The Social Cognition of Misinformation and Implications for Psychological Interventions

Cecilie Steenbuch Traberg
St Edmund's College

Social Decision-Making Lab
Department of Psychology
University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy,
January 2025

DECLARATION

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted, or is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

SUMMARY

Misinformation poses a significant threat to society, distorting public understanding of critical issues and undermining informed decision-making. From political falsehoods to health-related conspiracies, the pervasive spread of misleading information has far-reaching consequences. While much research on misinformation susceptibility has focused on cognitive factors - such as reasoning ability or analytical thinking - it often overlooks the social contexts in which judgements are made. This thesis addresses this gap by examining the role of social influence, specifically social proof, political identity and source effects, in misinformation susceptibility. Through a series of experiments and interventions, the thesis explores how the social context in which misinformation is consumed affects individual judgements of its veracity (Part I) and how this, in turn, influences the effectiveness of psychological interventions designed to reduce its impact (Part II).

Chapter 1 introduces the problem of misinformation and offers a social-psychological perspective on how individuals process information. Building on this, the chapter presents inoculation theory as a potential solution to the misinformation problem while highlighting its limited testing in social environments. **Chapter 2** presents two online experiments on source effects, demonstrating that news consumers are more susceptible to misinformation from (real) politically aligned media outlets, with this effect mediated via perceived source credibility. **Chapter 3** presents a two-by-two experiment that demonstrates the causal impact of both source credibility and source similarity on misinformation susceptibility - even for fictitious news sources. **Chapter 4** experimentally examines the effect of social cues, showing that individuals are more likely to believe misinformation when social proof - such as social media engagement - suggests that others find it reliable.

Chapter 5 presents a gamified intervention study demonstrating that inoculation protects news consumers against misinformation from political ingroup news outlets. **Chapter 6** presents an intervention study showing that a text-based inoculation intervention is effective in the presence of persuasive social cues, though even inoculated participants were influenced by these social cues. **Chapter 7** discusses the broader implications of these findings and situates them within the existing literature, offering directions for future research.

The research highlights how social context plays a powerful role in shaping people's susceptibility to misinformation. This thesis further demonstrates that while inoculation theory is effective in reducing misinformation susceptibility, certain contextual cues continue

to influence judgements, even in inoculated individuals. By providing a framework that integrates social identity theory, social influence research, and cognitive approaches, this thesis advances our understanding of how individuals process misinformation within social environments. These findings have significant implications for developing scalable interventions deployed across digital and social media environments, where much of today's news consumption and interaction occur.

ACKNOWLEDGEMENTS

This doctoral thesis marks the end of a journey that has taken longer than I initially imagined but one that has been infinitely richer for the experiences along the way. A week into my PhD, I discovered I was pregnant with my first child, and since then, I've had two more beautiful sons. Balancing a growing family with academic pursuits brought its fair share of challenges, pauses, and restarts - including three maternity leaves - but it also filled this journey with purpose and perspective.

I am deeply grateful to my supervisor, Professor Sander van der Linden, for his flexibility, encouragement, and unwavering support. Sander created an ideal environment that allowed me to explore my own ideas while providing the guidance and structure I needed to keep moving forward - even when I wasn't entirely sure where forward was. Thank you for giving me the space to thrive without losing my footing.

To Dr. Walther Christensen Traberg, who started this journey as my boyfriend and is now my husband and the father of our three wonderful sons: thank you for walking this path with me every step of the way. Your partnership, love, and unwavering belief in me made this possible, and I will always treasure the life we've built together.

To my dear friends and Cambridge PhD fellows, Sakshi Ghai and Laura Cashman: you made this PhD bearable in ways I cannot express. From lockdown phone calls that lasted for hours to long discussions about research, careers, and life in general, your companionship and insights kept me grounded and inspired. Thank you for being my sounding board, my confidantes, and my partners in crime during this crazy journey.

I am also grateful to the Social Decision-Making Lab, both past and present members, for creating a space where I could present my work - unfinished, confusing, or otherwise - with humour (bad humour included!) and honesty. Thank you for helping me make sense of my results and for creating an atmosphere of collaboration and genuine camaraderie.

To my parents: thank you for always believing in me and encouraging me to reach this milestone. Your love and support have been my foundation.

And finally, to my three sons - Theodore, Atlas, and Harrison - you are the heart of everything I do. Being your mother reminded me daily that life is so much bigger than a thesis and that my contribution to the world goes far beyond these pages. But I am proud of this thesis and what it represents, and I hope it is a meaningful contribution in its own small way. Thank you to everyone who has been part of this journey - it's been an unforgettable ride.

FUNDING

I would like to extend my sincere gratitude to the Harding Distinguished Postgraduate Scholarship and David and Claudia Harding for funding my Ph.D. I am especially grateful that the foundation offered additional support for student parents in my final year, which has been invaluable for my family.

I would further like to thank the Economic and Social Research Council for sponsoring my tuition fees during my Ph.D. and for offering additional training through the ESRC Doctoral Training Programme.

In addition, I would like to thank the Augustinus Foundation, the Direktør N. Bang og hustru Camilla Bang, født Troensegaards Legat, the Cambridge University Psychology Department and St. Edmund's College for helping me fund my research stays at Princeton and Harvard as well as my attendance at international conferences throughout my degree.

PUBLICATIONS

The work discussed here is published in peer-reviewed outlets. Below is an overview of the papers published based on this thesis:

Traberg, C. S., Morton, T., & van der Linden, S. (2024). Counteracting socially endorsed misinformation through an emotion-fallacy inoculation. *Advances in Psychology*, 2, e765332. https://doi.org/10.56296/aip00017

Traberg, C. S., Roozenbeek, J., & van der Linden, S. (2024). Gamified inoculation reduces susceptibility to misinformation from political ingroups. *Harvard Kennedy School (HKS) Misinformation Review*, 5(2). https://doi.org/10.37016/mr-2020-141

Traberg, C. S., Harjani, T., Roozenbeek, J., & van der Linden, S. (2024). The persuasive effects of social cues and source effects on misinformation susceptibility. *Nature Scientific Reports*, 14(1). https://doi.org/10.1038/s41598-024-42005-7

Traberg, C. S. (2024). Coercion by misinformation: Challenges and solutions. In *Coercion and Trust* (pp. 123–140). Routledge. https://doi.org/10.4324/9781003398233-8

Traberg, C. S., Harjani, T., Basol, M., Biddlestone, M., Maertens, R., Roozenbeek, J., & van der Linden, S. (2023). Prebunking against misinformation in the modern digital age. In T. D. Purnat & T. Nguyen (Eds.), *Managing Infodemics in the 21st Century*. Springer, Cham. https://doi.org/10.1007/978-3-031-27789-4_8

Traberg, C. S. (2022). Misinformation: Broaden definition to curb societal influence. *Nature*, 606, 653. https://doi.org/10.1038/d41586-022-01700-4

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, 136–151. https://doi.org/10.1177/00027162221087936

Traberg, C. S., & van der Linden, S. (2022). Birds of a feather are persuaded together: Perceived source credibility mediates the effect of political bias on misinformation susceptibility. *Personality and Individual Differences*, 185, 111269. https://doi.org/10.1016/j.paid.2021.111269

TABLE OF CONTENTS

PART I: THE SOCIAL COGNITION OF MISINFORMATION SUSCEPTIBILITY

1. INTRODUCTION

The reviewed literature, arguments and theory presented in this introductory chapter is published as the following theoretical, peer-reviewed pieces:

Traberg, C. S. (2024). Coercion by misinformation: Challenges and solutions. In *Coercion and Trust* (pp. 123–140). Routledge. https://doi.org/10.4324/9781003398233-8

Traberg, C. S., Harjani, T., Basol, M., Biddlestone, M., Maertens, R., Roozenbeek, J., & van der Linden, S. (2023). Prebunking against misinformation in the modern digital age. In T. D. Purnat & T. Nguyen (Eds.), *Managing Infodemics in the 21st Century*. Springer, Cham. https://doi.org/10.1007/978-3-031-27789-4_8

Traberg, C. S. (2022). Misinformation: Broaden definition to curb societal influence. *Nature*, 606, 653. https://doi.org/10.1038/d41586-022-01700-4

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, 136–151. https://doi.org/10.1177/00027162221087936

1. General Introduction

The Conceptual Complexity of Misinformation

In 2018, the World Economic Forum (WEF) identified misinformation as a significant global risk (World Economic Forum, 2018). Now, in 2024, the WEF's Global Risks Report has elevated misinformation to the foremost short-term global threat (World Economic Forum, 2024). This increasing recognition reflects its ability to distort public perceptions of critical societal issues, from climate change (van der Linden et al., 2017) to recent public health crises such as COVID-19 (van der Linden et al., 2023). Worryingly, misinformation can not just manipulate people's beliefs and perceptions about the world (Ecker et al., 2022) and spread fear and panic (Naeem & Ozuem, 2022); it can also have downstream effects on attitudes and behaviour. For instance, misinformation has been linked to reduced support for climate change mitigation (Cook et al., 2017; van der Linden, 2015), self-reported willingness to use violence to achieve political means (Jolley & Paterson, 2020), vaccine hesitancy and non-compliance with public health guidelines (Loomba et al., 2021; Roozenbeek, Schneider, et al., 2020; van der Linden, 2022) and even disruptions to election integrity (Green et al., 2022; Watts et al., 2021). Given these impacts, it is crucial to understand the psychological and social mechanisms that drive belief in misinformation in the first place, as such insights are essential for developing effective countermeasures (Ecker et al., 2022; van der Linden, 2022).

To examine the psychological underpinnings of belief in misinformation and the social and cognitive mechanisms facilitating its spread, it is necessary first to define what constitutes misinformation (sometimes referred to as 'fake news') - an issue that remains the subject of scholarly debate (Vraga & Bode, 2020). Lazer et al. (2018) define 'fake news' as *"fabricated information that mimics news media content in form but not in organisational process or intent"* (Lazer et al., 2018, *p.* 1094), highlighting the defining element of fake news, or misinformation, to be at the level of the publisher. Researchers who adopt this definition tend to operationally classify content as misinformation if it originates from publishers deemed untrustworthy - often through credibility ratings provided by organisations such as NewsGuard or Politico (Grinberg et al., 2019). However, attributing "fakeness" not at the level of the story but at that of the publisher risks overlooking both accidental misinformation and the more insidious role of misleading - but not entirely false - information from otherwise credible or popular outlets (van der Linden & Kyrychenko, 2024). Mainstream and popular media sources may benefit from higher levels of trust (C. S.

Traberg, 2022), and so misinformation spread by those sources may have worse downstream effects on news consumers' beliefs, opinions, and behaviours, as it may be more persuasive (van der Linden & Kyrychenko, 2024). As Traberg (2022) notes, *"misleading information from popular sources could be more convincing than extreme falsehoods, and so potentially more harmful to beliefs and behaviour."* This perspective is supported by recent findings demonstrating that during the COVID-19 pandemic, ambiguous misinformation about vaccines - factually accurate but deceptive content - was 46-fold more consequential for driving vaccine hesitancy than flagged misinformation (Allen et al., 2024). Contrasting with source-based definitions, other scholars focus on the content level, defining misinformation based on the factual accuracy of the information itself, regardless of the source (Ludwig & Sommer, 2024). Under this approach, any content that is fact-checked and determined to be false is considered misinformation, even if it originates from mainstream or reputable media outlets (Guess et al., 2019; Pennycook et al., 2018).

However, both source-based and content-based operationalisations risk oversimplifying the complex nature of misinformation. As highlighted by van der Linden & Roozenbeek (2020), misinformation may be better described as a spectrum ranging from entirely fabricated content, often termed "fake news", to information that is technically accurate but misleading, distorted, or presented without appropriate context (Traberg & van der Linden, 2022; van der Linden & Roozenbeek, 2020). Entirely fake information can originate from various sources such as satire websites like the *Onion* or even individual *TikTok* users looking to be controversial or claim to offer "good advice" (Mulcahy et al., 2024); from which the information may be picked up by other social media users or bogus news outlets causing it to go viral (Blunt, 2023). Examples of this type of misinformation include the rumour that *Disney World* in Florida was lowering its drinking age to 18 (Snopes, 2022) or the infamous conspiracy rumour "PizzaGate" that suggested Hilary Clinton was running a child-prostitution ring out of a Washington, D.C. pizzeria (New York Times, 2020).

With entirely factual information at the opposite end of the spectrum (e.g., true reportings that do not attempt to mislead, spin or decontextualise the truth), in the middle of the spectrum exists misleading information, representing a distortion of the truth (Traberg, 2024). Misleading information originates from facts but has been altered to the extent that it either intentionally or unintentionally conveys a message no longer interpreted in the same way as the original facts. For example, a misleading video, initially showing poll workers in Russia during the 2018 election, resurfaced in 2021 on social media with false claims that it depicted poll workers in Flint, Michigan, committing voter fraud in the US election of 2020

(Reuters, 2020). The video circulated widely and was shared nearly 2,000 times on Facebook, contributing to false allegations of voter fraud in Michigan. In this case, there was nothing fake about the video; however, it was placed in a misleading context that led people to reach inaccurate conclusions about reality.

In support of a broader definition of misinformation, recent surveys of misinformation experts (Altay, Berriche, Heuer, et al., 2023) reveal a broad consensus that misinformation comprises both false and misleading information. Indeed, the recent American Psychological Association's consensus statement on misinformation emphasises that misinformation is not just about falsehoods but also includes misleading information that can cause misunderstanding or misinterpretation (van der Linden et al., 2023). Similarly, Lewandowsky et al. (2017) argue that misinformation can encompass a range of inaccuracies, from simple errors to complex distortions that exploit cognitive biases.

Intent is another significant aspect when it comes to defining misinformation. As van der Linden & Roozenbeek (2020) note, false or misleading information can be created and shared without any deliberate attempt to deceive, for example, as a result of human error or misinterpretation. For instance, a medical writer might mistakenly misconstrue findings from a peer-reviewed study, or a journalist could base a story on information published by another outlet, which later proves incorrect. During the height of the COVID-19 pandemic in 2020, a widely-read Danish news outlet misinterpreted the findings of a controversial scientific study on mask efficacy (Rosenbaum, 2020), leading it to publish a headline suggesting that face masks were ineffective against COVID-19 (Bundgaard et al., 2020). Other national and international outlets soon echoed this interpretation, and the resulting headlines spread widely on social media when public demand for precise and accurate information was at its peak. Although the Danish outlet later apologised and amended the misleading headline (Rysgaard Møller, 2020), such retractions are rarely fully effective (Lewandowsky et al., 2012).

Further detailing the role of intent, disinformation typically refers to cases in which false or misleading information is disseminated with deliberate intent to deceive, confuse, or cause harm – with propaganda representing a form of disinformation with a specific political agenda (Benkler et al., 2018). One prominent example of disinformation occurred during the 2016 U.S. Presidential Election, where fabricated articles claimed that Pope Francis had endorsed Donald Trump for president, a false story intentionally created and disseminated by content farms to manipulate public opinion and generate ad revenue (Allcott & Gentzkow, 2017). Although it was entirely untrue, the story spread rapidly across social media, reaching millions of users before being debunked.

Despite the existence of a multitude of terms to describe falsehoods circulating online, such as misinformation, disinformation, and propaganda - their psychological impact on receivers may be driven by similar cognitive mechanisms, as the intent behind the information is often inaccessible to receivers at the moment they encounter it. For this thesis, therefore, *misinformation* will, in line with APA (van der Linden et al., 2023), serve as an umbrella term encompassing all forms of misleading and false information, regardless of the intent behind its creation or dissemination.

Much like the scholarly debates surrounding the definition of misinformation, there has also been considerable discussion about the extent to which it should be a societal concern. Scholars like Acerbi et al., (2022) and Altay et al., (2023) adopt what Ecker et al. (2024) call the 'minimising position', arguing that fears around misinformation's influence are overstated, mainly highlighting its supposed low prevalence. However, Ecker et al. (2024) challenge these claims, arguing that such conclusions often rely on overly narrow definitions that exclude misleading content disseminated through mainstream channels, closed groups, and political rhetoric. By broadening the scope of what constitutes misinformation - including misleading arguments and decontextualised facts, as done in this thesis, it becomes clear that its effects are far from negligible – from fuelling vaccine hesitancy during global health crises (Allen et al., 2024), igniting violence against minorities through coordinated campaigns (Yanagizawa-Drott, 2014), and exacerbating climate change denial (Ejaz et al., 2024). As such, understanding the psychological and social drivers of belief in misinformation remains a crucial step in mitigating its effects.

From Scrolls to Screens

Misinformation is far from a recent phenomenon. Ancient Roman politicians, such as Caesar and his rivals, weaponised rumours to discredit opponents through speeches, letters, and early scrolls (Morstein-Marx, 2004), a tactic echoed by powerful figures throughout history (Jowett & O'Donnell, 2018). During wars, state-sponsored propaganda campaigns were used to manipulate public sentiment and maintain morale (Ellul, 1973), and ideologically charged misinformation has been used to destabilise opponents' influence (DeFleur & Ball-Rokeach, 1989). However, the digitalisation of information consumption has radically transformed misinformation's speed, reach, and impact (Del Vicario et al., 2016). Today, false information spreads further and faster than truths online, partly driven by algorithms prioritising sensational content (Vosoughi et al., 2018). This effect is further amplified by human tendencies to engage with headlines that exploit outrage (McLoughlin et

al., 2024) and outgroup derogation (Rathje et al., 2021) - features often seen in fake news headlines (van der Linden et al., 2023). Additionally, modern digital platforms overwhelm users with a flood of information, creating an "information overload" environment (Rodriguez et al., 2014). In such contexts, limited cognitive resources may make individuals less likely to critically evaluate content, potentially instead relying on heuristic-driven processing (Metzger et al., 2010) and leaving them more vulnerable to misinformation (Lu & Xiao, 2024). As such, both cognitive factors and group biases may contribute to why individuals fall for misinformation, and dual-process models of reasoning and motivated reasoning perspectives provide valuable insights into the underlying mechanisms, as discussed in the following section.

Cognitive and Motivated Reasoning Accounts of Misinformation Susceptibility

In seeking to explain why individuals "fall for fake news", two main perspectives have dominated the literature in recent years: A cognitive account (Pennycook & Rand, 2019b) and a motivated reasoning account (Kahan, 2013). Cognitive accounts (sometimes referred to as dual-process models, see Pennycook & Rand, (2021)) argue that misinformation susceptibility largely stems from a reliance on "fast" or heuristic-based thinking (Pennycook & Rand, 2019b) - also known as System 1 processing in dual-process theory (Stanovich & West, 2000). This perspective posits that individuals fall for misinformation because they rely on intuitive, non-analytical thinking, which requires less cognitive effort. In contrast, individuals who engage in "slow," reflective processing (System 2) are argued to be better equipped to discern misinformation from accurate information (Pennycook & Rand, 2019b). A widely used measure in this domain is the Cognitive Reflection Test (CRT), which assesses an individual's tendency to override intuitive responses in favour of more reflective reasoning (Frederick, 2005). Research has shown that individuals with higher cognitive reflection - those who tend to engage in analytic thinking rather than relying on initial, gut-level judgements - are less likely to believe and share misinformation (Bago et al., 2020; Pennycook & Rand, 2019b, 2020).

In support of this perspective, accuracy nudges, which serve as brief reminders to assess the accuracy of information before accepting or sharing it (theoretically prompting a shift from intuitive to reflective processing) have been found to reduce both sharing and believing in misinformation (Pennycook et al., 2020, 2021). However, replication studies have revealed effect sizes so small (Gavin et al., 2022; Roozenbeek, Freeman, et al., 2021), that they would be considered 'negligible' according to Cohen's benchmarks (Cohen, 1988).

Furthermore, their effectiveness is limited among conservatives (Rathje et al., 2021), who are more likely to believe (Garrett & Bond, 2021) and share misinformation (Guess et al., 2019) - potentially because they engage in less analytical reasoning (Talhelm et al., 2015), because they are disproportionately exposed to misinformation from right-wing media (Grinberg et al., 2019), or because misinformation in research studies often has a pro-right ideological slant (Guess et al., 2019).

