to examine whether critical thinking can be an effective means of measuring and challenging unfounded conspiracy beliefs. Across a systematic review, a psychometric scale development paper, and two experimental papers, the results were supportive of critical thinking as an effective means of measuring how people think about unfounded conspiracy theories, and as a means of teaching people to discern between unfounded conspiracy theories and plausible conspiracy theories.

6.1 Thesis Overview

In the Introduction of this thesis, I introduced several research questions that I addressed throughout this work. I first posed the questions “*How effective are existing interventions at reducing conspiracy beliefs?*” and “*How effective are existing interventions at reducing conspiracy beliefs?*”. In Chapter 2, I explored the existing literature surrounding interventions that have been designed to challenge conspiracy beliefs. The review had three key findings. The first was that the majority of existing interventions were ineffective in terms of reducing belief in conspiracy theories. Second, three main categories of interventions showed some promise: analytical priming, informational inoculations, and critical thinking interventions. Third, the review highlighted that there was little consistency in the way that conspiracy beliefs were measured across interventions, making comparisons difficult. Measures were either topic-specific (anti-vaccine beliefs) or were too broad to be meaningful (general suspicions that the government is up to no good). This leads to my next research questions: “*How can critical appraisal of conspiracy theories be conceptualised and measured reliably?*” and “*How well does this novel assessment (CTAC) correlate with general conspiracy ideation and critical thinking?*”. Based on the limitations of current measures highlighted in Chapter 2, in Chapter 3 I sought to develop a more sensitive measure of

conspiracy thinking called the Critical Thinking about Conspiracies assessment (CTAC). The CTAC offered a more nuanced measure of conspiracy thinking by focusing on the underlying rationale of conspiracy beliefs and had the novel contribution of using both plausible and implausible items in contrast to existing assessments that encourage participants to reject all conspiracy theories. My findings also indicated that the CTAC correlated well with related measures of conspiracy ideation and cognitive reflection. My next questions were “*How do potential conspiracy interventions compare in terms of increasing critical appraisal of conspiracy theories?*” My review in Chapter 2 highlighted that it was difficult to compare the efficacy of promising interventions as they used different outcome measures to assess conspiracy beliefs. Additionally, these interventions had never been directly compared to one another in an experimental study. In Chapter 4 I addressed this gap by comparing three promising interventions using an experimental design. The findings suggested that existing interventions increased participants' abilities to reason about implausible conspiracy theories but not plausible ones. This led to my next research question: “*Does encouraging reasoning over blind scepticism improve discernment between plausible and implausible conspiracy theories?*”. I found that a novel discernment-based intervention was effective in terms of increasing participants' ability to critically distinguish between plausible and implausible conspiracy theories (albeit with overall small main effects). Finally, I tackled my final research question: “*Can a serious game effectively encourage critical evaluation of conspiracy theories?*”. To address this question, in Chapter 5 I designed a critical thinking game called *Nashville PI*, to condense the learning outcomes of the discernment interventions of Chapter 4 into a scalable and enjoyable game. The results suggested that the concept, mechanics, and narrative were well received among players, but our empirical evaluation suggests that *Nashville PI* was not successful as an effective intervention to increase people's ability to discern between implausible and plausible conspiracy theories by applying critical

thinking skills. Again, the overall effects of the game were found to be small. Feedback from players suggested that several mechanical issues were likely responsible for limiting the game's efficacy as an intervention, which I shall address in my future research activities. In summary, the studies that form this thesis have demonstrated the utility and applicability of critical thinking skills as a means of both measuring and challenging unfounded conspiracy theories, addressing all the aims outlined in the Introduction.

6.2 Contributions

As noted in the introduction of this thesis, examining conspiracy thinking through the lens of critical thinking offers several contributions to the field of conspiracy research, with several theoretical and applied implications.