Reliance on different cognitive processing styles is often discussed as contrasting with the motivated reasoning account (Van Bavel et al., 2021). The motivated reasoning account of misinformation posits that individuals process information in ways that confirm their pre-existing beliefs and identities, leading them to judge information supporting their views as more accurate and to dismiss information that challenges their views as inaccurate (Charness & Dave, 2017; Kahan, 2013; Kahan et al., 2017; Van Bavel et al., 2021). This ideological belief bias, sometimes referred to as "identity-protective cognition" (Kahan, 2013), is suggested to lead people to accept information aligned with their partisan identity while rejecting information from opposing perspectives. Research across disciplines (Bisgaard, 2019; Kahan, 2013; Schaffner & Roche, 2017; Van Bavel & Pereira, 2018) suggests that people engage in motivated reasoning when assessing the veracity of information, often influenced by confirmation bias - a tendency to favour information that confirms one's beliefs (Charness & Dave, 2017). Relatedly, the "You are fake news" effect describes the tendency for individuals to label information from opposing political sides as fake simply because it originates from the political out-group (van der Linden et al., 2020).

Interestingly, research has shown that when individuals are incentivised to be accurate, partisan biases in misinformation judgements can be reduced by as much as 30% (Rathje et al., 2023). This suggests that accuracy motivations can attenuate some of the effects of motivated reasoning, although partisan biases remain a significant factor in misinformation susceptibility. Supporting this, Roozenbeek et al. (2022) found that myside bias and political partisanship were among the strongest predictors of susceptibility to misinformation, outweighing the role of analytical thinking. These findings highlight that motivated reasoning plays a key role in misinformation susceptibility, as individuals are more likely to believe and share information that aligns with their ideological positions (Pereira et al., 2021).

While both cognitive and motivated reasoning perspectives offer valuable insights into why individuals fall for misinformation, they largely conceptualise information consumption as a unidimensional process between the message producer and consumer, as if

individuals engage with content in a social vacuum (Traberg et al., 2024). However, the rapid spread of misinformation online reflects the inherently social nature of information consumption in the digital age, as news consumers increasingly turn to social media for news (Atske, 2021) and information (Anderson, 2021; Matsa, 2021). Unlike traditional media, where information flows through controlled channels with editorial gate-keepers (Rhodes, 2022; Williams & Carpini, 2011), social media platforms allow both news and misinformation to be embedded into dynamic ecosystems designed initially for interpersonal communication (Cheung et al., 2011), where each individual acts as both a potential recipient and source of misinformation (Del Vicario et al., 2016). Modern misinformation, therefore, operates in socially embedded ecosystems, where the visible social cues surrounding misinformation – such as the source of the message and the reactions of other users, are technically inseparable from the information (Traberg et al., 2024).

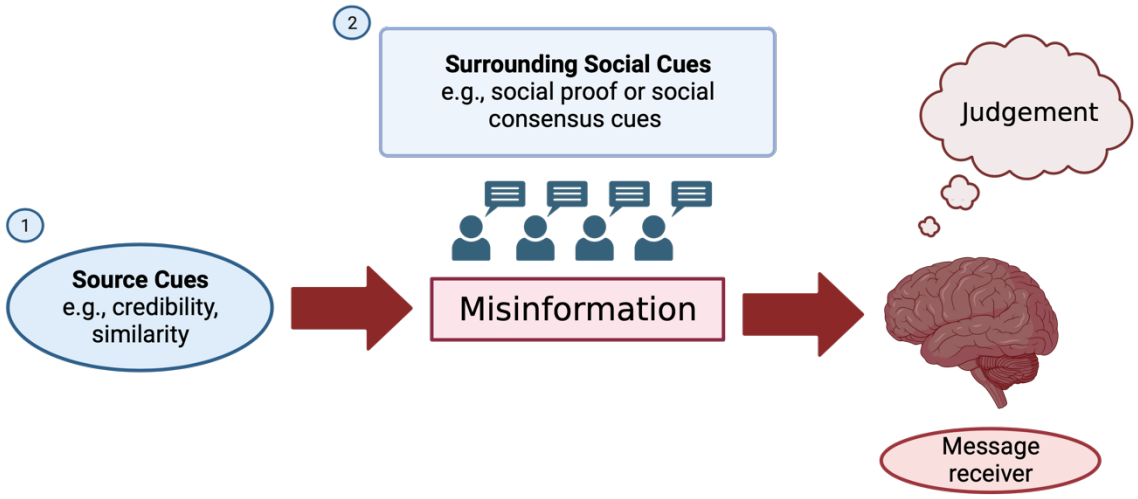
The pervasive nature of social information on these platforms means that people are consistently exposed to contextual cues that accompany the news they consume, which may influence judgements and assessments of information reliability (Dvir-Gvirsman, 2019; Kim, 2018; Lee et al., 2021; Luo et al., 2020). Therefore, understanding the psychological drivers of misinformation requires examining individual-level cognitive processes and broader social dynamics that shape engagement with misinformation in digital contexts. The following section conceptualises the digital 'social context' and proposes a lens through which to examine its effects on misinformation susceptibility.

The Social Context of Information Consumption

When individuals encounter misinformation in digital environments, they are exposed not just to the content but also to the social context surrounding the message. This thesis proposes that the social context can be broadly conceptualised through two primary cues: (1) *source cues* - characteristics of the message sender and (2) *social cues* - visible or inferred judgements from others, which may or may not reflect social consensus or approval (see Figure 1).

Figure 1

The Social Context of Misinformation



Source: Reprinted with permission from Traberg et al. (2024).

Note: (1) Misinformation content originates from a source for which source cues are present. (2) Surrounding social cues may be present in the form of social proof (e.g., 'likes' or comments) or social consensus cues (e.g., explicit judgements of others).

To illustrate, consider a typical social media headline by a news organisation in Figure 2. This example demonstrates that when individuals encounter a headline on social media, it is accompanied by contextual signals about the source and the judgements of others. While prior work has explored the effects of sources (Chaiken & Maheswaran, 1994; Petty & Wegener, 1998) and social cues (Cialdini & Goldstein, 2004) on persuasion, their impact in digital misinformation environments remains underexplored. Furthermore, although recent models have begun to incorporate the informational context into belief formation (Van Bavel et al., 2024), much prior empirical research has primarily focused on individual-level cognitive (Fazio et al., 2015; Pennycook et al., 2018, 2021; Pennycook & Rand, 2019b) or motivational processes (Kahan, 2013), often neglecting an examination of the role of social contexts in shaping how people may engage with misinformation. Recent scholarship has therefore called for *"a more integrated framework that is sensitive to context and can account for varying susceptibility to misinformation"* (van der Linden, 2022, p. 465)

Figure 2

Example of a News Headline in a Social Media Context

Note: This figure shows a news headline accompanied by *source cues* (e.g., “Fox News”) and *social cues* (e.g., comments, likes, and shares). The source provides information about the origin of the content, while visible user reactions reflect social judgements that may influence how the content is perceived and evaluated.

Moreover, while decades of research in persuasion and attitudes has uncovered the social psychological underpinnings of influence and message acceptance, with a strong focus on the role of information sources (Briñol & Petty, 2009; Chaiken et al., 1996; Hovland et al., 1953; Maio & Haddock, 2010; O’Keefe, 2016; Petty & Cacioppo, 1986; Petty & Wegener, 1998), this literature remains surprisingly underutilised in studying misinformation (Mang et

al., 2024). Building on the role of sources, research on social identity theory (SIT) (Tajfel & Turner, 1979) and self-categorisation theory (SCT) (Turner & Oakes, 1989) demonstrates that internalised group memberships can influence information processing as individuals are more likely to accept information presented by ingroup members (Ellemers et al., 1997). In addition, social influence research illustrates that the presence of others impacts cognition and judgements (Baron et al., 1996; Burnstein & Vinokur, 1977; Cialdini & Goldstein, 2004; Cialdini & Trost, 1998; Deutsch & Gerard, 1955; Maio & Haddock, 2010; Moussaïd et al., 2013). These fields offer highly relevant and important lenses through which to examine the potential influence of source and social cues in social media environments.

This thesis aims to bridge this gap by drawing on key social psychology perspectives - attitudes and persuasion (Petty & Cacioppo, 1986), social influence (Cialdini & Goldstein, 2004), and social identity research (Tajfel, 1978; Tajfel & Turner, 1979) - to develop a more comprehensive understanding of the social processes that underpin susceptibility to misinformation in digital environments. The following section explores this research in greater depth, highlighting its potential application to understanding and addressing susceptibility to misinformation.

2. Towards a Social Psychological Account of Misinformation Susceptibility

Psychological Models of Persuasion

Over the past 40 years, psychological models of persuasion have primarily been rooted in dual-process theories, positing that persuasion can occur through two routes: a peripheral or heuristic route, which involves less mental processing, and a central or systematic route, which requires more cognitive effort (Chaiken et al., 1989; Petty & Cacioppo, 1986). The Elaboration Likelihood Model (Petty & Cacioppo, 1986) is one such model that suggests that the likelihood of individuals processing a message through either route depends on factors such as motivation and ability. Motivation may stem from personal relevance or need for cognition (Cacioppo & Petty, 1982), while distractions, prior experience, or message complexity could influence the ability to process the message (Cacioppo & Petty, 1982). In highly motivated and able individuals, central route processing is argued to occur, leading them to examine information and arguments carefully. Conversely, when motivation or ability is low, individuals may rely on peripheral cues, such as sources (Pornpitakpan, 2004) or social cues (Metzger et al., 2010), to assess a message (Petty & Cacioppo, 1984).

To reiterate, dual-process theories argue that to engage in the systematic processing of content, individuals must be motivated and in environments that enable them to use a high level of mental effort in processing information (Petty & Cacioppo, 1986). However, this contrasts with how people are likely to engage with online information (Feng et al., 2015), where news consumers feel overloaded with the amount of news they are confronted with (Holton & Chyi, 2012), and where people may cope with information overload by minimising their cognitive effort and time through the use of heuristic-based processing (Metzger et al., 2010). Indeed, previous research challenges the assumption that people carefully assess online information in isolation and instead shows that users frequently rely on group-based tools like reviews or testimonials to make credibility judgements regarding information (Metzger et al., 2010). More recent evidence reveals that up to 75% of social media content is shared without users clicking or reading further, highlighting a potential lack of systematic engagement (Sundar et al., 2024).

Social Cues

Dual-mode models also provide a framework for understanding attitude change in social settings (Wood, 2000). Employing a modified version of the classic Asch conformity study (1956), Baron et al., (1996) demonstrated that when participants were highly motivated to be accurate, they were more likely to rely on their own evaluations of stimuli. In contrast, when accuracy motivation was moderate, individuals were more likely to rely on the judgements of others - even when they could determine the correct answer themselves. This finding raises questions about how individuals evaluate on social media: Do they employ heuristic-driven evaluations influenced by visible social judgements of others, or do they systematically assess content despite the social context?

The distinction between informational and normative influence, as outlined by Deutsch & Gerard, (1955) provides a useful lens for understanding how social cues may operate. Informational influence occurs when individuals adopt group judgements because they perceive them as more accurate. In contrast, normative influence reflects the desire to conform to group norms to gain social approval or avoid rejection. While these two types of influence may interact in digital contexts (Price et al., 2006), examining normative influence causally can be challenging within an experimental paradigm, as participants are not exposed to real social groups and thus lack the social stakes - such as fear of rejection - that typically drive normative conformity. Although real social interaction can be mimicked in social games

or simulations (Hertz et al., 2017), this thesis primarily focuses on informational influence and individual assessments of (mis)information reliability.

Social proof represents a specific form of informational social influence, referring to the tendency to rely on the actions or judgements of others when making judgements (Cialdini & Goldstein, 2004). Its impact has been documented across various domains, from e-commerce (Walther et al., 2012) to movie selection (Xu & Fu, 2014). In the context of misinformation, social proof may be manifested through engagement metrics such as 'likes'or 'comments'. Some studies have examined whether both contextual cues impact misinformation susceptibility, but results are highly mixed and lack generalisability due to the limited scope of misinformation stimuli used (Ali et al., 2022; Butler et al., 2023; Kim, 2018; Mena et al., 2020; Pehlivanoglu et al., 2021). That is, while some studies find that such cues increase the perceived credibility of misinformation (Butler et al., 2023; Luo et al., 2020), other research reports no significant effect of such cues (Mena et al., 2020; Traberg et al., 2024). These mixed results suggest that social proof may not always have the same influence on judgements, indicating the need for further investigation into the mechanisms that mediate these effects.

One potential mechanism is the role of social proof in shaping perceptions of social consensus. Social consensus refers to the perception of agreement or shared judgement within a group (Clarkson et al., 2013). Scholars have distinguished between two forms of consensus information: *implicit* consensus, referring to beliefs about what others would do if they were present, and *explicit* consensus, being the more direct or actual behaviour of others (Kassin, 1979, p 1967). Both forms have been shown to influence individual level beliefs (Hewstone & Jaspars, 1988; Kassin, 1979), as well as attitudes and behaviours in areas such as climate change (Lewandowsky et al., 2019) and regarding social issues (Gardikiotis et al., 2005). In fact, social consensus has been named one of the most important factors in determining whether people conform to others (Gardikiotis et al., 2005; Gerard et al., 1968). Particularly under conditions of cognitive overload or low motivation online, it is therefore possible that visible social cues online may act as proxies for consensus (Clarkson et al., 2013), potentially creating a "false consensus" effect wherein individuals overestimate the prevalence of certain beliefs or opinions (Gardikiotis et al., 2005). However, this theory has yet to be empirically tested.

To provide a preliminary summary of the current understanding and gaps in knowledge regarding the influence of social cues: While prior research has explored the role of social cues in shaping susceptibility to misinformation, findings remain inconsistent, and

the mechanisms through which social cues exert their influence are yet to be fully understood. For example, why do social cues enhance the perceived reliability of misinformation in some contexts but not in others? This thesis seeks to address these gaps by examining whether individuals are influenced by both *implicit* consensus cues (e.g., comments that implicitly endorse a headline's veracity) and *explicit* group consensus cues (e.g., explicit numerical representations of group judgements). Additionally, it investigates whether such 'local' group cues shape perceptions of broader public consensus and, in turn, how these perceptions impact individual-level judgements of misinformation.

While social cues, such as group consensus, may signal collective approval or disapproval, they represent only one dimension of the contextual signals that this thesis proposes may influence information processing. Indeed, previous research in persuasion suggests that the associated source of information may serve as a powerful heuristic cue in determining the likelihood of persuasion (Pornpitakpan, 2004). The following section explores the influence of source cues on influence processes and its potential applications to understanding misinformation susceptibility.

Source Factors

In the field of attitude research, there is considerable evidence that source credibility increases the likelihood of persuasion (Albarracín et al., 2017; Chaiken & Maheswaran, 1994; Kumkale et al., 2010; Metzger et al., 2003; Pornpitakpan, 2004; Sternthal et al., 1978). Specifically, the more credible a source is perceived to be, the more likely people are to evaluate persuasive messages from those sources positively (Briñol & Petty, 2009). While several factors contribute to credibility evaluations, the perceived similarity between the source and message receiver plays a particularly important role in persuasion (Chaiken & Maheswaran, 1994; Metzger et al., 2003), potentially because individuals assume that like-minded others are more likely to be correct (Aune & Kikuchi, 1993; Gino et al., 2009) or because messages linked to a similar source increase their self-relevance (Briñol & Petty, 2009). For instance, Marks et al. (2019) showed that people prefer to get advice from advisors who have a similar political leaning, even when they have evidence that they are less competent at the task at hand.

In a separate but related stream of research, the Social Identity Theory Approach to Communication (Greenaway et al., 2015) suggests that social identity and group memberships play a large role in how people process information. This approach draws on Social Identity Theory (SIT) (Tajfel & Turner, 1979) and Social Categorisation Theory (SCT)

(Turner & Oakes, 1989), two theories which both assert that individuals internalise group membership as part of their self-identity, and that they evaluate the world around them in accordance with these identities. Similar to attitude researchers, SIT researchers assert that individuals evaluate information based on whether or not the source is perceived to be an ingroup or outgroup member (Ellemers & Haslam, 2012; Pereira et al., 2021; Van Bavel & Pereira, 2018). The group identity literature suggests that social categories to which we belong can imply similarity, and under uncertainty, such as when evaluating information online, we may think ingroup member judgements can be informative because we assume that they will have a similar view to us (Spears, 2021). Indeed, there is good evidence to suggest that ingroup sources are more persuasive than outgroup sources (Mackie et al., 1990), but only when group membership is made salient (McGarty et al., 1994) and, in particular, when the source is a prototypical group member (van Knippenberg et al., 1994). Individuals are also more likely to cognitively elaborate on messages when they stem from a source with the same group membership (Mackie et al., 1990). Together, these findings raise the question of whether individuals are more likely to be influenced by misinformation when the associated source shares the recipient's group membership and whether this serves as a credibility cue.

Researchers have already started to consider how shared group membership and source similarity impact our judgements of misinformation (Pereira et al., 2021; Van Bavel et al., 2024). Studies have found that individuals are more likely to believe misinformation if it comes from sources that have previously published attitudinally congruent content (Bauer & Clemm von Hohenberg, 2021; Nadarevic et al., 2020). This effect is especially pronounced when the source aligns with the individual's political or ideological identity, as seen in research where Republican supporters of former President Trump were more likely to believe misinformation attributed to him compared to the same information presented without source attribution (Swire et al., 2017). This work underscores the persuasive influence of sources perceived as 'like-minded,' and that source characteristics can shape assessments of information reliability independently of the content itself.

However, past research has not systematically examined the (potential) interplay between source credibility and similarity in shaping susceptibility to misinformation. Specifically, it remains unclear whether ingroup news sources are inherently perceived as more credible and whether this perception drives susceptibility. Furthermore, what happens when a source is explicitly labelled as lacking credibility? Will partisan source biases persist under such conditions? This thesis addresses these gaps by experimentally manipulating

source cues, such as credibility and group similarity, to investigate their independent and potential interactive effects on the perceived reliability of misinformation, providing a more nuanced understanding of how sources shape belief formation.

Bridging the Gap: Contextual Cues and Implications for Interventions

To summarise the discussed literature and highlight future research directions, it is evident that individuals increasingly consume information within social contexts enriched by both source cues (Bauer & Clemm von Hohenberg, 2021) and social cues (Lee et al., 2021). Despite this, much previous research on misinformation susceptibility has examined the phenomenon in isolation, neglecting the potential influence of social context on judgements. Existing theories of judgement formation and message elaboration suggest (Briñol & Petty, 2009; Chaiken & Maheswaran, 1994; Petty & Cacioppo, 1986) - but do not confirm - that these social contextual cues play a role in shaping susceptibility to misinformation. The first goal of this thesis is to address this gap by investigating how the *social context* - operationalised here as source cues (e.g., political identity, perceived credibility) and social cues (e.g., perceived consensus, social proof) - influences individuals' susceptibility to misinformation. Specifically, this thesis seeks to uncover how these cues operate, through which mechanisms they exert their influence, and what mediating variables make them impactful.

Implications for Psychological Interventions

If, as hypothesised, the social context influences susceptibility to misinformation, this finding will have significant implications for intervention strategies to reduce misinformation influence. A promising and growing body of research aims to pre-emptively build psychological resistance to misinformation (Banas & Rains, 2010; Compton et al., 2021; Traberg et al., 2022; van der Linden, 2022). However, these interventions have largely conceptualised news consumption as a one-dimensional relationship between (mis)information producers and consumers (Cook et al., 2017; Mason et al., 2023; Roozenbeek & Linden, 2018; Roozenbeek & van der Linden, 2019; van der Linden, 2022). This approach overlooks the inherently social nature of information environments, where judgement formation rarely occurs in isolation.