6.2.1 Novel Review Comparing Efficacy of Interventions

This thesis contributes a novel systematic review of the psychological literature investigating conspiracy belief interventions. Surprisingly, such a crucial review had not been conducted before this project, despite the emergence of various forms of conspiracy interventions designed by psychology researchers. In Chapter 1 I proposed two research questions to be answered with this review. The first question was “*How effective are existing interventions at reducing conspiracy beliefs?*”. My review has highlighted that the majority of existing interventions are ineffective, with few showing statistically significant changes, and only a handful of interventions were capable of inducing medium to large changes in conspiracy beliefs. My review has highlighted several variations of conspiracy intervention that show promise and may direct future conspiracy interventions in the right direction. Furthermore, my findings highlighted some limitations of current conspiracy interventions. The second question posed in Chapter 1 was “*How effective are current measures of conspiracy thinking?*”. My review was that there was a large amount of variation in the way that

conspiracy beliefs were measured, which I cite as a factor that limits direct comparison of their efficacy as interventions. This review also highlighted the lack of longitudinal studies and cross-cultural comparisons among current interventions. Overall, this review has greatly contributed to the theoretical understanding of the current state of conspiracy belief interventions.

6.2.2 Operationalising Conspiracy Thinking

As highlighted in Chapter 2 there is a lack of consistency in the way that researchers have been measuring conspiracy thinking. Existing measures are either topic-specific (anti-vaccine conspiracy theories) or prompt participants with broad statements such as “A lot of important information is deliberately concealed from the public out of self-interest”. I argue that taking either of these approaches limits their scalability and that more focus needs to be directed to measuring the underlying cognitive processes of conspiracy beliefs. Furthermore, the majority of assessments that have been used in experimental studies on conspiracy beliefs have only measured endorsement of implausible conspiracy theories, with no aim of measuring whether participants can discern between plausible and implausible conspiracy theories. To that end, in Chapter 1 I posed the question “*How can critical appraisal of conspiracy theories be conceptualised and measured reliably?*”. To answer this question, in Chapter 3 I developed a novel assessment called the Critical Thinking about Conspiracies assessment (CTAC) to address these concerns. The CTAC has the novel benefit of measuring a participant's ability to critically evaluate a conspiracy theory rather than whether they merely endorse the theory itself. Additionally, the CTAC contains both plausible and implausible items, which allows researchers to measure whether participants can accurately distinguish plausible conspiracy theories from implausible conspiracy theories. As I demonstrated in Chapters 4 and 5, the CTAC's ability to measure discernment has a unique advantage in that it moves beyond measuring a potential response bias in terms of endorsing

conspiracy theories and detects whether participants are making tangible logical errors in evaluating conspiracy theories. Another advantage of the CTAC is that it is not constrained by time and culture as it uses novel conspiracy theories to populate items, in contrast to existing measures of conspiracy ideation. The second research question I proposed for Study 2 was “*How well does this novel assessment (CTAC) correlate with general conspiracy ideation and critical thinking?*”. My findings indicated that the CTAC showed a moderate negative correlation with measures of conspiracy ideation, and a positive correlation with cognitive reflection, as we predicted. Overall, the CTAC has contributed a novel, scalable and widely functional assessments to the field of conspiracy research.

6.2.3 Highlighting Oversights in Current Conspiracy Intervention Research

The work in this thesis has highlighted and addressed several limitations of existing conspiracy intervention research. A major contribution of this thesis has been highlighting the oversights of existing conspiracy intervention research. In Chapter 1, I note that a prevailing approach to studying conspiracy ideation is to treat all conspiracy theories as unfounded. As such, much research has prioritised reducing belief in conspiracy theories without much consideration regarding discernment between plausible and implausible information. The work in this thesis is the first series of experimental studies to assess participants' ability to assess plausible and implausible theories. Using an assessment that is capable of measuring discernment has allowed my experimental work to detect that many conventional conspiracy interventions do in fact decrease endorsement of implausible conspiracy theories, but also make participants more sceptical of plausible conspiracy theories. Because previous intervention research has predominantly measured endorsement of implausible conspiracy theories, it could not be determined whether these interventions were merely encouraging a response bias to reject all conspiracy theories blindly. In particular, my experimental work in Chapter 4 has highlighted that conspiracy interventions run the risk of

increasing blind scepticism, and my work in Chapter 5 suggests that conspiracy interventions can risk more harm than good if not designed with care. These findings highlight a crucial takeaway message; that future research ought to emphasise teaching and measuring discernment as opposed to merely measuring belief in unfounded conspiracy theories. Suppose future studies neglect to measure discernment as an outcome measure. In that case, they run the risk of possibly encouraging blind scepticism without having the means of measuring whether they are actually teaching participants to improve their ability to appraise conspiracy theories critically.

6.3 Strengths

A particular strength of this research was that most experimental work adhered to strong open-science practices. My experimental studies were pre-registered in advance, and the data collected throughout all four papers in this research have been made publicly available on the Open-Science Framework. As noted in Chapter 2, many psychological studies on conspiracy thinking to date have not been pre-registered. By adhering to strong open-sciences practices, my data, methods, and analyses have been transparent, and replicable, facilitating reproducibility. The second key strength of this research was the focus on teaching and measuring critical thinking as a discrete skill, as opposed to measuring general conspiracy ideation of belief in specific conspiracy theories. Focusing on discrete skills allowed for assessments and interventions that were scalable in that they were not limited by time or cultural context

Finally, an important strength underlying the work in this thesis was the focus on measuring discernment between plausible and implausible conspiracy theories as opposed to measuring conspiracy beliefs. As highlighted by previous misinformation research, there is a considerable lack of work measuring discernment between plausible and implausible

information in general misinformation research. Guay notes that this oversight might mean that many common interventions that appear successful may merely be causing spillover effects and introducing response bias rather than teaching participants to properly distinguish between information that is reliable or unreliable. The work conducted in this thesis takes a fundamentally different approach to the psychological conspiracy research that precedes it by moving beyond measurements of conspiracy beliefs and focusing on the underlying thinking associated with endorsing conspiracy theories.

6.4 Limitations

Naturally, there are several limitations of this research. The first limitation is that this research primarily used cross-sectional designs, and did not conduct any longitudinal analysis. Consequently, no conclusions can be drawn regarding whether any of the interventions tested in this research have effects that last after a delay. This is an important factor regarding the findings of both Chapters 4 and 5, but unfortunately, it fell outside the scope of this thesis. I plan to address this limitation in my future studies which will examine whether these interventions' effects persist after periods varying from two weeks to four months. I also intend to examine whether “booster interventions” where participants receive a second intervention after some time has elapsed, can help ensure that the interventions lead to more lasting changes in participants' critical thinking skills – as has been demonstrated previously in other misinformation interventions (Maertens et al., 2021).

A major limitation seen in Chapter 5 with the development of *Nashville PI* was a lack of rigorous usability testing when the final game was developed. While I had conducted some pilot playtesting sessions to prototype various versions of the game in Chapter 5, there was not sufficient time during the project to conduct the playtesting necessary when the final version of *Nashville PI* was developed. As a result, there were several issues with the game's mechanics and user experience that hindered the player's ability to adequately develop the

desirable learning outcomes by the end of the game. This limitation is one that I aim to address in my future research, which will consist of several in-depth human-computer interaction-centred research with the primary aim of improving the game's playability in order to better facilitate learning among players.

Another limitation is that my research was confined to examining English-speaking, Western populations. In Chapter 2, I note that one limitation of existing conspiracy intervention research is a focus primarily on WEIRD populations. Due to the relatively higher word count of the CTAC developed in Chapter 3 compared to existing conspiracy thinking measures, and its emphasis on comprehension, I made a voluntary trade-off to focus exclusively on those who spoke English as a first language. This decision was made to mitigate the risk that reading ability or comprehension was not interfering with participants' ability to solve the problems posed in each item effectively. While the vignettes in the CTAC were novel, as a European researcher my exposure to Western conspiracy theories may have influenced how I wrote these vignettes. Consequently, the items used in my assessment may be more catered to Western audiences than other cultures. Future work could adapt the CTAC to other languages so that researchers can examine how various interventions impact critical appraisal of conspiracy theories using samples outside WEIRD populations.