Interventions that demonstrate promising results under controlled conditions (Basol et al., 2020, 2021; Roozenbeek, Traberg, et al., 2022; Roozenbeek, van der Linden, et al., 2022, 2022), where misinformation is studied in isolation, may yield uncertain outcomes when applied to real-world social contexts. To examine the true efficacy of inoculation strategies, it

is essential to account for the role of social context in shaping how individuals process and resist misinformation. The following section introduces inoculation theory as a potential solution to misinformation and explores how social contexts may moderate its effectiveness in practice.

3. Psychological Inoculation against Misinformation

Theoretical Foundations

Just as researchers have been concerned with understanding the underpinnings of misinformation susceptibility, so too have they sought to understand how individuals can better resist the influence of misinformation (Basol et al., 2020, 2021; Compton et al., 2021; Cook et al., 2017; Roozenbeek, Traberg, et al., 2022; Roozenbeek & van der Linden, 2019; Traberg et al., 2022). A promising area of research which has witnessed increasing attention in recent years is centred around inoculation theory or so-called psychological 'vaccines' against persuasion (Compton et al., 2021; Lewandowsky & van der Linden, 2021; Traberg et al., 2022; van der Linden, 2022). Introduced by McGuire in the 1960's, inoculation theory draws parallels between psychological resistance to persuasive attacks and the immunisation process against viral contagion (McGuire, 1964; McGuire & Papageorgis, 1962). The theory employs a biological metaphor, suggesting that resistance to persuasion functions similarly to immune responses in medical immunisation: by experiencing and successfully refuting weaker versions of persuasive attacks, individuals develop cognitive resistance that helps them withstand stronger yet still misleading arguments (McGuire, 1964). McGuire & Papageorgis (1961) hypothesised that it would be more effective to expose individuals to a weakened version of an impending persuasive attack (and preemptively refute it) than to simply bolster their existing beliefs with supportive facts. Their experiments validated this hypothesis, confirming the relative efficacy of inoculation over traditional correction methods (McGuire, 1964).

An inoculation message typically consists of two components: threat and refutational preemption. The *threat* component involves making individuals aware of an imminent persuasive attack, such as forewarning them that political actors may attempt to mislead them on issues like climate change or energy technologies . *Refutational preemption* (or prebunking) refers to providing individuals with counterarguments they can use to defend their beliefs when exposed to future persuasive attempts (Compton & Pfau, 2005). These counterarguments are stored cognitively and retrieved when similar (but stronger) arguments arise later. Early work on inoculation theory found evidence that these interventions could in

fact protect individuals against persuasive attacks against non-controversial 'cultural truisms' (McGuire & Papageorgis, 1962), such as the common knowledge that 'toothbrushing helps prevent tooth decay'. McGuire argued that testing inoculation on these uncontested beliefs provided a "cleaner" test of the theory, as individuals had generally not encountered counterarguments against these widely accepted truths (McGuire, 1964). Over time, however, researchers have demonstrated that inoculation can be applied to more polarised and controversial issues, including climate change (van der Linden et al., 2017), genetically modified organisms (GMOs) (Wood, 2007), extremism (Braddock & Dillard, 2016; Lewandowsky & Yesilada, 2021), and conspiracy theories (Banas & Miller, 2013). A meta-analysis across forty studies involving over ten thousand participants found an average effect size of $d = 0.43$ for inoculation interventions (Banas & Rains, 2010), a medium effect that Cohen (1988) describes as "large enough to be visible to the naked eye" (p. 26). A more recent meta-analysis has found that inoculation interventions specifically targetting misinformation had an average effect size of $d = 0.20$ for truth discernment (the ability to discern between true and false information) as the main outcome variable (Lu et al., 2023).

Building on this foundational work in inoculation theory, researchers have explored the use of gamified interventions to scale inoculation theory in a way that is both engaging and effective. One prominent example is the award-winning *Bad News* game (Basol et al., 2020; Roozenbeek, Traberg, et al., 2022; Roozenbeek & van der Linden, 2019), which allows players to assume the role of misinformation producers, encouraging them to critically reflect on the tactics used to manipulate public opinion. By engaging with these tactics firsthand, players develop mental defences against similar strategies in the real world.

Field studies and randomised controlled trials have demonstrated the efficacy of *Bad News* in inoculating players against various misinformation techniques, such as impersonation, polarisation, and the use of moral outrage (Basol et al., 2020; Roozenbeek & van der Linden, 2019). Similar interventions, such as the *Go Viral!* game and *Harmony Square* have also shown success in inoculating individuals against COVID-19 misinformation and political disinformation, respectively (Basol et al., 2021; Roozenbeek & Linden, 2020). This form of inoculation, where players actively engage in creating misinformation within a controlled setting, aligns with the belief of the original proponents of inoculation theory (McGuire & Papageorgis, 1962), who proposed that such involvement may enhance learning and retention (Tyler et al., 1979). In fact, work by Maertens et al. (2021) has shown that inoculation effects can remain stable for at least three months.

Inoculation interventions thereby offer a promising approach to countering misinformation susceptibility, providing individuals with cognitive resistance against persuasive attacks. A clear benefit of inoculating news consumers against misinformation relative to other misinformation mitigation strategies, such as fact-checking or debunking (Chan et al., 2017), is that inoculation represents a pre-emptive strategy (a form of 'pre-bunking') (Traberg et al., 2023), which is crucial, as the influence of misinformation can persist even after formal retractions or debunking (Johnson & Seifert, 1994; Lewandowsky et al., 2012).

Open Questions and Research Gaps

Despite the promise of inoculation theory in combating misinformation, critical questions remain about its effectiveness within the socially embedded nature of digital misinformation environments. As recent reviews highlight (Traberg et al., 2022, 2023), the majority of inoculation research has been conducted in controlled settings that isolate misinformation from the complex social dynamics typical of media consumption. However, as discussed in the previous section, social media and online platforms have become primary sources of information (Atske, 2021), where individuals rarely encounter misinformation in isolation from social cues. Consequently, despite the theoretical and empirical groundwork laid in the preceding sections on the influence of social and source cues, it remains largely unknown whether traditional inoculation interventions can effectively confer resistance to misinformation in these socially mediated environments (Traberg, 2024; Traberg et al., 2022, 2023). Building on existing inoculation research, this thesis will investigate whether inoculation interventions retain their efficacy in contexts where individuals are simultaneously exposed to social and source cues. Specifically, it will test whether inoculation can cognitively protect individuals from the influence of socially endorsed misinformation or misinformation disseminated by politically aligned news outlets.

To investigate the influence of social context on misinformation susceptibility and its implications for inoculation interventions, this dissertation adopts a social psychological perspective, emphasising the social cognition underlying susceptibility to misinformation. Moving beyond the content of misinformation itself, the primary aim is to explore how the social context in which misinformation is presented shapes its persuasive power. Specifically, this research seeks to examine the interplay between social cues, misinformation susceptibility, and psychological interventions. By embedding social cues within simulated information environments, this dissertation bridges the gap between real-world social media

studies - often limited by uncontrolled variables that obscure causal relationships - and controlled experimental research that primarily examines cognitive mechanisms in isolation from social factors. This approach offers a more comprehensive understanding of how social contexts influence misinformation susceptibility and the potential for targeted interventions to mitigate its effects.

4. Overview of Thesis Contribution

This thesis is structured across 7 chapters. Broadly speaking, this thesis employs experimental paradigms to explore how social and source cues influence misinformation susceptibility and the efficacy of inoculation interventions in social environments. This thesis specifically asks: what are the cognitive processes through which individual misinformation susceptibility is impacted via social and source cues? Under which circumstances do social cues impact individual misinformation susceptibility? When do sources bias news consumer judgements of misinformation? Building on these findings, the thesis further tests the efficacy of prominent psychological interventions rooted in inoculation theory under conditions in which persuasive social and source cues are present.

Part I: The Social Cognition of Misinformation Susceptibility

Part I of this thesis examines the impact of social context on misinformation susceptibility by directly manipulating the source cues and social cues surrounding misinformation headlines. It further seeks to uncover which underlying psychological mechanisms drive the influence of social contextual cues.

Chapter 2 investigates the impacts of source similarity, political slant, and perceived source credibility through two experiments that manipulate the source attribution of true and false headlines using mainstream media outlets differing in political slant. This work is published in *Personality and Individual Differences* (Traberg & van der Linden, 2022). Results show that individuals are more likely to judge misinformation as reliable when it is attributed to political ingroup news sources as opposed to political outgroup news sources. In addition, mediation analyses suggest that the effect of source similarity (ingroup source bias) on misinformation susceptibility is mediated via the credibility of those sources. **Chapter 3** deepens this investigation of sources by directly manipulating the credibility of *fictitious* news sources to uncover the causal mechanisms of source similarity and credibility on the perceived reliability of true and false headlines. This work is published in *Nature Scientific Reports* (Traberg et al., 2024). Here, results show that source credibility influenced the

perceived reliability of misinformation, and political ingroup sources increased susceptibility, but only if the source was high in credibility.

Chapter 4 examines the second contextual cue - the role of explicit and implicit social proof and perceived consensus (in misinformation reliability) in influencing misinformation susceptibility. Here, results show that when explicit or implicit cues signalled endorsement of misinformation, this increased individual misinformation susceptibility. Mediation analyses revealed that perceived public consensus mediated the effect of social cues on misinformation susceptibility.

Part II: Inoculation in Social Contexts

Part II of this thesis explores whether inoculation interventions are robust in settings in which the above-established social cues and source effects impact misinformation susceptibility. **Chapter 5** presents a mixed between-within experiment investigating whether the popular inoculation game – *Bad News* – (Roozenbeek & van der Linden, 2019) can protect news consumers against misinformation even from politically similar sources. Results show that after gameplay, inoculated are significantly less likely to judge misinformation headlines as reliable, and show increased truth discernment (the ability to distinguish between true and fake news correctly). **Chapter 6** presents a 2 (inoculation vs control) x 2 (social cues vs control) experiment investigating the efficacy of a new text-based emotion-fallacy inoculation message. Specifically, it tests whether this intervention protects news consumers against misinformation even when persuasive social cues indicate that others judge it as reliable. Results show that inoculated individuals are significantly less likely to judge misinformation as reliable but that even inoculated individuals are impacted by the influence of social cues.

The studies presented in this thesis aim to challenge and advance the prevailing understanding of misinformation susceptibility by addressing the overlooked role of social and source cues in shaping individual judgements. By integrating insights from social influence, persuasion, and inoculation theory, this thesis questions the assumption that misinformation susceptibility can be understood purely through individual-level processes, advancing a more holistic framework that considers the inherently social nature of digital information environments. Specifically, this research contributes to the growing field of misinformation studies by filling critical gaps in our understanding of how social cues and source factors influence misinformation susceptibility. Moreover, it experimentally tests

whether these socio-cognitive mechanisms impact the effectiveness of interventions, such as inoculation, designed to counter misinformation.

In challenging conventional perspectives on misinformation susceptibility, this thesis underscores the need to rethink existing theoretical frameworks that conceptualise news consumption as an isolated cognitive process. Instead, it underscores how contextual factors may influence perceptions of social consensus and source credibility, shaping judgement formation in social settings. By doing so, it establishes a foundation for developing interventions that are not only theoretically robust but also practically effective in addressing misinformation within social environments. Lastly, the applied nature of this research offers insights for practitioners, including policymakers, educators, and platform designers. By addressing the efficacy, limitations, and future directions of inoculation strategies within social contexts, this thesis aims to inform efforts to build psychological resilience against misinformation, offering tools to tackle this pressing societal challenge.

2. PERCEIVED SOURCE CREDIBILITY AND SIMILARITY

This chapter features Study 3 from the comprehensive paper published in *Nature Scientific Reports* in the article below (in the chapter referred to as Study 1):

Traberg, C. S., Harjani, T., Roozenbeek, J., & van der Linden, S. (2024). The persuasive effects of social cues and source effects on misinformation susceptibility. *Scientific Reports, 14*, 4205. https://doi.org/10.1038/s41598-024-54030-y

It further features the main study published in *Personality and Individual Differences* in the article below (in the chapter referred to as Study 2):

Traberg, C. S., & van der Linden, S. (2022). Birds of a feather are persuaded together: Perceived source credibility mediates the effect of political bias on misinformation susceptibility. *Personality and Individual Differences, 185*, 111269. https://doi.org/10.1016/j.paid.2021.111269

1. Abstract

The viral spread of misinformation poses a threat to societies around the world. Recently, researchers have begun to study how motivated reasoning about news *content* influences misinformation susceptibility. However, the importance of source credibility in the persuasion process is well-documented, and given that source similarity may contribute to credibility evaluations, this raises the question of whether individuals are more susceptible to misinformation from ideologically congruent news sources because they find them more credible. In a large-scale between-subject study (N = 10,539) and a pre-registered online mixed-subject experiment with a US sample (N = 150) using simulated social media posts, we find evidence that both liberals and conservatives judge misinformation to be more accurate when the source is politically congruent and that this effect is mediated by perceived source credibility. We show that source effects play a greater role in veracity judgements for liberals than conservatives but that individuals from both sides of the spectrum judge politically congruent sources as less slanted and more credible. These findings add to the current understanding of source effects in online news environments and provide evidence for the influential effect of source similarity and perceived credibility in misinformation susceptibility.

2. Introduction

The source of a persuasive message plays an important role in determining whether the recipient is persuaded (Pornpitakpan, 2004). Building on prior research in persuasion (Petty & Cacioppo, 1986), social influence (Spears, 2021) and social identity research (Pereira et al., 2021; Van Bavel & Pereira, 2018), this chapter investigates whether individuals are more likely to believe misinformation from political ingroup sources and whether this effect is mediated by perceived source credibility.

In the context of misinformation, some researchers argue that the source of misinformation may not significantly impact belief (Dias et al., 2020), while others suggest that sources play an influential role (Pehlivanoglu et al., 2021). The Identity-based Model of Belief (Van Bavel et al., 2024; Van Bavel & Pereira, 2018) provides an important framework for understanding how group membership shapes individuals' evaluations of information. This model posits that people are more likely to accept information from sources that align with their ingroup identity, as these sources are perceived to share their values and goals. In contrast, information from outgroup sources may be viewed with suspicion, regardless of its accuracy. This chapter empirically explores how political alignment between the source of information and the recipient affects belief in misinformation and whether perceived source credibility mediates this relationship. Specifically, it examines how politically aligned news outlets may influence individuals' perceptions of both misinformation and factual information, with a focus on the role of source bias.

The Role of Source Credibility

There is considerable evidence that source credibility impacts how people assess persuasive messages (Briñol & Petty, 2009; Chaiken & Maheswaran, 1994; Pornpitakpan, 2004), and recent research suggests source credibility also impacts news judgements. For example, Nadarevic et al. (2020) found that source credibility (manipulated by presenting information from either expert or lay sources) had a significant effect on the perceived truth value of news statements on social issues. Furthermore, Bauer & Clemm von Hohenberg (2021) found that participants were more likely to believe and share news from real sources over fake sources, an effect which varied most strongly for different levels of trust in the mainstream media. Source evaluations may also play a role in the sharing of misinformation as Domenico et al. (2021) find that when individuals are exposed to a (fake) source before misinformation by that same source, they are less likely to share it, in part because they perceive the source as having the intent to deceive.

Recent evidence further suggests that individuals may rely on political similarity with a source to evaluate the veracity of information. For instance, a 2017 study investigated the impact of source credibility on the assessment of factual and false statements made by President Trump (Swire et al., 2017). Republican supporters of Trump were more likely to believe information if it was attributed to the President compared to when the same information was presented without source attribution. The opposite pattern held for Democrats, suggesting that individuals use political sources to assess the accuracy of (mis)information. Given this, if a conservatively (liberally) slanted news source publishes misleading information, are conservative (liberal) individuals more likely to be persuaded by it? While previous research has attempted to investigate the role of partisan bias in susceptibility to online misinformation, these studies have largely used content congruence (either Democrat-consistent or Republican-consistent content) rather than source congruence as an operationalisation of partisan bias (Pennycook & Rand, 2019b). Research, therefore, still has to examine the role of political source bias in susceptibility to misinformation.

Although Pennycook & Rand (2019b) suggest that susceptibility to misinformation was driven more by lack of analytical thinking than partisan bias, a critical examination of the key measure, the *media discernment score*, reveals that participants were rating all news with opposing politically slanted headlines as less reliable. Therefore, the *media discernment score* may illusorily indicate that individuals were better at discerning between real and fake news for similar headlines when, in fact, individuals were simply rating *true* news from politically similar sources as significantly more reliable than *true* news from politically dissimilar sources. Indeed, a recent re-analysis of Pennycook and Rand's (2019b) study found that participants were more likely to judge politically incongruent headlines (compared to congruent) as false regardless of headline accuracy (Batailler et al., 2021).

Accordingly, this thesis hypothesises that partisan bias may play a role when news is presented from politically similar news sources. Specifically, it is predicted that (a) individuals will be more likely to judge misinformation from politically similar sources as accurate, and equally, (b) more likely to judge factual information from dissimilar sources as inacurrate.

Ideology and Media Bias

To assess whether the political slant of news sources influences persuasion by misinformation, it is necessary to quantify media bias. While news outlets may claim (or strive) to be politically neutral, research suggests that US news outlets differ ideologically

(Gentzkow & Shapiro, 2010). Research aimed at quantifying media bias has either employed content-based approaches, finding that news sources differ in the way they filter or frame news issues (Budak et al., 2016; Nelson et al., 1997), or audience-based approaches, where researchers have identified differences in the political ideologies of news source audiences (Zhou et al., 2011). Using a content-based approach, Larcinese et al. (2011) found that news outlets with pro-Democratic endorsement patterns gave more coverage to high unemployment when the incumbent president was a Republican compared to when the president was Democratic. Using an audience-based approach, Iyengar and Hahn (2009) found that conservatives and Republicans preferred to read news reports attributed to Fox News and avoided news from CNN, whereas Democrats and liberals turned to CNN but avoided Fox News; a pattern the authors identified for both controversial, but also "soft" subjects including travel.

Although researchers may disagree about the degree of source slant (D'Alessio & Allen, 2000), the evidence points to a divided news audience across partisan lines (Grieco et al., 2018). A cognitive consequence of this may be greater lenience when it comes to misinformation from those preferred sources, potentially due to these sources being perceived as more credible, increasing the likelihood of persuasion. Moreover, recent research has demonstrated that source bias in itself is a unique aspect of source credibility and, indeed, a predictor of persuasion (Wallace et al., 2020). While popular news media may engage in classic editorial practices, they can still provide biased news perspectives, as unintentionally biased sources provide skewed perspectives without underlying intent to deceive (Wallace et al., 2020), making them either more or less credible in the eyes of partisan news consumers. This increased credibility may, in turn, make news consumers more or less likely to believe any information, including misleading information, from sources that have the 'right' kind of bias (Swire-Thompson et al., 2020).

Similarly, partisan biases have been shown to play a role in the tolerance of misinformation spread by politicians: Democrats show a partisan bias in tolerance of lying by commission, whereas Republicans show a partisan bias in tolerance of paltering (De keersmaecker & Roets, 2019). Likewise, political misalignment with a news source may equally lead individuals to be more likely to judge even factual information from those sources as inaccurate. For instance, van der Linden et al. (2020) found that the first media association that comes to mind when hearing the term "fake news" for liberals is Fox News versus CNN for conservatives. Known as the "You are fake news" effect, this may cause a

partisan bias in the assessment of both real and fake news from politically similar versus dissimilar sources.

The Current Studies

This chapter evaluates whether political alignment with news sources impacts participants' ability to evaluate the accuracy of (non-political) news. Specifically, this thesis investigates whether political alignment with news sources increases susceptibility to misinformation and whether political misalignment with news sources decreases people's ability to identify factual information. Study 1 was exploratory, using a between-subject design and a large-scale sample size (N = 10,539) to investigate the impact of news source and source similarity on misinformation susceptibility. Study 2 was confirmatory and pre-registered with a mixed-subject design (N = 150) to further unpack the effects of source similarity, credibility, and slant on misinformation susceptibility.