Finally, another limitation of this thesis is that it focuses primarily on the cognitive factors of conspiracy thinking independent of political ideology (Imhoff et al., 2021; Uscinski et al., 2022), personality (Stasielowicz, 2022), socioeconomic factors (Enders et al., 2024; Salvador Casara et al., 2022), and culture (Bruder et al., 2013; Salali & Uysal, 2021). Naturally, all the factors listed above will undoubtedly have an impact on the cognitive factors underlying conspiracy thinking. Again, there was a voluntary trade-off made in the current thesis to “zoom in” to focus solely on critical thinking as a process that is important in the evaluation of conspiracy theories. Subsequent work might build on this by “zooming out”

and examining how critical thinking is (and is not) put into practice, depending on other factors such as social influence.

6.5 Future Directions

This thesis has contributed to understanding how critical thinking can be used to challenge unfounded conspiracy beliefs. As mentioned above, there are several limitations that I intend to address in my future work.

Given the cross-sectional nature of the experimental studies in Chapters 4 & 5, my work will use longitudinal study designs to examine whether the effects of the interventions I have designed persist over time. Additionally, due to the limitations encountered during the validation of *Nashville PI*, I am particularly interested in expanding the functionality and application of the game developed in Chapter 5. I plan to conduct more human-computer interaction research to improve the overall design of *Nashville PI* as an effective serious game. This work will include comparing how effective different forms of tutorials and walkthroughs are at familiarising players with the mechanics of the game in order to maximise player enjoyment and ease of play. I will also conduct design-focused playtesting interviews, where I will use deductive analysis to identify which aspects of the game players find challenging, and will make amendments to the game's narrative, mechanics, and user interface to improve the playability of the game.

My future work will also focus on testing how *Nashville PI* can be disseminated to those who need it the most (i.e. those who are not adept at critically appraise conspiracy theories). As those who believe in conspiracy theories tend to be distrustful of institutions, such as universities and their researchers, naturally one could expect that those who would be in most need of the interventions would be the most reluctant to voluntarily participate in

them. Therefore, one of my future studies will employ A/B testing to compare different means of disseminating and advertising *Nashville PI* to online communities that are centred around discussing conspiracy theories. My early work in this regard has shown that, somewhat contrary to our expectations as a research team, even those with extreme conspiracy theory beliefs were not overly deterred by the role of Google in funding the research reported in this thesis. What seems to be more important is not making the intervention sound dull (“*come and learn critical thinking skills*” is not enticing for most internet users). Instead, pitches that spark a sense of challenge and social comparison seem to be more effective (“*can you spot the real conspiracies that are being hidden? See if you can beat your friends’ scores*”). By achieving these goals, I will have not only identified an effective and scalable means of increasing critical appraisal of conspiracy theories but will also demonstrate whether it can lead to persistent changes among participants and determine the most effective ways to disseminate this intervention to those that are most in need of it.

In terms of more general future directions, the findings of this thesis provide a foundation for future research to focus on measuring discernment and the underlying process of conspiracy ideation, as a way forward to challenge unfounded conspiracy beliefs. At the time of writing, the majority of research investigating conspiracy ideation has focused primarily on belief in what are considered “false” claims as the primary outcome measure. Throughout this work, I have addressed why I believe that this approach is misguided. I argue that future research should focus on addressing three main issues in the field of conspiracy research. The first is to shift the focus of future research from measuring false beliefs and focus on teaching and measuring discernment and critical thinking. The second is to emphasise and encourage more longitudinal studies, which are largely sparse in the field. Finally, research should focus on the distinction between efficacy (performance of

interventions in controlled lab settings) and effectiveness (performance of interventions in practical ‘real-world’ settings) as highlighted by Roozenbeek et al, (2024).