To investigate this, participants were presented with news headlines in the form of 'tweets' from either conservative (e.g., Fox News) or liberally slanted (e.g., CNN) news sources, where headlines contained factual information or misinformation. In both studies, participants were asked to rate the accuracy of these. In study 2, they were also asked to rate their intention to share the headlines online. As the main research question surrounded perceptions of source slant and credibility, an audience-based approach to quantifying media bias was adopted, categorising the political slant of news sources according to the political leaning of their (main) *audience*.

Study 1

Study 1 sought to uncover whether individuals are more likely to judge non-political misinformation as accurate when the associated news source shares their political slant. Using non-political misinformation was important to isolate the potential effects of partisan source bias. Data was collected via an online platform (N = 10,539, 60% between 18 and 29, 44% female, 49% higher educated, 74% liberal). The source of the misinformation headlines was experimentally manipulated such that it either contained sources (liberal or conservative US news outlets) or a control condition without any source. Following data collection, sources were coded as being similar or dissimilar to the participants' indicated political orientation to enable factorial analyses. Moderates were excluded, as they could not be assigned to a similarity condition.

Study 1 sought to test three exploratory hypotheses. First, the study examined whether the inclusion of news sources would impact judgements of misinformation:

H1: The presence of news sources influences the perceived reliability of misinformation.

Second, it explored the effects of political source similarity on assessments of misinformation, and based on previous literature, the following hypothesis was tested:

H2: Source similarity increases the perceived reliability of misinformation.

Although the sample was unbalanced in political ideology, an exploratory analysis was run to test the hypothesis:

H3: The effect of source similarity on the perceived reliability of misinformation varies depending on political ideology.

Third, while the focus of this thesis is on *mis*information, it is important to assess whether judgements of factual information are subject to the same biases as judgements of misinformation. As such, this study tested the following hypothesis:

H4: Source similarity increases the perceived reliability of factual information.

3. Method

The purpose of Study 1 was to investigate whether source cues impact misinformation susceptibility using a large sample size and a between-subject design. The study was approved by the Cambridge Psychology Research Ethics Committee (PRE.2019.104). All methods were performed in accordance with the relevant guidelines and regulations.

Participants

Data for Study 1 was collected on the *Bad News* platform (Roozenbeek & van der Linden, 2019) as part of a larger intervention experiment on source effects, and consisted of data from the pre-intervention questions. This study relied on press coverage for the *Bad News* game, which allowed anyone with access to the internet to visit the website and participate in the study. As such, this study used a convenience sample. Data collection was set to run for as long as the game platform was available to host the study - between 1st

November 2019 and 15th January 2021. The final sample was $N = 10,539$, 60% between 18-29, 44% female, 49% higher educated, and 74% liberal.

Design and Materials

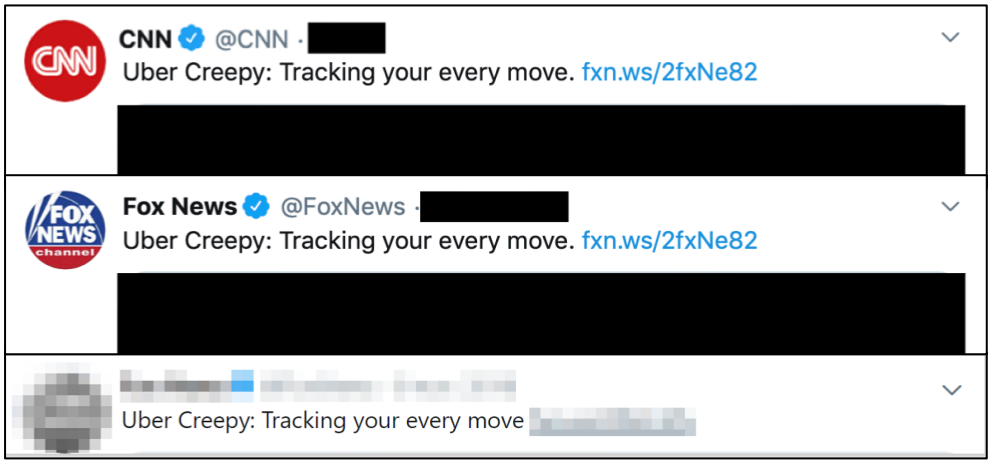
This study employed a between-subjects design and manipulated source slant across three conditions (liberal ($n = 6,184$), conservative ($n = 2,608$), and control ($n = 1,749$)).

Source Manipulation

The news source on Twitter (now known as X) headlines was manipulated across three conditions: Traditionally liberally slanted news sources (CNN, the Washington Post or the New York Times), conservatively slanted news sources (Wall Street Journal, Fox News and Breitbart) or a control condition where the source was blurred out. See Figure 3 for an example of a headline in each condition.

Figure 3

Example of Source Manipulations in Study 1



Note: News headline example in left-wing condition (top), right-wing condition (middle) & control (bottom).

Headlines

To select misinformation news headlines, Hoaxy - a platform developed by researchers at Indiana University that visualises the spread of claims and fact-checking – was used (Shao et al., 2016). A search was carried out to identify misleading headlines based on their use of one of three

misleading strategies used by misinformation producers identified by previous research (Roozenbeek & van der Linden, 2020). These three misinformation tactics included: 1) using exaggeratedly emotional language to distort the news story to generate strong emotional responses, 2) creating or inspiring conspiratorial thinking to rationalise current events, and 3) discrediting otherwise reputable individuals, institutions, or facts to instil doubt in audiences. The non-misleading headlines did not make use of any manipulation techniques and were based on factual news. See Table 1 below for a full overview of the headlines employed.

Table 1

Overview of Items in Study 1

Misinformation vs Facts	**Category**	**Content**	**Source Name**
Misinformation	Conspiracy	Uber Creepy: Tracking your every move	Fox News/CNN/ Control
Misinformation	Conspiracy	Exposing the shadow "elite" controlling the world	Wall Street Journal/The Washington Post/Control
Misinformation	Emotion	Horrific TV show inspiring suicide, says filmmaker	Breitbart News/ New York Times/Control
Misinformation	Emotion	Cop arrested after bragging about video of unprovoked attack on helpless man	Fox News/CNN/Control
Misinformation	Discredit	Expert: Scientific studies no longer trustworthy	Wall Street Journal/The Washington Post/ Control
Misinformation	Discredit	Scandal: MRI Brain imaging completely unreliable	Breitbart News/ New York Times/Control
Factual information	-	Angela Merkel to step down as German chancellor in 2021	CNN/Fox News/Control
Factual information	-	Apple, Google and Amazon named as most valuable brands in the world	The Washington Post/Wall Street Journal/Control
Factual information	-	Physical fitness keeps your brain in good shape	New York Times/Breitbart News/Control
Factual information	-	Olympics: Tokyo finishes building stadium for 2020	CNN/Fox News/Control
Factual information	-	The new Brexit deadline will be January 31	The Washington Post/Wall Street Journal/Control
Factual information	-	Grammy 2020 Awards: Taylor Swift, Lizzo and Beyonce nominated	New York Times/Breitbart News/Control

Measures

Perceived Reliability of Misinformation

To measure misinformation susceptibility, participants were asked to rate each headline's reliability on a standard 7-point Likert scale: *"How reliable is the above news headline"* (1 = Very Unreliable, 7 = Very Reliable). Participants judged the reliability of six misinformation headlines ($M = 3.51$, $SD = 1.48$, $\alpha = 0.80$).

Perceived Reliability of Factual Information

Participants were asked to rate each headline's reliability on a standard 7-point Likert scale: *"How reliable is the above news headline"* (1 = Very Unreliable, 7 = Very Reliable). Participants judged the reliability of six factual headlines ($M = 4.83$, $SD = 1.38$, $\alpha = 0.80$).

Source Similarity

Source similarity was coded according to whether or not the participants' reported political ideology matched the political slant of the source. As political affiliation was originally collected on a 7-point scale, this was converted to a binary (liberal vs conservative) measure. For the analysis, which focused specifically on source similarity, participants who had reported a 'moderate' political stance (a '4' on the 1-7 scale) were excluded from analyses.

Demographics

Age was measured categorically: "18-29" (58.17%), "30-49" (28.72%) and "Over 50" (13.11%). Highest level of completed education was categorised as "High School or Less" (16.86%), "Some College" (31.51%) or "Higher Degree" (51.61%). Gender was categorised as "Female" (44.83%), "Male" (47.90%) and "Other" (7.28%). Finally, political affiliation was measured on a 7-point scale where 1 is very left-wing and 7 is very right-wing ($M = 3.21$, $SD = 1.81$).

Procedure

Participants who visited the website for the *Bad News* game between 1st November 2019 and 15th January 2021 were asked if they would like to take part in a scientific study prior to gameplay. If participants agreed to take part in the study, they were asked to provide informed consent. Following this, they answered socio-demographic questions. As automatic randomisation to conditions was not possible on the game platform, participants were assigned to a condition by selecting a number from 1 to 3, which redirected them to one of the three source conditions. Once they were assigned a condition, they were exposed to news headlines and informed that these were screenshots of real headlines; and asked to report on the above measures (perceived reliability). After completing the study, participants were debriefed regarding the purpose of the study and informed that the headlines were fictitious.

4. Results

H1: The presence of news sources influences the perceived reliability of misinformation.

Study 1 assessed the general influence of the inclusion of sources on misinformation headlines. An independent samples t-test showed that participants who viewed false headlines without sources rated the reliability of these as significantly lower (M = 3.35, SD = 1.65) than those who viewed the headlines with sources (M = 3.57, SD = 1.39, t(10,539) = -6.71, $p <$ 0.001, d = -0.15). As such, *H1* was supported.

H2: Source similarity increases the perceived reliability of misinformation.

A one-way ANOVA was run with similarity as the IV and perceived reliability of misinformation headlines as the DV, showing a statistically significant effect of source similarity on perceived reliability, F(2,7695) = 31.31, $p < 0.001$, $\eta_p^2 = 0.01$. Tukey's post-hoc tests revealed significant differences between when the source was similar (M = 3.65, SD = 1.30) compared to when it was dissimilar (M = 3.38, SD = 1.48, $p < 0.001$, d = 0.19), and between when the source was similar and the control condition (M = 3.40, SD = 1.71, $p <$ 0.001, d = 0.17), but no significant difference between when the source was dissimilar and the control condition (p = 0.83). H2 was thereby supported as participants rated misinformation from politically similar sources as significantly more reliable than misinformation from politically dissimilar sources.

H3: The effect of source similarity on the perceived reliability of misinformation varies depending on political ideology.

To examine whether the effect of source similarity on the perceived reliability of misinformation varied based on political ideology, a multiple regression analysis was conducted. The predictors included source similarity, the continuous political ideology variable (excluding moderates) and their interaction. The dependent variable was the perceived reliability of misinformation headlines. The model revealed a significant interaction between source similarity and political ideology, F (5,7692) = 241.8, $p < 0.001$, with an adjusted R^2 = 0.14. The analysis showed that the interaction between source similarity and political ideology was significant and negative when comparing the "similar" condition to the "dissimilar" condition ($\beta = -0.401$, $p < 0.001$). This indicates that as participants became more conservative (higher values on the political ideology scale), the effect of source similarity on perceived reliability decreased. For the control condition (where no sources

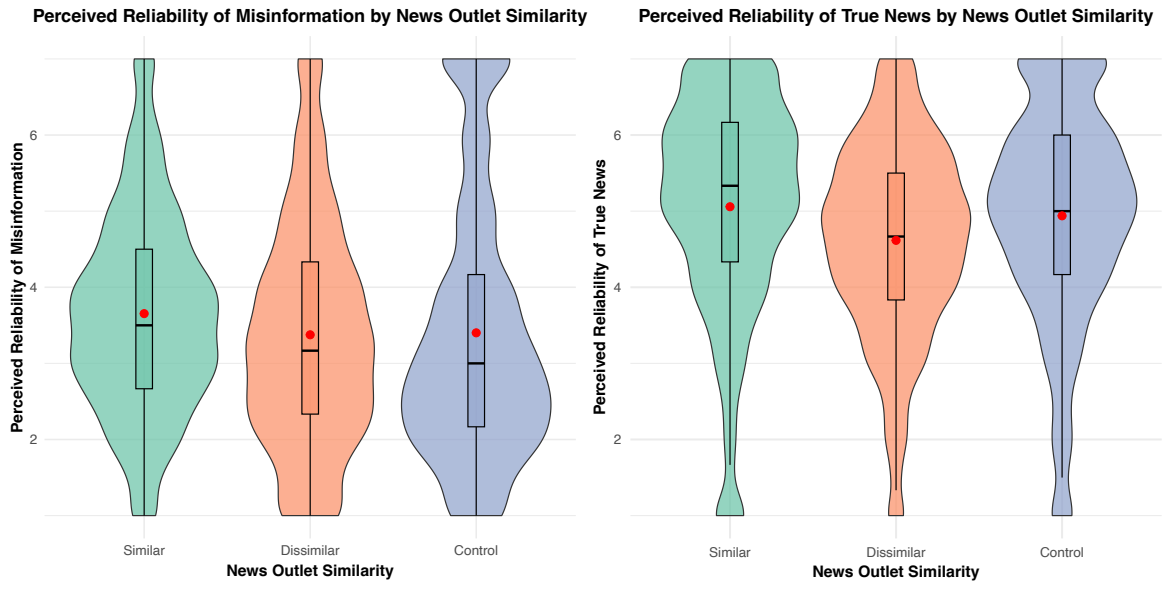
were included), the interaction between political ideology and perceived reliability was significant and positive ($\beta = 0.070$, $p < 0.001$). This suggests that conservatives were slightly more likely than liberals to perceive misinformation as reliable in the absence of source information. These findings suggest that the influence of source similarity on the perceived reliability of misinformation may vary by political ideology. Specifically, liberals were more strongly influenced by political similarity, whereas conservatives showed less variation in their judgements of misinformation across source conditions. However, due to the unequal sample sizes, this result is interpreted with caution.

H4: Source similarity increases the perceived reliability of factual information.

A one-way ANOVA with source similarity as the IV and perceived reliability of factual headlines as the DV was found significant ($F(2,7695) = 80.12$, $p < 0.001$, $\eta_p^2 = 0.02$), with post-hoc tests revealing that individuals judged factual headlines as more reliable when they were attributed to a politically similar source compared to dissimilar ($M_{diff} = -0.44$, $p < 0.001$, $d = -0.32$). Participants judged factual headlines as more reliable in the similar condition compared to the control condition ($M_{diff} = 0.12$, $p = 0.01$, $d = 0.09$), and as more reliable in the control compared to the dissimilar condition ($M_{diff} = 0.32$, $p = 0.01$, $d = 0.24$). The results from H2 and H4 are visualised in Figure 3 below.

Figure 4

Perceived Reliability of (Left) Misinformation and (Right) True News by News Outlet Similarity



5. Discussion

This study provides several important insights into the role of source presence and source similarity in shaping perceptions of misinformation reliability. First, the inclusion of a source significantly increased the perceived reliability of misinformation headlines. Participants rated misinformation as more reliable when it was accompanied by a media source, suggesting that mere source attribution may lend an air of legitimacy to information, regardless of its accuracy. This aligns with prior research, which indicates that audiences often interpret source information as a heuristic cue for credibility, particularly in settings where deeper cognitive processing is limited (Metzger, Flanagin & Medders, 2010; Petty & Cacioppo, 1968; Chaiken, 1980; Pornpitakpan, 2004).

Second, this study provides some initial evidence for the influence of source similarity on perceptions of misinformation reliability. When participants encountered misinformation attributed to a politically similar source, they judged it as more reliable compared to misinformation from a politically dissimilar source or when no source was provided. This suggests that ingroup affiliation may enhance reliability perceptions of misinformation to a greater extent than a lack of source or even an outgroup source diminishes it, which is in line with the Identity-based Model of Belief. This pattern further aligns with Tajfel and Turner's (1979) Social Identity Theory, which posits that people exhibit favourability toward information from ingroup sources due to identity reinforcement. Additionally, Hogg & Terry (2000) and Mackie et al (1992) discuss how social identity impacts information processing, indicating that individuals may default to scepticism toward information from outsiders or unsourced information.

For factual information, a similar pattern emerged. Participants rated factual headlines as more reliable when attributed to a politically similar source, with reliability ratings decreasing when the source was politically dissimilar. This finding suggests that political similarity influences the perceived reliability of both misinformation and factual information, underscoring the role of social identity and in-group favouritism in information processing. This tendency to evaluate even factual content through a partisan lens suggests that source similarity serves as a powerful cue, potentially overriding the objective accuracy of information. These findings contribute to our understanding of how source characteristics influence judgements about information reliability.

6. Limitations

While this study sheds light on the influence of source presence and similarity, it has several limitations which warrant discussion. First, the sample was highly skewed towards liberals (74%), and an issue with the recruitment platform meant that the sample size was unbalanced, such that the final sample had very few conservatives who saw in-group misinformation (3.2% or 236 out of 7471). For this reason, the exploratory analysis of political ideology needs to be interpreted with great caution. This also severely limits the generalisability of the findings. Second, the study's participants were not exclusively from the U.S., yet the sources used in the headlines were U.S.-based. This mismatch may have affected how non-US participants perceived the sources, potentially impacting the study's findings on source similarity. Finally, this study did not explore the underlying mechanisms driving the observed effects of source similarity. While the findings indicate that individuals are more susceptible to misinformation from ingroup sources, it remains unclear whether this effect is primarily driven by perceived credibility. Eagly et al. (1978) emphasise that perceived credibility often hinges on source similarity, which may underlie this effect. Study 2 will address this gap by investigating these underlying mechanisms and exploring how credibility assessments contribute to the influence of ingroup sources on misinformation susceptibility.

Study 2

Study 2 builds on findings in Study 1 and evaluates whether political alignment with news sources interferes with participants' ability to evaluate the accuracy of (non-political) headlines. The design and main analyses in this study were pre-registered. The pre-registration is available online (https://osf.io/4suyb?view_only=e048edbe34244df58ca2d84ec1c80f4d).

As in Study 1, we ask whether political alignment with news sources increases susceptibility to misinformation and, equally, whether political misalignment with news sources decreases people's ability to identify factual information. To investigate this, a social media feed was simulated in which participants were presented with news headlines in the form of 'tweets' from either conservative (e.g. Fox News) or liberally slanted (e.g. CNN) news sources, where headlines contained either factual information or misinformation. Participants were asked to rate the accuracy of headlines, as well as their intention to share them online. As the main research question surrounded perceptions of source bias and

credibility, an audience-based approach to quantifying media bias was adopted, categorising the political slant of news sources according to the political leaning of their (main) audience.

Based on results from a pilot study showing a significant effect of source ideology on misinformation susceptibility (see Appendix A), we pre-registered testing the effect of the main manipulation variable (source ideology) on several dependent variables. Although results from the pilot revealed that news from liberal sources was judged as more reliable overall, the sample was also skewed towards liberals (for further information about the pilot, see Appendix A), and as such, we did not necessarily expect to replicate the same result here. Accordingly, the first hypothesis was non-directional:

H1. There will be a main effect of source ideology on (a) accuracy judgements of misinformation, (b) accuracy judgements of factual information, (c) sharing of misinformation, and (d) sharing of factual information.

As previous research has found that individuals can, at least to some extent, distinguish between misinformation and factual information overall (Basol et al., 2020; Roozenbeek & van der Linden, 2019) we test the following hypothesis:

H2. There will be a difference between the perceived accuracy of misinformation and factual information regardless of source.

Because individuals tend to have a 'bias blind spot', that is, being unable to recognise (or at least underestimate) the influence of biases on their own judgements (Pronin, 2007), we also test whether liberals and conservatives differ in their judgements of the political slant of news sources. We expect that both groups will perceive politically similar sources as having little to no political slant while perceiving politically dissimilar sources as having a significant political slant. Accordingly, we test the following hypothesis:

H3. There will be a significant difference between the perceived political slant of liberal vs conservative sources such that (a) liberal participants are more likely to judge conservative sources as more extreme than conservatives and (b) vice versa.

Similarly, we hypothesise that political ideology impacts judgements of (mis)information, such that individuals will judge both misinformation and factual information from sources with opposite political slants as less accurate.

H4. Political ideology will moderate the effect of source ideology on accuracy judgements such that individuals will be more likely to judge both (a) misinformation and (b) factual information from sources with opposite political slants as inaccurate.

Furthermore, we predict that credibility judgements of news sources will depend on the individual's political ideology. Specifically, we test the following hypotheses:

H5. Political ideology of participants predicts their perceived source credibility such that (a) liberal participants will judge conservative sources as being less credible than conservative participants and (b) vice versa.

Moreover, based on research in the attitudes field(Briñol & Petty, 2009; Kumkale et al., 2010; Metzger et al., 2010), we test whether these effects are mediated by the perceived credibility of politically (dis)similar sources:

H6. Source credibility will mediate the effect of political ideology on accuracy judgements overall.

7. Method

Design

This study's design and main analyses were pre-registered (https://osf.io/4suyb?view_only=e048edbe34244df58ca2d84ec1c80f4d). Exploratory analyses were not pre-registered and are labelled accordingly. This study employed a 2 × 2 mixed-subjects design (N = 150) with the independent variables being the political ideology of the participants (Liberal n = 76, Conservative n = 74) and the within-subject factor being the political slant of source (Liberal vs Conservative). Dependent variables included: Accuracy judgements of 6 misleading headlines and 6 factual headlines following a similar format to Roozenbeek and van der Linden (2019) (see Table 2 for full list of headlines), perceived source credibility, and perceived slant of the six sources present in the headlines. A power analysis for the main

experiment was conducted with G*power using $\alpha = 0.05$, $dz = 0.3$ and power of 0.95, indicating that a total sample size of 147 participants would be required to detect the main effect in a mixed-subject experiment. The full pilot study results and discussion can be found in Appendix A.

Participants

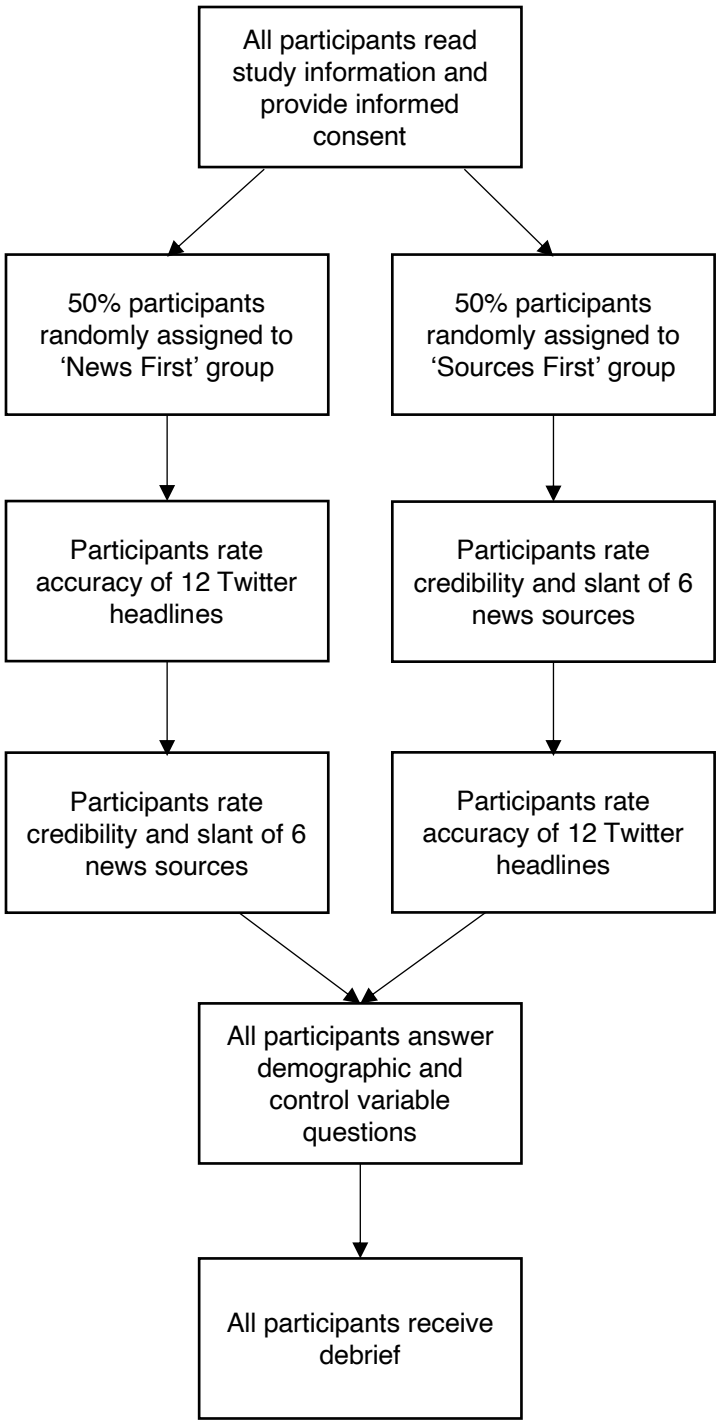
A total of 156 participants were recruited through the crowdsourcing platform Prolific. As opposed to the previous study, only U.S. participants were recruited to ensure familiarity with U.S. news outlets, which were part of the manipulation. As a criterion for participation on Prolific was identifying as either liberal or conservative, 6 participants who identified as moderates were excluded (in line with the pre-registration). The final sample consisted of 150 participants (53% female, age range 19–80, $M_{age} = 32$, 51% liberals, 55% higher educated).

Procedure

Upon providing informed consent, participants proceeded to the study. To counterbalance for order effects, 50% of participants rated the credibility and slant of sources prior to being exposed to the headlines. The remaining participants were exposed to headlines prior to completing credibility and slant ratings. Participants were randomly presented with 12 fictitious headlines which contained either misinformation (6) or factual information (6) and were depicted as being posted by liberal news outlets (6) or conservative news outlets (6). For each headline, participants were asked to report how accurate they perceived each headline to be as well as how likely they would be to share them on social media. Following the completion of the outcome measures, all participants completed a series of demographic and control measures. A flowchart of the study procedure is provided in Figure 5 below. This study was approved by the Cambridge Psychology Research Ethics Committee (PRE.2019.104).

Figure 5

Flowchart Detailing Study Procedure



Materials

Fake vs Real Headlines

Identified through the independent fact-checking and misinformation visualisation platform "Hoaxy" (Shao et al., 2016) six misinformation headlines were selected. To the extent possible, all headlines were non-political; that is, they did not discuss or refer to politics, nor did they contain information or news about topics shown to divide partisans (such as, but not limited to, immigration, crime rates, taxes, elections, religion, climate change, gun policy, death penalty, abortion, etc.). All headlines were chosen based on three common and key deceptive strategies identified by Roozenbeek and van der Linden (2019) (spreading conspiracy theories, using emotional language and discrediting opponents), and formatted to resemble Twitter (X) posts to enhance ecological validity. To ensure images or dates would not influence participants, these aspects were blacked out. An example headline classified as misinformation, specifically spreading conspiracy theories, was: "*Exposing the shadow 'elite' controlling the world*" (see Figure 6, top). Criteria for factual headlines were that the information presented could be confirmed by multiple news outlets, was not subjectively interpretable, and did not use any misinformation strategies (Roozenbeek et al., 2020; Roozenbeek & van der Linden, 2019). An example headline classified as factual information was: "*Angela Merkel to step down as German chancellor in 2021*" (see Figure 6, bottom).

Figure 6

News Items Included Containing (Top) Misinformation and (Bottom) Factual Information

Source Slant

Source slant was manipulated by altering the source of the headlines. To categorise news media slant, we adopted an audience-based approach, as our main research questions pertained to perceptions of sources and not the objective content or coverage bias displayed by news media. Three sources shown to be trusted by more Republicans than distrusted were selected: Fox News, the Wall Street Journal, and Breitbart (Jurkowitz et al., 2020) and three sources shown to be trusted by more Democrats than distrusted were selected: CNN, the New York Times, and the Washington Post (Jurkowitz et al., 2020). This categorisation is also in line with a previous crowdsourced ideological ordering of US news outlets (Budak et al., 2016). Each headline was manipulated to resemble a post from one of the liberal/conservative sources. A total of 12 items were created, with 6 from the identified conservative sources and 6 from the identified liberal sources.

The final set of items consisted of 12 headlines. These included 3 misinformation headlines from liberal sources, 3 misinformation headlines from conservative sources, 3 factual news headlines from liberal sources and 3 factual news headlines from conservative sources (see Table 2 for all headlines). See Figure 7 (misinformation) and Figure 8 (factual information) for all headlines in Twitter format.

Table 2

Misinformation/Factual Headlines in Study 2

Misinformation/Factual	Source Slant	Source Name in Study	Content	Original Source
Misinformation	Right-wing	Fox News	Uber Creepy: Tracking your every move	InfoWars
Misinformation	Left-wing	The Washington Post	Exposing the shadow "elite" controlling the world	Humans Are Free
Misinformation	Right-wing	Breitbart News	Horrific TV show inspiring suicide, says filmmaker	WorldNetDaily
Misinformation	Left-wing	CNN	Cop arrested after bragging about video of unprovoked attack on helpless man	The Free Thought Project
Misinformation	Right-wing	The Wall Street Journal	Expert: Scientific studies no longer trustworthy	WorldNetDaily
Misinformation	Left-wing	The New York Times	Scandal: MRI Brain Imaging completely unreliable	InfoWars
Factual	Right-wing	Fox News	Angela Merkel to step down as German chancellor in 2021	BBC
Factual	Left-wing	The Washington Post	Apple, Google and Amazon named as most valuable brands in the world	The Guardian
Factual	Right-wing	Breitbart News	Physical fitness keeps your brain in good shape	Psychology Today

Factual	Left-wing	CNN	Olympics: Tokyo finishes building stadium for 2020	Reuters
Factual	Right-wing	The Wall Street Journal	The new Brexit deadline will be January 31	Vox
Factual	Left-wing	The New York Times	Grammy 2020 Awards: Taylor Swift, Lizzo and Beyonce nominated	The Daily Mail

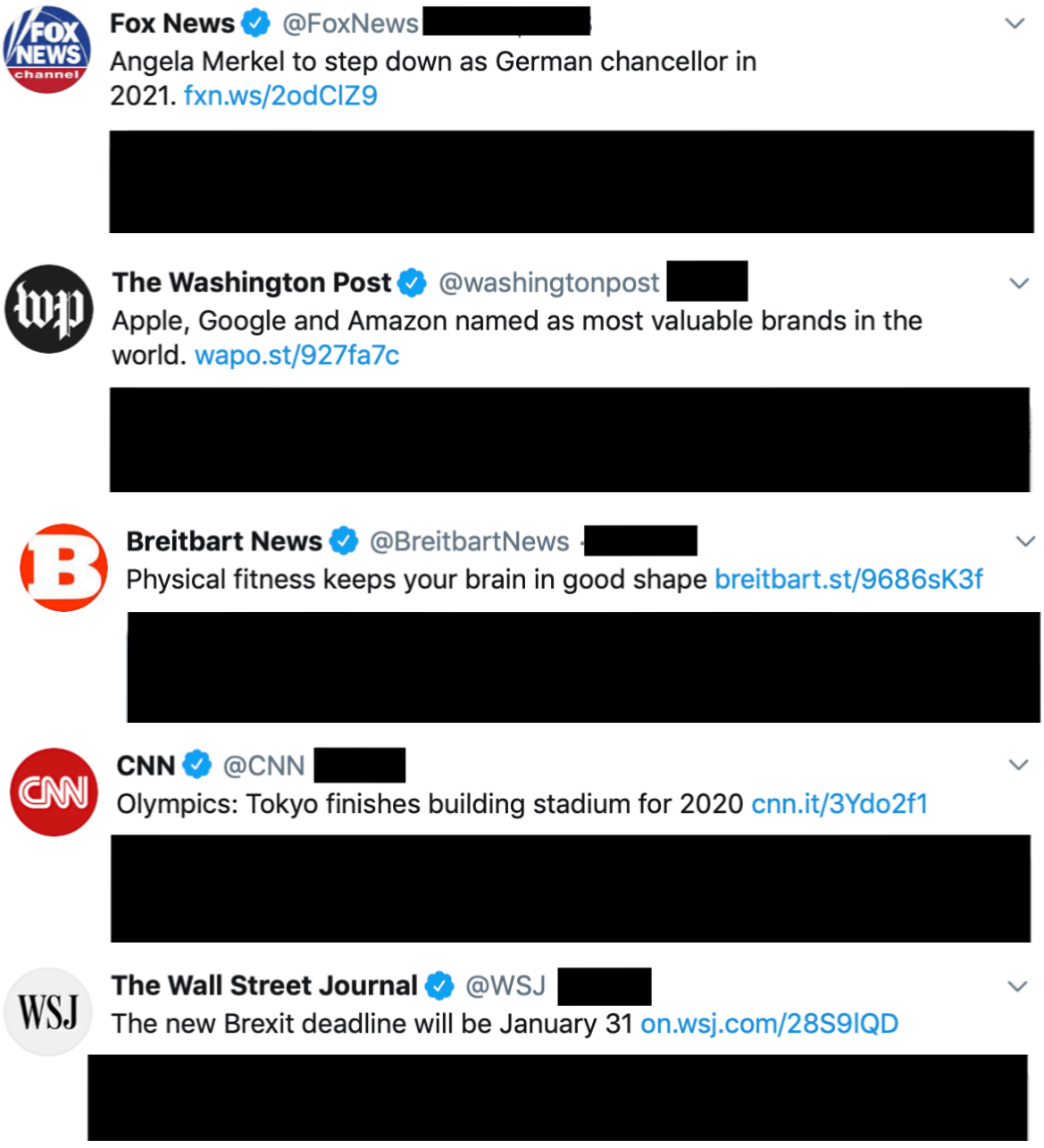
Figure 7

Misinformation Headlines in Twitter Format

Note. Headlines categories from the top: Right-wing, right-wing, right-wing, left-wing, left-wing, left-wing.

Figure 8

Factual Headlines in Twitter Format



Note. Headlines categories from the top: Right-wing, right-wing, right-wing, left-wing, left-wing, left-wing.

Measures

Perceived Accuracy of Misinformation

To assess perceived accuracy of misinformation, a single-item measure was presented alongside each of the Twitter headlines. Participants were asked to report on the perceived accuracy of the tweet on a 7-point Likert scale from Very Accurate (1) to Very Inaccurate (7). Participants judged the accuracy of six misinformation headlines ($M = 3.68$, $SD = 0.91$, $\alpha = 0.53$).

Perceived Accuracy of Factual Information

To assess the perceived accuracy of factual information, a single-item measure was presented alongside each of the Twitter headlines. Participants were asked to report on the perceived accuracy of the tweet on a 7-point Likert scale from Very Accurate (1) to Very Inaccurate (7). Participants judged the accuracy of six factual headlines ($M = 5.62$, $SD = 0.87$, $\alpha = 0.73$).

Intent to Share Misinformation

Similar to perceived accuracy, a single-item intent-to-share measure was presented alongside the Twitter posts, where participants reported how likely they would be to share the headline with their network if it came up on their news feed. This was reported on a 7-point Likert scale from Very Unlikely (1) to Very Likely (7). Participants rated how likely they would be to share six misinformation headlines ($M = 2.30$, $SD = 1.27$, $\alpha = 0.84$).

Intent to Share Factual Information

Similar to perceived accuracy, a single-item intent-to-share measure was presented alongside the Twitter posts, where participants reported how likely they would be to share the headline with their network if it came up on their news feed. This was reported on a 7-point Likert scale from Very Unlikely (1) to Very Likely (7). Participants rated how likely they would be to share six factual information headlines ($M = 2.28$, $SD = 1.33$, $\alpha = 0.87$).

Source Credibility and Slant

Participants were asked to report on the perceived source credibility of the three liberal sources (New York Times, Washington Post, CNN) (M = 4.18, SD = 1.75) and the three conservative sources (Fox News, Breitbart News, Wall Street Journal) (M = 3.64, SD = 1.23) on a 7-point Likert scale from Not Credible (1) to Credible (7). Similarly, participants were asked to report on the perceived slant of the three liberal sources (M = 2.91, SD = 1.29) and the three conservative sources (M = 4.90, SD = 1.19) on a 7-point Likert scale from Very Liberal (1) to Very Conservative (7).

Political ideology

Political ideology was assessed using a self-placement scale ranging from Very Liberal (1) to Very Conservative (7) (M = 3.89, SD = 2.14). As outlined in the pre-registration, for certain analyses, political ideology was recoded as a categorical variable, where responses 1–3 were collapsed into a single liberal category (n = 76), and responses 5–7 were collapsed into a single conservative category (n = 74).

Demographic Variables

As previous research has shown a significant correlation between self-reported sharing of misinformation and actual sharing of misinformation on Twitter / X (Mosleh et al., 2020), Twitter use was assessed as a potential predictor of the main DV. Participants were asked to indicate their current level of Twitter usage, and responses were recorded across the following categories: I have an account, but I hardly ever use it (27.3%); I have an account and I use it occasionally (31.3%); I have an account and I use it often (7.3%); I have an account and I use it on a daily basis (13.3%) and; I don't have an account (20.8%). Primary news source was also included as a demographic variable, for which participants reported the source they get the majority of their news from. Responses were recorded across the following categories: I don't really follow the news (4.0%); Social media (28.7%); TV & radio (22.7%); Print media (newspapers, magazines) (2.0%); Word of mouth (0.7%); Online media (excluding social media) (42.9%).

8. Results

News Source Impacts Accuracy Judgements

H1. There will be a main effect of source ideology on (a) accuracy judgement of misinformation, (b) factual information and (c) sharing of misinformation and (d) factual information.

As pre-registered to test H1a, a paired samples *t*-test showed a main effect of source ideology on accuracy judgements of misinformation, with misinformation from liberal sources being judged as significantly more accurate ($M = 3.81$, $SD = 1.12$) than misinformation from conservative sources ($M = 3.54$, $SD = 1.15$), $t(149) = 2.42$, $p = 0.02$, $d = 0.24$, $M_{diff} = 0.27$, 95% CI [0.05, 0.48]. Similarly, a paired samples *t*-test showed a main effect of source ideology on accuracy judgements of factual information in support of H1b, with liberal factual information being rated as significantly more accurate ($M = 5.88$, $SD = 0.96$) than conservative factual information ($M = 5.36$, $SD = 1.06$), $t(149) = 6.39$, $p < 0.001$, $d = 0.53$, $M_{diff} = 0.53$, 95% CI [0.37, 0.69]. Furthermore, H1c was supported as a paired samples *t*-test showed that intent to share misinformation from liberal sources was higher ($M = 2.40$, $SD = 1.42$) than intent to share misinformation from conservative sources ($M = 2.19$, $SD = 1.4$), $t(149) = 2.0917$, $p = 0.04$, $d = 0.15$, $M_{diff} = 0.20$, 95% CI [0.01, 0.4]. Finally, a paired samples *t*-test showed that intent to share factual information from liberal sources ($M = 2.29$, $SD = 1.44$) was not significantly different than intent to share information from conservative sources ($M = 2.27$, $SD = 1.41$), $t(149) = 0.21$, $p = 0.83$, $d = 0.01$, $M_{diff} = 0.02$, 95% CI [− 0.15, 0.18] and as such, H1d was not supported.

Participants Distinguish Between Fact and Misinformation Overall

H2. There will be a difference between the perceived accuracy of misinformation and factual information.