6.6 Conclusion

This thesis has successfully shown that focusing on critical thinking is an effective and viable means of improving how psychologists measure and challenge unfounded conspiracy thinking. The studies in this thesis have aimed to improve our understanding of how people reason about conspiracy theories, and how we can best design interventions that challenge flaws and shortcomings associated with both unwarranted acceptance and dismissal of conspiracy theories. I have designed a novel outcome measure (Chapter 3), that has three advantages: (i) It focuses on the underlying cognitive processes of conspiracy thinking rather than just belief, (ii) it uses novel conspiracy theories, so researchers need not worry about prior beliefs or bias towards items, and (iii) it consists of both plausible and implausible conspiracy items as opposed to awarding points for dismissing all conspiracy theories. I have also conducted two experimental studies with applied benefits (intervention experiments in Chapters 4 & 5). These findings have important applied benefits, as they pave the way for more engaging and scalable interventions that can be applied to educational settings to challenge unfounded conspiracy theories within the public sphere. My work has also raised a number of theoretical considerations, mainly that existing approaches of challenging and measuring conspiracy thinking to date may have been unintentionally encouraging blind scepticism of conspiracy theories. My work highlights that it is important to both teach and measure discernment between plausible and implausible conspiracy theories in order to avoid merely creating response bias among participants. Given both the novel assessments and interventions, and the applied and theoretical implications, this thesis makes unique contributions to the psychological research surrounding conspiracy theories.

6.7 References

- Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring Individual Differences in Generic Beliefs in Conspiracy Theories Across Cultures: Conspiracy Mentality Questionnaire. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00225>
- Enders, A., Klofstad, C., Diekman, A., Drochon, H., Rogers De Waal, J., Littrell, S., Premaratne, K., Verdear, D., Wuchty, S., & Uscinski, J. (2024). The sociodemographic correlates of conspiracism. *Scientific Reports*, 14(1), 14184. <https://doi.org/10.1038/s41598-024-64098-1>
- Imhoff, R., Dieterle, L., & Lamberty, P. (2021). Resolving the Puzzle of Conspiracy Worldview and Political Activism: Belief in Secret Plots Decreases Normative but Increases Nonnormative Political Engagement. *Social Psychological and Personality Science*, 12(1), 71–79. <https://doi.org/10.1177/1948550619896491>
- Maertens, R., Roozenbeek, J., Basol, M., & Van Der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, 27(1), 1–16. <https://doi.org/10.1037/xap0000315>
- Roozenbeek, J., Remshard, M., & Kyrychenko, Y. (2024). Beyond the headlines: On the efficacy and effectiveness of misinformation interventions. *advances.in/psychology*, 2, e24569. <https://doi.org/10.56296/aip00019>
- Salali, G. D., & Uysal, M. S. (2021). WHY SOME HESITATE MORE: CROSS-CULTURAL VARIATION IN CONSPIRACY BELIEFS, BELIEF IN SCIENCE, AND VACCINE ATTITUDES. *medRxiv*, 2021.07.09.21260228. <https://doi.org/10.1101/2021.07.09.21260228>

Salvador Casara, B. G., Suitner, C., & Jetten, J. (2022). The impact of economic inequality on conspiracy beliefs. *Journal of Experimental Social Psychology*, 98, 104245.

<https://doi.org/10.1016/j.jesp.2021.104245>

Stasielowicz, L. (2022). Who believes in conspiracy theories? A meta-analysis on personality correlates. *Journal of Research in Personality*, 98, 104229.

<https://doi.org/10.1016/j.jrp.2022.104229>

Uscinski, J., Enders, A., Diekmann, A., Funchion, J., Klofstad, C., Kuebler, S., Murthi, M., Premaratne, K., Seelig, M., Verdear, D., & Wuchty, S. (2022). The psychological and political correlates of conspiracy theory beliefs. *Scientific Reports*, 12(1), 21672.

<https://doi.org/10.1038/s41598-022-25617-0>