A paired samples *t*-test showed that participants were able to distinguish between factual information and misinformation overall, and in support of H2, perceived accuracy of misinformation was significantly lower ($M = 3.68$, $SD = 1.14$) than perceived accuracy of factual information ($M = 5.62$, $SD = 1.04$), $t(299) = -26.659$, $p < 0001$, $d = 1.78$, $M_{diff} = -1.90$, 95% CI [− 2.1, − 1.8]. Although not pre-registered, a post-hoc independent samples *t*-test found no difference between the ability of liberals ($M = 3.72$, $SD = 1.2$) and conservatives ($M = 3.63$, $SD = 1.07$) to identify misinformation as inaccurate overall, $t(298) =$

− 0.71, $p = 0.48$, $d = -0.081$, 95% CI [− 0.35,0.17]. Similarly, an independent samples *t*-test found no difference between the ability of liberals ($M = 5.65$, $SD = 1.01$) and conservatives ($M = 5.59$, $SD = 1.08$) to identify facts as accurate, $t(298) = -0.49$, $p = 0.6$, $d = -0.06$, 95% CI [− 0.3,0.18].

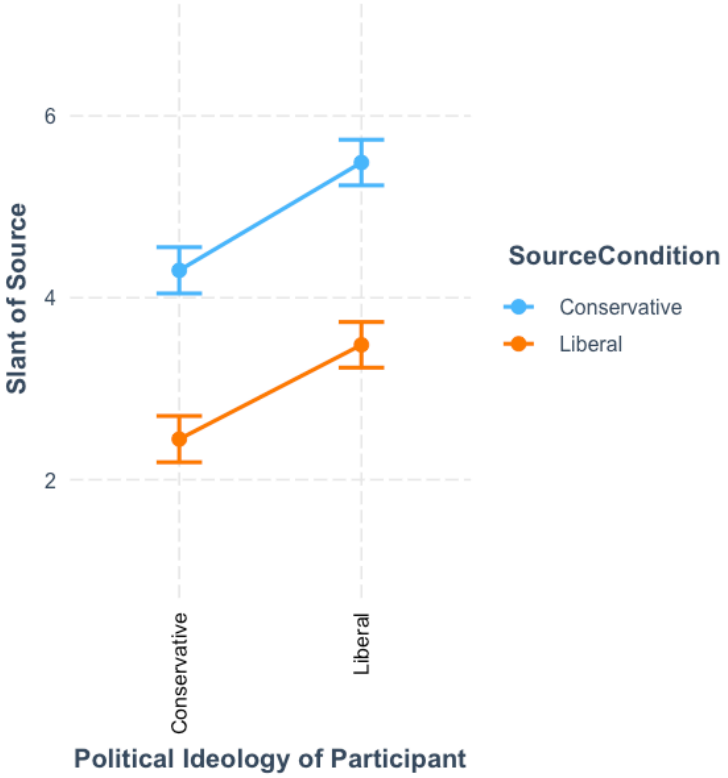
Participants Perceive Politically Similar Sources as Less Slanted

H3. There will be a significant difference between the perceived political slant of liberal vs conservative sources such that (a) liberal participants are more likely to judge conservative sources as more extreme than conservatives (a) and (b) vice versa.

Supporting H3a, an independent samples *t*-test showed that liberal participants judged conservative sources to be more politically slanted ($M = 5.49$, $SD = 0.93$) than conservative participants ($M = 4.3$, $SD = 1.13$), $t(148) = 7.03$, $d = 1.15$, $p < 0.001$. Supporting H3b, an independent samples *t*-test showed the reverse effect was seen for liberal sources, where these were rated as significantly more politically slanted by conservative participants ($M = 2.45$, $SD = 1.28$) than by liberal participants ($M = 3.48$, $SD = 1.08$), $t(148) = 5.36$ $p < 0.001$, $d = 0.87$. Results for H3 are visualised in Figure 9.

Figure 9

Slant of Source by Participants' Political Ideology and News Source Ideology



Note: Participants reported perceived slant of sources on a 7-point Likert scale ranging from (1) very liberal to (7) very conservative. Error bars represent 95% confidence intervals.

Political Ideology Moderates the Source Effect

H4a. Political ideology will moderate the effect of source ideology on accuracy judgements of misinformation such that individuals identifying as conservative will be more likely (than liberals) to judge misinformation from liberal sources as being inaccurate, and individuals identifying as liberals will be more likely (than conservatives) to judge misinformation from conservative sources as being inaccurate.

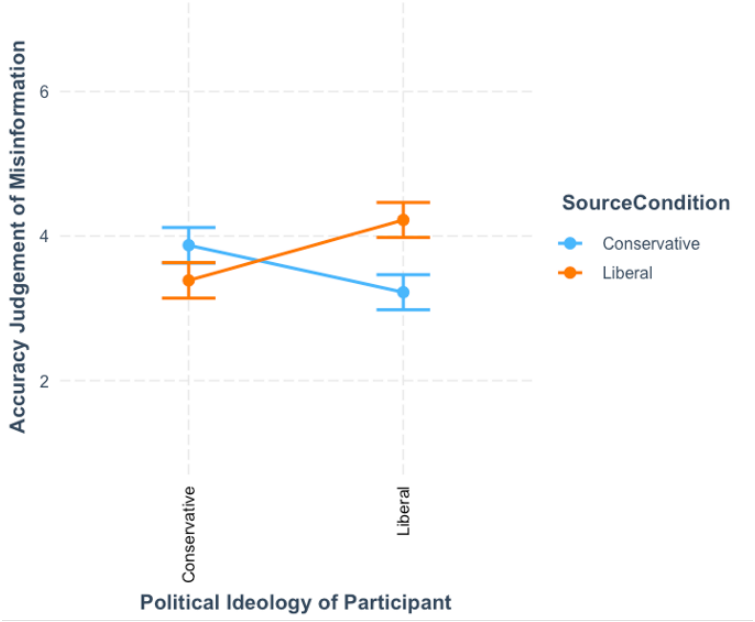
In accordance with the pre-registration and to test H4a, a factorial ANOVA was run with accuracy judgements of misinformation as DV, and source ideology and political ideology as IVs, yielding a main effect for source ideology, $F(1,296) = 4.62$, $p = 0.03$, $\eta_p^2 = 0.02$ and no main effect of political ideology, $F(1,296) = 0.56$, $p = 0.45$, $\eta_p^2 = 0.002$. The interaction effect between source ideology and political ideology was significant, $F(1,296) = 35.86$, $p < 0.001$, $\eta_p^2 = 0.11$, and a simple slopes analysis further revealed that political

ideology was significant for both levels of source: Perceived accuracy of misinformation from conservative sources was significantly higher for conservative participants (M = 3.87, 95% CI [3.63,4.12]) than for liberal participants, (M = 3.22, 95% CI [2.98,3.47]), $p < 0.001$, d = 0.61, whereas perceived accuracy of misinformation from liberal sources was significantly higher for liberal participants (M = 4.22, 95% CI [3.98,4.47]) than for conservative participants, (M = 3.39, 95% CI [3.14,3.63]), $p < 0.0001$, d = 0.78. Among conservatives, source had a significant impact on perceived accuracy, with conservative participants judging misinformation from conservative sources as more accurate (M = 3.87, 95% CI [3.63,4.12]) than misinformation from liberal sources (M = 3.39, 95% CI [3.14,3.63]), p = 0.006, d = 0.32, where the opposite effect was seen for liberals participants, who judged misinformation from liberal sources to be significantly more accurate (M = 4.22, 95% CI [3.98,4.47]) than that from conservative sources (M = 3.22, 95% CI [2.98,3.47]), $p < 0.001$, d = 0.66.

A post-hoc ANOVA (which was not preregistered) with political ideology and political similarity between source and participant as IVs and perceived accuracy of misinformation as DVs found that liberal judgements of liberal misinformation were significantly higher (M = 4.22, 95% CI [3.98, 4.47]) than conservative judgements of conservative misinformation (M = 3.87, SE = 0.125, 95% CI [3.63, 4.12], d = 0.33, p = 0.047), suggesting liberals were slightly more ‘biased’ in their judgements of misinformation from politically congruent sources than conservatives. Results for H4a can be seen in Figure 10.

Figure 10

Perceived Accuracy of Misinformation by Participants' Political Ideology and Source Ideology



Note: Participants reported accuracy judgements on a 7-point Likert scale ranging from very inaccurate (1) to very accurate (7). Error bars represent 95% confidence intervals.

H4b. Political ideology will moderate the effect of source ideology on accuracy judgements of factual information such that individuals identifying as conservative will be more likely (than liberals) to judge factual information from liberal sources as being inaccurate and individuals identifying as liberal will be more likely (than conservatives) to judge factual information from conservative sources as inaccurate.

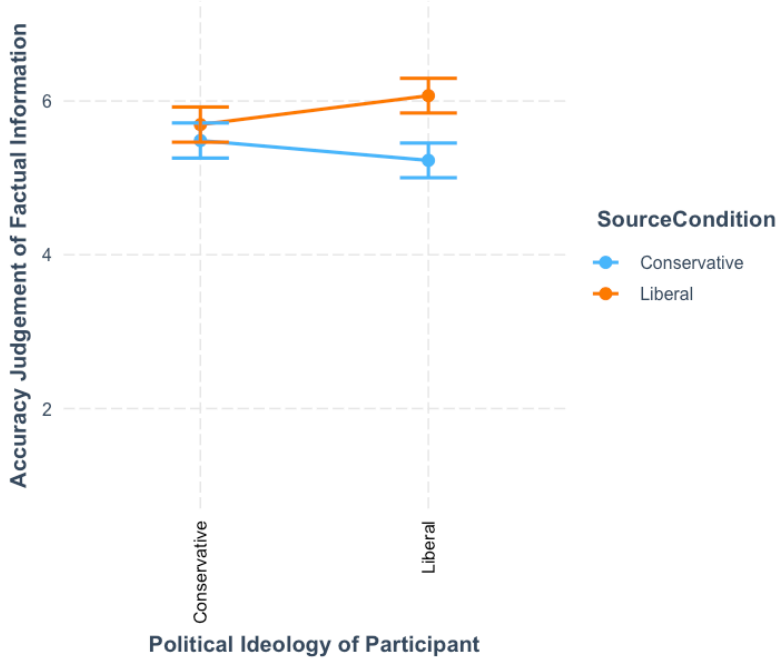
In adherence with the pre-registration and to test H4b, a factorial ANOVA was run with perceived accuracy of factual information as DV, and source ideology and political ideology as IVs, yielding a main effect for source ideology, $F(1,296) = 20.99$, $p < 0.001$, $\eta_p^2 = 0.07$, but no main effect of political ideology, $F(1,296) = 0.21$, $p = 0.61$, $\eta_p^2 < 0.01$. However, the interaction effect between source ideology and political ideology was significant, $F(1,296) = 7.56$, $p < 0.01$, $\eta_p^2 = 0.02$. As such, a simple slopes analysis was run, finding that when the source was conservative, political ideology was not significant: That is, conservative ($M = 5.49$, 95% CI [5.26,5.72]) and liberal participants' ($M = 5.23$, 95% CI [5,5.45]) perceived accuracy of factual information presented by conservative sources was not significantly different, $p = 0.11$, $d = 0.26$. However, when the source presenting factual information was liberal, liberal participants' perceived accuracy was significantly higher ($M =$

6.07, 95% CI [5.84,6.3]) than conservative participants' perceived accuracy (M = 5.69, 95% CI [5.46,5.92]), p = 0.02, d = 0.38. Among conservatives, source did not have a significant impact on the perceived accuracy of factual information as they did not differ in their judgements of factual information from liberal (M = 5.69, 95% CI [5.46,5.92]) or conservative sources (M = 5.49, SE, 95% CI [5.26,5.72]), p = 0.21, d = 0.15. In contrast, for liberal participants, the source had a significant impact on the perceived accuracy of factual information, as they judged factual headlines to be significantly more accurate when they came from liberal sources (M = 6.07, 95% CI [5.84,6.3]) than from conservative sources (M = 5.23, 95% CI [5,5.45]), $p < 0.001$, d = 0.6. These results indicate an asymmetry between liberal and conservative participants in that liberal participants were significantly worse at identifying factual information as accurate when the news source was conservative compared to when the news source was liberal. This was not the case for conservatives, whose perceived accuracy of factual news did not depend on the slant of the associated source. Results for H4b can be seen in Figure 11.

Again, although not pre-registered, a post-hoc ANOVA was run to examine differences in the perceived accuracy of factual information based on political ideology and political source similarity. Results showed that when sources were politically different from participants' ideology, liberal participants' perceived accuracy of facts was significantly lower (M = 5.23, 95% CI [5,5.45]) than conservative participants' (M = 5.69, 95% CI [5.46,5.92]), $p < 0.01$, $d = -0.47$.

Figure 11

Accuracy Judgement of Factual Information by Participants' Political Ideology and Source Ideology



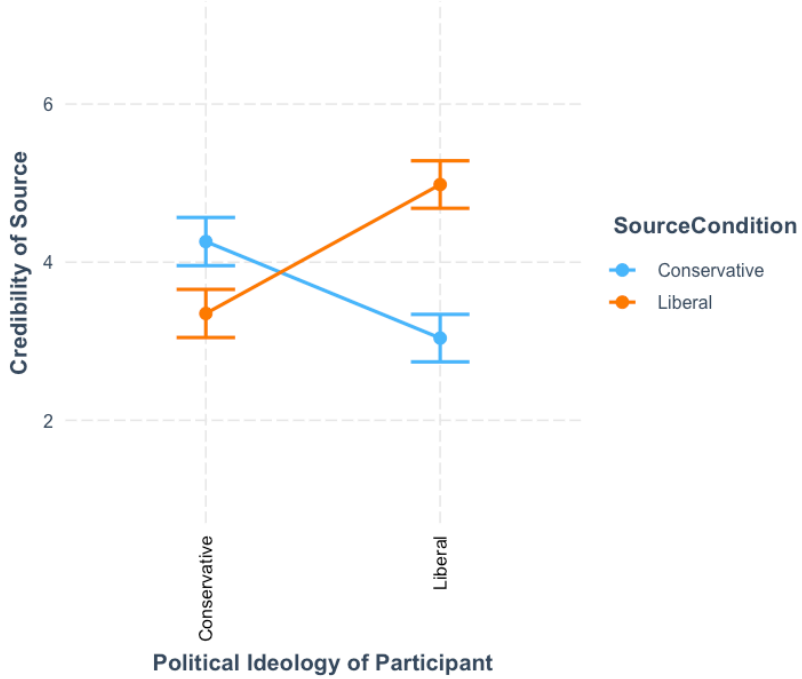
Note: Participants reported accuracy judgements on a 7-point Likert scale ranging from very inaccurate (1) to very accurate (7). Error bars represent 95% confidence intervals.

H5. Political ideology of participants predicts their perceived source credibility such that liberal participants will judge conservative sources as being less credible than conservative participants (a) and vice versa (b).

An independent samples *t*-test supported H5a, with results showing that conservative sources were rated as significantly less credible by liberal participants ($M = 3.04$, $SD = 1.08$) than by conservative participants ($M = 4.26$, $SD = 1.07$), $t(148) = 6.98$, $p < 0.001$, 95% CI [0.88, 1.57], $d = 1.14$. Similarly, an independent samples *t*-test supported H5b, with results showing that liberal sources were rated as significantly less credible by conservative participants ($M = 3.35$, $SD = 1.67$) than by liberal participants ($M = 4.98$, $SD = 1.42$), $t(148) = -6.45$, $p < 0.001$, 95% CI [− 2.13, − 1.13], $d = -1.05$. Results for H5 are illustrated in Figure 12.

Figure 12

Perceived Source Credibility by Participants' Political Ideology and Source Ideology



Note: Participants reported credibility of sources on a 7-point Likert scale ranging from not credible (1) to credible (7). Error bars represent 95% confidence intervals.

Source Credibility Mediates the Effect of Political Ideology on Accuracy Judgements

H6. Source credibility will mediate the effect of political ideology on accuracy judgements overall.

To examine H6, a parallel mediation model was tested. Where previous analyses excluded moderates (as we pre-registered exclusion of moderates for analyses employing contrasts between liberals and conservatives) we here included the 6 moderates from the sample ($N = 156$) to remain consistent with our pre-registration. Using the Lavaan package in R with 10,000 bootstrapped confidence intervals, a structural equation model (SEM) was tested with political ideology as the independent variable, two parallel mediators: 1) credibility of liberal sources and 2) credibility of conservative sources, and two parallel DVs: 1) accuracy judgements of misinformation from liberal sources and 2) accuracy judgements of misinformation from conservative sources. The unstandardised regression coefficient for the total effect of political ideology on accuracy judgements of liberal misinformation was $b = -0.20$, $p < 0.001$, 95% CI [0.28, − 0.12]. Of the total effect, the direct effect of political

ideology on accuracy judgements of liberal misinformation was significant with unstandardised regression coefficient $b = -0.11$, $p = 0.01$, 95% CI [− 0.20, − 0.01]. The unstandardised regression coefficient for the mean bootstrapped indirect effect was $b = -0.1$, $p < 0.001$, 95% CI [− 0.17, − 0.05], with 48.5% of the total effect of political ideology on accuracy judgements of liberal misinformation mediated via credibility of liberal sources.

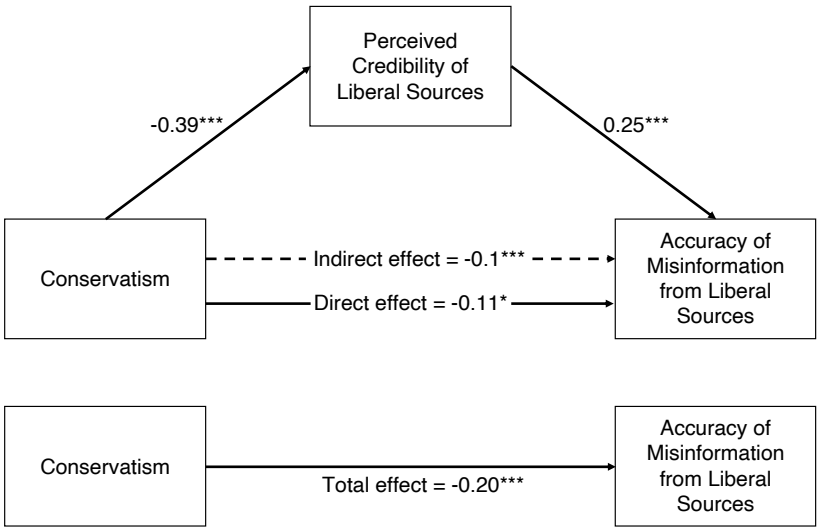
For the conservative misinformation path, the unstandardised regression coefficient for the total effect of political ideology on accuracy judgements of conservative misinformation was $b = 0.17$, $p < 0.001$, 95% CI [0.08, 0.25]. The direct effect of political ideology on accuracy judgements of conservative misinformation was not significant ($b = 0.04$, $p = 0.38$, 95% CI [− 0.06, 0.12]). However, there was a significant indirect effect of political ideology via the credibility of conservative sources, as the unstandardized regression coefficient for the mean bootstrapped indirect effect was $b = 0.13$, $p < 0.001$, 95% CI [0.08, 0.19], mediating 78.2% of the total effect. As such, the effect of political ideology on accuracy judgements of *liberal* misinformation was partially mediated via the credibility of liberal sources and the effect of political ideology on accuracy judgements of *conservative* misinformation was fully mediated via the credibility of conservative sources. That is, the more liberal participants were, the more credible they rated liberal sources to be and resultingly, the more accurate they judged misinformation from those sources to be. Similarly, the more conservative participants were, the more credible they rated conservative sources to be, and, in turn, the more accurate they perceived misinformation from those sources to be. The mediation model path for liberal misinformation is visualised in Figure 13 and the path for conservative misinformation is visualised in Figure 14.

Demographic Variables

Several demographic variables were tested as potential predictors of the main DV (accuracy judgements of misinformation). Of these, neither Twitter use ($F(4,295) = 1.32$, $p = 0.26$), gender ($F(2,297) = 0.54$, $p = 0.58$) nor birthyear ($F(1,298) = 0.88$ $p = 0.35$) were significant predictors of perceived accuracy of misinformation. Although education was found to be a significant predictor in a main effects model ($F(5,294) = 2.50$, $p = 0.03$), Tukey's post-hoc tests showed no significant differences between any group contrasts (all p's > 0.05). Finally, the same was true for primary news source, which was found to be a significant predictor of perceived accuracy of misinformation in a main effects model ($F(5,294) = 2.72$, $p = 0.02$), but Tukey's post-hoc tests revealed no significant contrasts (all p's > 0.05).

Figure 13

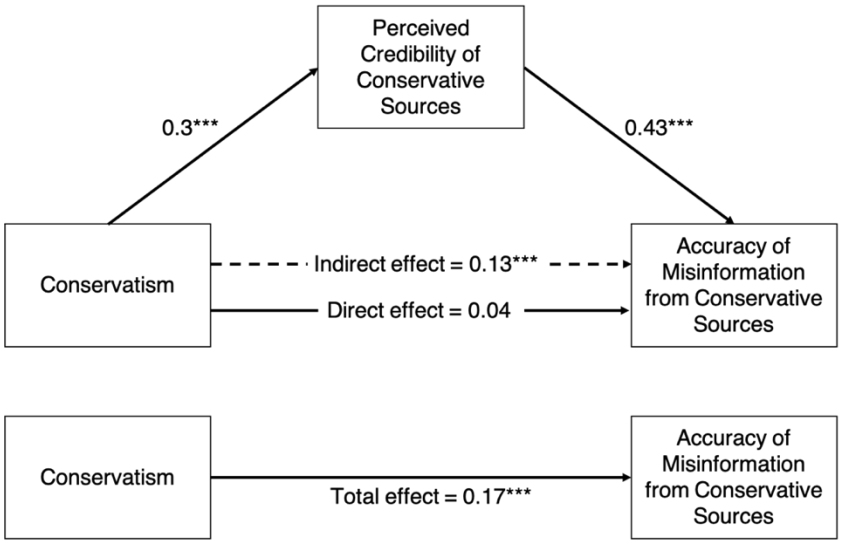
The Effect of Political Ideology on Accuracy Judgements of Misinformation from Liberal Sources via Credibility of Liberal Sources



Note: Participants reported their political ideology on a 7-point Likert scale ranging from (1) very liberal to (7) very conservative. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Figure 14

The Effect of Political Ideology on Accuracy Judgements of Misinformation from Conservative Sources via Credibility of Conservative Sources



Note: Participants reported their political ideology on a 7-point Likert scale ranging from (1) very liberal to (7) very conservative. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

9. Discussion

The main underlying question posed in this study was whether ideological congruence with news sources impacts people's ability to evaluate (mis)information from those sources. This chapter further asked whether this evaluation is driven by credibility assessments and whether individuals judge ideologically incongruent sources to be more politically slanted and less credible than ideologically congruent sources.

Across both studies, ideological congruence with news sources played a significant and considerable role in judgements of (mis)information. The results demonstrated that both conservative and liberal participants were less likely to (incorrectly) judge misinformation as being reliable (Study 1) and accurate (Study 2) when the associated source was ideologically congruent, and liberal participants were more likely to (incorrectly) judge factual information as being inaccurate when the associated source was conservatively slanted. Secondly, the mediation analyses demonstrated that these effects were mediated via credibility assessments of sources, particularly for conservative misinformation. Finally, our findings illustrated that participants found ideologically congruent sources to be less politically slanted than ideologically incongruent sources.

Ideological Congruence with Sources

Classic theories of persuasion highlight the importance of sources (Brinol & Petty, 2009) - in particular source similarity - in the persuasion process (Chaiken & Maheswaran, 1994; Metzger et al., 2003). This chapter demonstrates that regardless of political ideology, individuals use sources to evaluate the accuracy of news headlines. Both liberals and conservatives rated misinformation from politically congruent sources as significantly more accurate than misinformation from politically incongruent sources, suggesting a symmetry between the political bias of the left and right, although the effect size was larger for liberals ($d = 0.66$) than conservatives ($d = 0.32$).

In recent years, great attention has been paid to studying the (a)symmetries between motivated social cognition of conservatives and liberals regarding politically salient information (Harper & Baguley, 2019; Jost et al., 2003; Sindermann et al., 2020; Swire et al., 2017). While original work highlighted conservatism as a driver of motivated social cognition (Jost et al., 2003) and evidence pointed to conservatives being more influenced by source similarity than liberals (Jost et al., 2018), recent evidence suggests that both sides of

the political spectrum have a tendency to regard politically disconcordant news as illegitimate and fake (Harper & Baguley, 2019), potentially due a general tendency of individuals to associate politically incongruent media sources themselves with the term “fake news” (van der Linden et al., 2020), delegitimising any news they may present.

Liberals: Biased or Justifiably Critical?

An interesting asymmetric finding that emerged in this study was that only liberal participants judged factual information as being more inaccurate when the associated source was politically incongruent as opposed to politically congruent. That is, when factual news was presented by traditionally conservative sources, liberal participants were significantly worse at identifying the information as accurate than when the factual news was presented by traditionally liberal sources. Conversely, this was not the case for conservatives, whose perceived accuracy of factual news did not depend on the slant of the associated source. A relevant question to pose is whether this difference exists because conservatives were equally ‘bad’ at correctly judging the accuracy of information from all sources or because a liberal bias prevented liberal participants from seeing truth value in any information presented by conservative sources. That is, are conservatives ‘lazy’ (not engaging in analytical thinking about sources) or are liberals ‘biased’? Although we cannot causally differentiate between the two explanations, we argue that the ‘bias’ displayed by liberals is likely warranted; an argument we unpack further below.

Interestingly, our results showed that liberals and conservatives did not differ statistically in their ability to correctly identify the accuracy of neither fake nor true news headlines overall. However, this may be explained by the fact that liberals were, generally speaking, rating news from liberal sources as highly accurate and news from conservative sources as inaccurate - regardless of content. Conservatives, on the other hand, paid relatively less attention to sources, and while they did rate misinformation from conservative sources as more reliable than misinformation from liberal sources, this difference was smaller ($d = 0.32$) in comparison to liberals, who rated misinformation from liberal sources as considerably more accurate than misinformation from conservative sources ($d = 0.66$).

As bias is a word often associated with negative connotations, we ask whether this ‘bias’ is necessarily bad. On the one hand, overreliance on sources may lead liberals to delegitimise true and accurate information from politically incongruent sources and to be ‘blind’ to misinformation from politically congruent sources. In this way, liberal bias may be

harmful, as even legitimate (including liberal) news sources have been found to publish misleading information occasionally (Iyengar & Massey, 2019) and to have a political slant (Gentzkow & Shapiro, 2010). However, paying attention to news sources has, by many researchers, been credited as one of the most reliable ways of spotting misinformation, which may explain why research consistently finds that liberals are better at discerning between fake and real news (Basol et al., 2020; Pennycook & Rand, 2019a). Furthermore, traditionally right-wing media sources have been shown to publish more misinformation than their ideological counterparts. For example, recent analyses of media consumption data and beliefs in misinformation during the early months of the COVID-19 pandemic showed a clear relationship between far right-wing media consumption and misinformed views (Motta et al., 2020). As such, it may be a beneficial 'bias' - or a reliable heuristic - to judge information from right-wing media sources more critically, as the real signal from these sources is arguably noisier and more unreliable. That being said, while relying on assessments of media sources to evaluate the accuracy of news information may generally serve as an advantage for liberals as it reduces their susceptibility to misinformation from illegitimate sources, it may increase their susceptibility to misleading information from sources they trust, in so far this situation presents itself.

The mediation analyses highlight that the tendency for participants to judge misinformation from conservative sources as accurate was fully mediated by credibility judgements of the corresponding sources. This suggests that liberal participants found misinformation from conservative sources to be inaccurate specifically because they judged conservatively slanted sources to lack credibility. The results further show that the tendency for participants to judge misinformation from liberal sources as accurate was only partially mediated by credibility judgements of the corresponding sources. This allows for the possibility that while liberals rely on credibility assessments of sources to judge information from them, conservatives may not do so to the same extent, as the assessment of liberal misinformation as inaccurate was only partly explained by credibility judgements of liberal sources.

These findings are explained in relation to the supposition above; that conservatives paid less attention to sources than did liberals, echoing previous research, which has found conservatives to be worse at discerning between reliable and unreliable news sources overall (Pennycook & Rand, 2019a). We posit that liberals are more likely to rely on source judgements, specifically assessing the credibility of sources, in order to assess the reliability of associated news. This may explain why conservatives judge (any) information from liberal

sources less ‘harshly’ than when liberals assess information presented by conservative sources.

Classic persuasion literature has long highlighted the importance of source similarity in persuasion, potentially due to source similarity contributing to perceptions of source credibility. Furthermore, perceiving a source to be biased has been shown to negatively impact both persuasion and source credibility (Wallace et al., 2020). These findings largely echo these effects, as in our study, participants rated politically incongruent sources as more biased than politically congruent sources, rated their credibility to be lower, and their news headlines to be less accurate. The findings highlight that these key persuasion processes are mirrored in news consumption on social media, with (political) source similarity largely influencing credibility judgements of news sources, leading individuals to judge news from those sources as either accurate or inaccurate. As individuals tend to pay less attention to media sources, they do not deem to be credible (Johnson & Kaye, 1998), this likely contributes to the increase in polarisation along partisan lines (Iyengar & Westwood, 2015), with individuals assigning partisan positions to news sources and filtering content based on political congruence (Coe et al., 2008; Iyengar & Hahn, 2009; Messing & Westwood, 2014; Stroud, 2008). Indeed, the findings demonstrate that both sides of the political spectrum perceive politically incongruent sources as being significantly more slanted than politically congruent sources, suggesting that each side perceives their ‘own’ news sources as being neutral. While some researchers have suggested that susceptibility to misinformation is driven more by a lack of analytical thinking than partisan bias (Pennycook & Rand, 2019b), this work demonstrates that partisan bias does indeed play a role when news consumers are presented with visually salient sources on social media.

10. Limitations

Of course, this research is not without limitations. Firstly, although the study used non-political headlines as content, it is impossible to eliminate politics from news content completely. Furthermore, to the extent that this was achieved, this is hardly the case in a real news environment, where the political slant of content potentially plays an equal, if not larger, role than the source in credibility judgements. It must also be acknowledged that although using real-world headlines enhances ecological validity, we cannot rule out prior exposure effects (Roozenbeek & van der Linden, 2019). However, to the extent that prior exposure plays a role, it tends to increase rather than decrease belief in misinformation

(Pennycook et al., 2018), and as such it is reassuring that our results showed that participants tended to rate all misinformation headlines as inaccurate.

Second, the distinction between full and partial mediation has become a topic of debate and other mediators not included in the current model may play a role (Rucker et al., 2011). It must be stressed that no causal conclusions can be drawn from the mediation analyses presented here and that the sample was not nationally representative of the U.S. population on political ideology. Future research would benefit from experimentally manipulating the credibility of the news source and replicating these findings using a larger set of headlines, a nationally balanced sample, and party ID instead of ideology. Furthermore, as this chapter focuses on *perceived* source slant and employs an audience-based approach to quantifying media slant, content-based approaches suggest that the conservative sources we included may have a stronger ideological slant in the conservative direction compared to the liberal sources we included (Budak et al., 2016). Finally, while using real-world news outlets increased the ecological validity of these results, this approach also introduces limitations. For instance, news outlets followed by conservative audiences have been shown to publish more misleading stories than news outlets followed by liberals (Sharockman, 2014), a point of consideration when interpreting the results of this study.

11. Conclusion

Despite the limitations, these results demonstrate that source credibility plays a role in susceptibility to persuasion by misinformation, challenging previous claims that sources have little influence on misinformation belief (Pennycook & Rand, 2020). The findings reveal that both conservatives and liberals are more likely to be persuaded by misinformation when it originates from politically congruent sources. Moreover, participants perceived politically aligned news outlets as more credible, which likely contributed to the increased reliability they attributed to news from these sources.

In summary, political alignment and source credibility significantly influenced how participants judged misinformation. While participants generally distinguished between factual and misleading information, misinformation from politically congruent sources was still perceived as more accurate. A strong correlation emerged between perceptions of source credibility and the political alignment of the source with the recipient, suggesting that political congruence amplifies trust in the source. However, further research is necessary to determine whether the observed effects are solely driven by political alignment or if prior

exposure to specific news sources also plays a role. Additionally, it remains unclear whether these effects persist when sources are known to lack credibility.

The next chapter will build on these findings by investigating whether source credibility has a causal effect on misinformation judgements and seeks to disentangle the effects of source credibility and similarity comprehensively.

3. DISENTANGLING THE EFFECTS OF SOURCE CREDIBILITY AND SIMILARITY

This chapter features Study 4 from the comprehensive paper published in *Nature Scientific Reports* in the article below:

Traberg, **C. S.**, Harjani, T., Roozenbeek, J., & van der Linden, S. (2024). The persuasive effects of social cues and source effects on misinformation susceptibility. *Scientific Reports, 14*, 4205. https://doi.org/10.1038/s41598-024-54030-y

1. Abstract

This chapter examines the independent and interactive effects of source credibility and political similarity on the perceived reliability of misinformation. Building on previous findings in Chapter 2 showing that political congruence between a news source and the individual increases susceptibility to misinformation, this study disentangles the roles of similarity and credibility to address whether shared identity or perceived credibility holds greater sway in judgements of misinformation. Using a 2 x 2 factorial design (N = 790), the study manipulated source similarity (similar vs dissimilar political orientation) and source credibility (high vs low) while measuring the perceived reliability of misinformation headlines. The results reveal that both source similarity and credibility independently increase the perceived reliability of misinformation, with credibility exerting a stronger influence. Importantly, similarity amplified the effects of credibility but did not compensate for low credibility. This pattern was particularly pronounced among liberals, who showed greater sensitivity to similarity effects than conservatives. Cognitive reflection emerged as a moderator, with lower scores associated with greater reliance on credibility cues, while higher scores predicted a more consistent scepticism across conditions. These findings contribute to understanding the nuanced interplay between credibility and similarity in shaping misinformation susceptibility.

2. Introduction

The previous chapter investigated the impact of sources on susceptibility to non-partisan misinformation, finding that individuals were more susceptible to misinformation from U.S. news outlets that shared their political identity, an effect mediated by perceived source credibility. While the findings suggested that source similarity and credibility both contribute to judgements of information reliability, their individual roles and relative influence remain unclear. Chapter 3 seeks to address this gap by isolating the effects of source credibility and similarity to better understand their independent (and potentially interactive) impacts on misinformation judgements. This introduction will first revisit foundational theories on source credibility and social identity to frame the potential distinct contributions of each factor to reliability judgements in the context of misinformation.

Expanding the Theoretical Framework

Source credibility and similarity have long been recognised as critical factors in persuasion research. Early work by Hovland & Weiss (1951) identified expertise and trustworthiness as essential components of source credibility, influencing how messages are perceived regardless of content. Trustworthiness refers to a source's perceived honesty or reliability, while expertise reflects the source's knowledge or competence in a specific area (Hovland et al., 1953). Decades of research has shown that sources perceived as trustworthy and knowledgeable are generally more persuasive, with information from such sources having a stronger impact on attitudes and beliefs (Petty & Cacioppo, 1986; Pornpitakpan, 2004). However, the similarity between the source and the audience - alignment in values, attitudes, or group identity - also plays a significant role in both cognition and persuasion. Research has demonstrated that similarity can enhance perceptions of credibility (Pornpitakpan, 2004). Indeed, as Chapter 2 showed, political source congruence predicted perceived source credibility, suggesting that people may judge a source as more credible when it aligns with their political identity. This echoes previous research showing that social identities can significantly influence our cognition (Van Bavel et al., 2014), with this finding suggesting this may be mediated via ingroup sources benefitting from higher perceived credibility.

Other empirical studies have explored these factors in digital contexts, where credibility cues are often less direct and more varied. For instance, Tormala & Petty, (2004) found that high-credibility sources prompted more favourable message evaluations, even

under low-effort processing conditions. Additionally, in online environments, cues like verified badges, institutional affiliations, and endorsements have been shown to shape credibility judgements (Flanagin & Metzger, 2007).

While the previous chapter demonstrated that source similarity predicts credibility perceptions (Traberg & van der Linden, 2022), one could imagine a scenario where the two factors may be contradictory. For instance, consider a news outlet that aligns closely with an individual's political beliefs, consistently reinforcing their ideological stance on contentious issues such as immigration, climate policy, or healthcare (i.e., high similarity/ in-group). The audience may perceive this alignment through the outlet's selective coverage, framing of events, and repeated emphasis on values that resonate with their worldview, fostering a sense of trust based on shared identity (Tanis & Postmes, 2005). At the same time, this outlet might have a documented history of spreading misinformation, with independent fact-checkers revealing instances of exaggerated claims, cherry-picked data, or even outright falsehoods. Credible media watchdogs could publicise such findings, or users might encounter fact-checking labels attached to the outlet's posts on social media platforms. This knowledge about the outlet's credibility issues may create tension for the audience. While they feel a natural affinity for the source due to shared values, they may also be wary of completely trusting its content due to its track record of inaccuracy.

This scenario illustrates a potential conflict between the influence of similarity and credibility, particularly in the context of misinformation. The similarity might make audiences receptive to the message, yet low credibility could prompt scepticism. Although the importance of both factors is well-documented in both persuasion research (Chaiken & Maheswaran, 1994; Hovland & Weiss, 1951; Pornpitakpan, 2004) and social identity research (Pereira et al., 2021; Van Bavel et al., 2024; Van Bavel & Pereira, 2018), few studies have examined their independent effects within the misinformation domain. This gap in the literature raises an important question: when faced with news sources that are ideologically aligned but lack credibility, are audiences more influenced by similarity or credibility? This study seeks to address this question by testing the independent impact of each factor on perceptions of misinformation credibility, providing a clearer understanding of their roles in a polarised information environment.

The Convergence Between Source Credibility and Similarity

This brings us to the current state of research, where there is often a convergence between similarity and credibility, effectively blurring the lines between the two. Recent studies suggest that while similarity may increase trust in a source, it does not always compensate for a lack of inherent credibility (Birnbaum & Stegner, 1979; Metzger et al., 2003). Social Identity Theory (SIT), developed by Tajfel and Turner (1979), helps explain why this convergence occurs, as it posits that individuals categorise themselves and others into social groups based on shared characteristics like political beliefs, cultural backgrounds, or values. According to SIT, people are inclined to exhibit ingroup favouritism, showing a preference for information that aligns with their own group identities. This can lead to a bias in favour of sources perceived as similar, even when those sources lack objective credibility (Spears, 2021). This tendency to favour ingroup sources may contribute to heightened trust in messages from similar sources, as people often perceive alignment with their values as a sign of reliability. For example, Van Bavel & Pereira (2018) proposed the "Identity-Based Model of Political Belief", arguing that partisanship and social identity can bias information processing. Reviewing extensive evidence showing that individuals often engage in motivated reasoning, the authors further explain how partisanship can even alter memory and perceptual judgements. In a recently updated model, Van Bavel et al. (2024) integrate the role of informational context into the Identity-Based Model of Political Belief, highlighting how identity goals can influence the informational context by shaping the sources individuals seek and trust, as well as how accuracy goals might be differentially activated depending on the social and informational environment. A key takeaway from this model is that social identity goals can override accuracy goals. This dynamic interplay between social identity and accuracy motivations underscores how group membership influences information processing, sometimes creating a convergence between source similarity and credibility that can complicate reliability judgements.

Indeed, prior research has shown that similarity cues, particularly those tied to shared identity, can lead individuals to trust and accept information more readily from ingroup sources than from outgroup or neutral sources (Turner et al., 1987). However, this preference may create tension when a source is politically congruent but lacks credibility or vice versa. This raises key questions about the interaction between these factors: Does similarity or objective credibility exert a stronger influence on susceptibility to misinformation? And what happens when these cues conflict?

This chapter extends the persuasion literature to address these questions by disentangling the effects of source similarity and credibility. It builds on SIT by investigating whether shared identity alone is sufficient to enhance perceptions of information reliability or if objective credibility remains dominant when these factors are examined independently. Understanding the interplay between similarity and credibility not only advances theoretical perspectives on persuasion but further holds practical implications for combating misinformation, where politically similar yet unreliable sources may have an unfortunate influence on individual news consumers.

In the previous chapter, another question emerged: As the study used real news outlets, whether the observed effects were driven by participants' prior experience and familiarity with these sources remains uncertain. First, was the increased effect of partisanship stronger for liberals simply because they are more driven by partisanship in their judgements of news reliability, or does this finding have more to do with the concrete liberal vs conservative sources used in the study paradigm, which may differ in their actual credibility (Sharockman, 2014)? This question will be explored by examining whether the same patterns would hold for sources with which individuals have no prior experience.

Prior Experience with News Outlets

Research suggests that credibility judgements about a source can have lasting effects, leading individuals to trust information from previously trustworthy sources even when credibility indicators are later mixed (Sterrett et al., 2019). For instance, one study found that participants were more likely to trust news articles shared on social media by elite sources they had previously deemed trustworthy, as opposed to those they had judged untrustworthy (Sterrett et al., 2019). Notably, trust ratings were unaffected by whether the article originated from a real or fake news outlet (Sterett et al., 2019), a result supported by other studies (Tsang, 2020), suggesting that initial credibility assessments can have lasting impacts. Indeed, people may be more likely to believe misinformation from real news sources that share their political identity based on past experience, as certain outlets have been shown to publish more misleading news than others (Motta et al., 2020; Sharockman, 2014). This raises a question central to this chapter: Do people fall for misinformation from similar sources simply due to a preference for similar sources, even when credibility is compromised?

The Current Study

The primary aim of this study is to disentangle the effects of source credibility and similarity on information judgements to determine each factor's impact independently and to examine potential interaction effects. By disentangling the effects of credibility and similarity, this study aims to enhance our understanding of how people assess information reliability in digital contexts. These insights are particularly relevant for designing interventions to counteract misinformation by addressing similarity-related cognitive biases. The study builds on Chapter 2 by examining source credibility and similarity independently and builds on these by examining further pre-registered hypotheses (https://doi.org/10.17605/OSF.IO/WK9YA):

H1: Source credibility will impact reliability judgements of misinformation such that participants in the high source credibility condition will have higher reliability judgements of misinformation than participants in the low source credibility condition.

H2: Source similarity will impact reliability judgements of misinformation such that participants will have higher reliability judgements of misinformation when the source is politically similar compared to when the source is politically dissimilar.

H3: There will be an interaction between source similarity and political ideology of participants such that the effect of source similarity on reliability judgements of misinformation is greater for liberals than for conservatives.

H4: There will be an interaction between source credibility and political ideology of participants such that the effect of source credibility on reliability judgements of misinformation is greater for liberals than for conservatives.

H5: There will be an interaction between source similarity and source credibility such that (a) the effect of source credibility on reliability judgements of misinformation will be greater when the source is politically dissimilar compared to when it is similar and (b) the effect of source similarity on reliability judgements of misinformation will be greater when source credibility is low compared to when source credibility is high.

H6: Cognitive Reflection Test scores will moderate the effect of (a) source credibility and (b) source similarity on reliability judgements of misinformation.

H7: Need for Cognition will moderate the effect of (a) source credibility and (b) source similarity on reliability judgements of misinformation such that when Need for Cognition is low, there will be a larger effect of both source manipulations on reliability judgements of misinformation.

3. Method

Participants

This study recruited $N = 790$ US participants from Prolific who either identified as liberal or conservative (65% female, $M_{age} = 35$, 62% completed higher education, 52% liberal, 86% had a Facebook account, 34% got the majority of their news from social media). All participants provided informed consent and were compensated for their time. The study was approved by the Cambridge Psychology Research Ethics Committee (PRE.2021.066). All methods were performed in accordance with the relevant guidelines and regulations.

Design and Materials

This study employed a 2x2 between-subjects factorial design manipulating source similarity (similar vs dissimilar) and source credibility (high credibility vs low credibility) (see Figure 15).

Source Manipulation

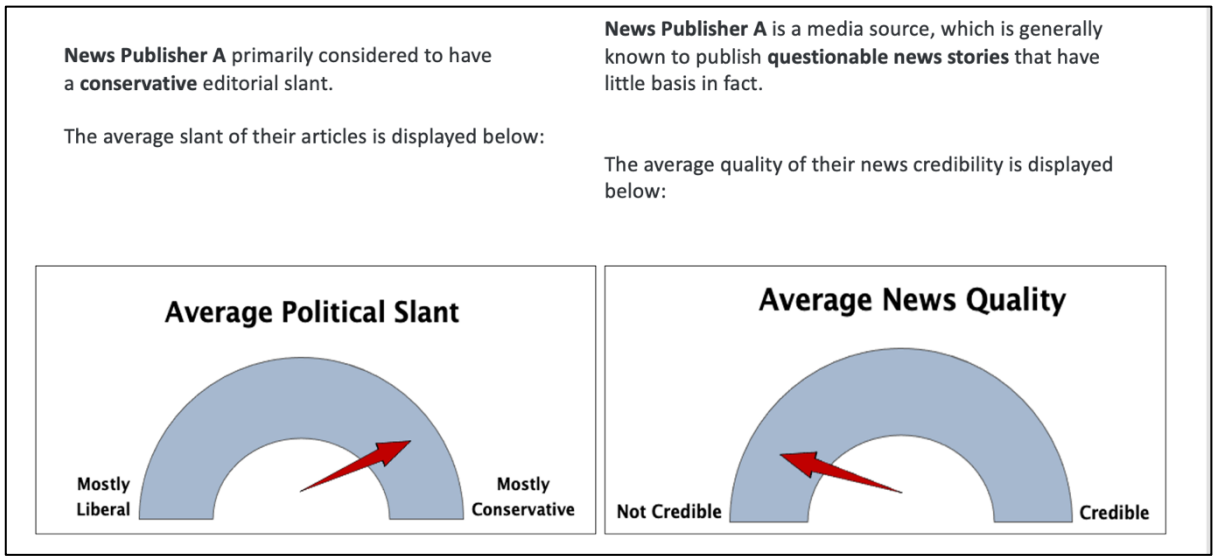
This study manipulated attributes of news sources using descriptive vignettes and graphical "metre" displays. Participants were presented with descriptions of a fictitious news source referred to as "News Publisher A," ensuring they had no prior familiarity with it. Each participant read two descriptions of the news source. The first described the source's credibility. In the low-credibility condition, the description read: *"News Publisher A is a media outlet generally known to publish questionable news stories with little basis in fact. The average quality of their news credibility is displayed below."* In the high-credibility condition, the description stated: *"News Publisher A is a mainstream media outlet generally known to publish news stories that adhere strictly to journalistic principles of truthfulness and accuracy. The average quality of their news credibility is displayed below."* The second description addressed the source's editorial slant. In the liberal condition, the source was described as having a liberal slant: *"News Publisher A is primarily considered to have a liberal editorial slant. The average slant of their articles is displayed below."* In the conservative condition, the description read*: "News Publisher A is primarily considered to have a conservative editorial slant. The average slant of their articles is displayed below."* After reading each description, participants viewed graphical metre visualisations that corresponded to the information provided (see Figure 15). The order of the credibility and editorial slant descriptions was counterbalanced to minimise potential order effects.

Source Similarity Coding

Source similarity was coded based on the participants' self-reported political ideology. A source was categorised as similar if its editorial slant aligned with the participant's political ideology (e.g., a liberal source and a liberal participant). Conversely, it was categorised as dissimilar if the editorial slant of the source opposed the participant's ideology (e.g., a conservative source and a liberal participant).

Figure 15

Example of Fictitious Source Description of 'News Publisher A'



Note: The above represents the conservative, low credibility condition

Headlines

Participants were exposed to five misleading headlines and five non-misleading headlines. These headlines were similar to those used in Chapter 2 and used one of three misleading tactics: 1) creating or inspiring conspiratorial thinking, 2) discrediting otherwise reputable individuals, institutions, or facts to instil doubt in audiences and 3) using exaggeratedly emotional language to distort the news story to generate strong emotional responses. The five non-misleading headlines included two headlines which had previously been used in Chapter 2. Still, this time, we also included three new non-misleading headlines based on current factual events, which did not use any misleading elements. An overview of all headlines used in this study can be found in Table 3.

Table 3

Full Overview of Headlines

Misinformation vs Facts	Category	Content
Misinformation	Conspiracy	Uber Creepy: Tracking your every move
Misinformation	Conspiracy	Exposing the shadow "elite" controlling the world
Misinformation	Emotion	Horrific TV show inspiring suicide, says filmmaker
Misinformation	Discredit	Expert: Scientific studies no longer trustworthy
Misinformation	Discredit	Scandal: MRI Brain imaging completely unreliable
Factual information	-	Physical fitness keeps your brain in good shape
Factual information	-	Apple, Google and Amazon named as most valuable brands in the world
Factual information	-	Netflix to include mobile games for subscribers
Factual information	-	The Mandalorian and The Crown Tie for Most Nods in This Year's Emmy Nominations
Factual information	-	Abba reunite for Voyage, first new album in 40 years

Measures

Perceived Reliability of Misinformation

Participants were asked to report their perceived reliability of misinformation headlines: *"How reliable is the above news headline?"* (1 = Very Unreliable, 7 = Very Reliable). Participants thereby judged the reliability of five misinformation headlines (M = 3.19, SD = 1.23, α = 0.85).

Perceived Reliability of Factual Information

Participants were asked to report their perceived reliability of factual headlines: *"How reliable is the above news headline?"* (1 = Very Unreliable, 7 = Very Reliable). Participants thereby judged the reliability of five misinformation headlines (M = 5.35, SD = 1.19, α = 0.86).

Cognitive Reflection

Cognitive Reflection was assessed using the 3-item Cognitive Reflection Test (Frederick, 2005), which measures participants' ability to override intuitive but incorrect answers in favour of reflective, correct responses. Participants could score between 0 (no correct answers) and 3 (all correct answers) (M = 1.80, SD = 0.82).

Need for Cognition

The Need for Cognition Scale (Lins de Holanda Coelho et al., 2020) assessed participants' tendency to engage in and enjoy effortful cognitive activities. The scale consists of six items (e.g., "I would prefer complex to simple problems" and "Thinking is not my idea of fun (reverse-scored)") rated on a 5-point Likert scale ranging from 1 (Extremely uncharacteristic of me) to 5 (Extremely characteristic of me). Total scores range from 0 to 30, with higher scores indicating greater need for cognition (M = 21.23, SD = 5.16, α = 0.89).

Demographics

Participants answered demographic questions including Age (year of birth) (M_{age} = 35), Education (Less than high school (0.25%), high school graduate (37.72%), Bachelor's degree (43.67%), Master's degree (14.43%), Doctoral degree (1.90%), Professional degree (2.03%)), Sex (Male (33.04%), Female (65.44%), Other (1.52%)), Political Affiliation (7-point Likert scale arranged from Extremely Liberal (left) to Extremely Conservative (right) (M = 3.79 , SD = 2.18)), Political Party Preference (Strongly Democratic (18.35%), Democratic (25.32%), Lean Democratic (8.73%), Lean Republican (15.70%), Republican (20.13%), Strongly Republican (11.77%)), Facebook use ("I don't have an account" (13.67%); "I have an account but I hardly ever use it" (25.57%); "I have an account, and I use it occasionally" (22.03%); "I have an account and I use it often" (14.43%); "I have an account and I use it on a daily basis" (24.30%)) and News Consumption source ("I don't really follow the news" (7.47%), "Social Media" (33.67%), "TV and radio" (17.10%); "Print Media" (newspapers, magazines) (2.91%); "Word of Mouth" (1.90%), and "Online news sites (excluding social media) (36.96%)".

Timing

Response times were recorded for each question across three metrics: (1) time to first click upon viewing the question page, (2) time to last click before moving on, and (3) total time taken to submit a response for each question.

Attention Check

At the end of the study, we included an attention check to ensure that participants had read the original source descriptions. Specifically, participants were asked two questions with the first being: *"How was the credibility of News Publisher A described?"* with response options being: 1) "News Publisher A was described as a mainstream media source, which is

generally known to publish news stories that adhere strictly to journalistic principles of truthfulness and accuracy", 2) "News Publisher A was described as a media source, which is generally known to publish questionable news stories that have little basis in fact" or 3) "I didn't see any such information". The second attention check question was: *"How was the political slant of News Publisher A described?"*with response options being: 1) "News Publisher A was described as primarily having a liberal editorial slant", 2) "News Publisher A was described as primarily having a conservative editorial slant" or 3) "I didn't see any such information". 53 out of 843 participants failed the attention checks and were therefore excluded from analyses, leaving 790 participants.

Procedure

Participants were recruited via Prolific, where the study was advertised as a study on "News Evaluation". Requirements for participating were that they had not previously participated in studies run by the Cambridge Social Decision-Making Lab (which often revolved around similar themes), that they were above 18 years old, that they were fluent English speakers, and that they identified as either Conservative or Liberal. Once they were redirected to the Qualtrics platform, where the study was hosted, they were provided with information about the study and provided informed consent prior to starting. Then, participants answered the demographic questions. Following this, participants were randomly assigned to one of the four conditions.

Before being asked to rate the reliability of headlines, participants were shown the source vignettes and metres in line with their condition. They were then presented with the 10 news headlines in randomised order and asked to rate the reliability of each of the headlines. Following the main experiment, participants proceeded to complete the Need for Cognition scale and the Cognitive Reflection Test, and then the demographic questions. At the end, they completed an attention check. Finally, participants were debriefed.

4. Results

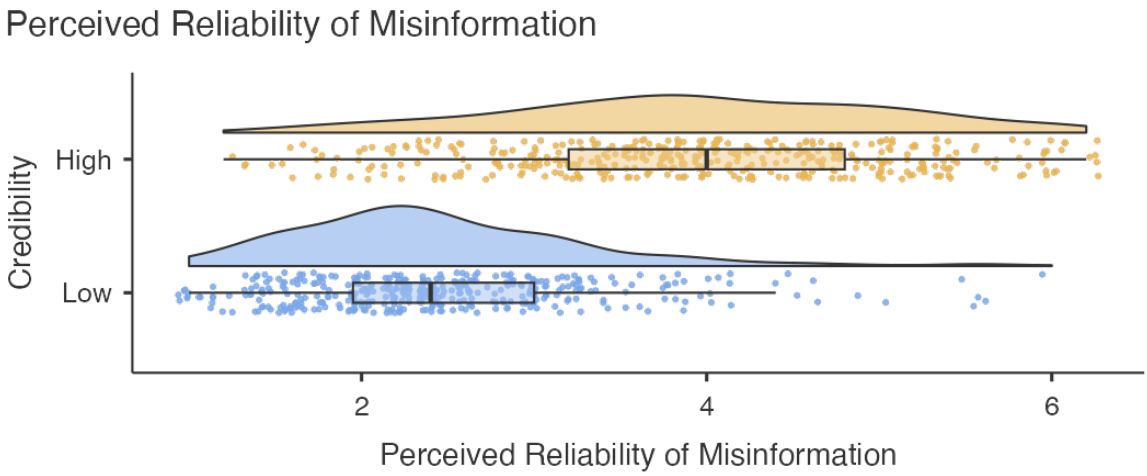
H1. Source credibility will impact reliability judgements of misinformation such that participants in the high source credibility condition will have higher reliability judgements of misinformation than participants in the low source credibility condition.

To test H1, a *t*-test was run comparing judgements of false headlines between the two source credibility conditions (high credibility vs low credibility). As hypothesised, the analysis showed that participants rated misinformation from credible sources as being more

reliable (M = 3.94, SD = 1.10) compared to misinformation from sources lacking credibility (M = 2.43, SD = 0.83, $t(788) = 21.78$, $p < 0.001$, d = 1.55) (see Figure 16). We thereby find support for H1.

Figure 16

Average Perceived Reliability of Misinformation Headlines by Credibility Condition

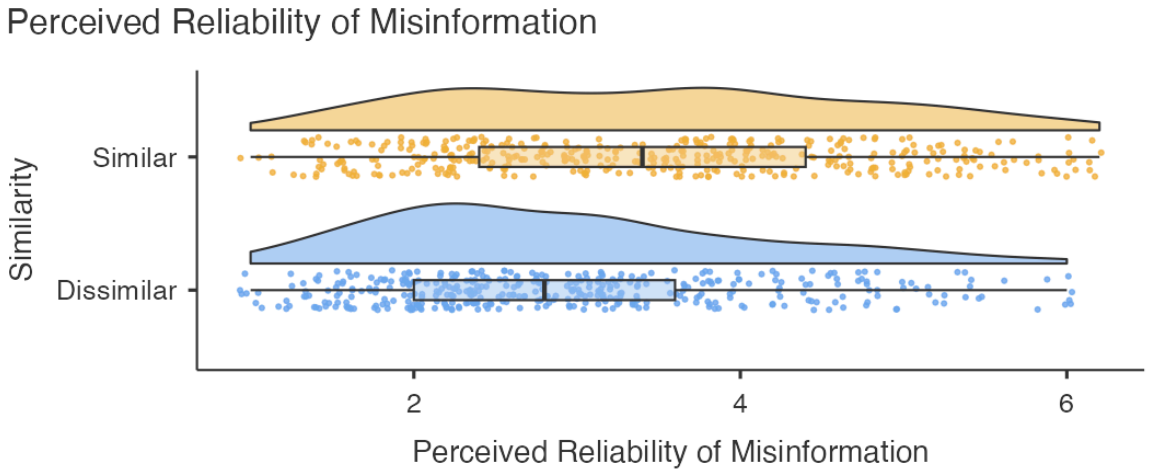


H2. Source similarity will impact reliability judgements of misinformation such that participants will have higher reliability judgements of misinformation when the source is politically similar compared to when the source is politically dissimilar.

To test H2, a t-test was run comparing judgements of false headlines between the two source similarity conditions (similar vs dissimilar). As hypothesised, participants rated misinformation from similar sources as being significantly more reliable (M = 3.46, SD = 1.28) compared to misinformation from dissimilar sources (M = 2.93, SD = 1.13, $t(788) = 6.18$, $p < 0.001$, d = 0.44). H2 was thereby supported as participants rated misinformation from politically similar sources as significantly more reliable than misinformation from politically dissimilar sources (see Figure 17).

Figure 17

Average Perceived Reliability of Misinformation Headlines by Similarity Condition



H3. There will be an interaction between source similarity and the political ideology of participants such that the effect of source similarity on reliability judgements of misinformation is greater for liberals than for conservatives.

An ANOVA was run with source similarity and participant political ideology (and their interaction) as independent variables and perceived reliability of misinformation as the dependent variable. Results showed a main effect of source similarity ($F(1, 786) = 40.82$, $p < 0.001$, $\eta_p^2 = 0.05$), a main effect of political ideology ($F(1,786) = 45.24$, $\eta_p^2 = 0.05$, $p < 0.001$), but no significant interaction between the two ($F(1, 786) = 0.66$, $p = 0.416$, $\eta_p^2 = 0.00$). As such, H3 was rejected. Exploratory post-hoc tests showed that conservatives were more likely to judge misinformation headlines as reliable compared to liberals ($t(786) = 6.73$, $M_{diff} = 0.56$, $p < 0.001$).

H4. There will be an interaction between source credibility and political ideology of participants such that the effect of source credibility on reliability judgements of misinformation is greater for liberals than for conservatives.

An ANOVA was run with source credibility and participant political ideology (and their interaction) as independent variables and perceived reliability of misinformation as the dependent variable. Results showed a main effect of source credibility ($F(1, 786) = 525.08$, $p < 0.001$, $\eta_p^2 = 0.40$), a main effect of political ideology ($F(1,786) = 65.00$, $p < 0.001$, $\eta_p^2 =$

0.09), but no significant interaction between the two ($F(1, 786) = 2.14$, $p = 0.144$, $\eta_p^2 = 0.00$). As such, H4 was rejected.

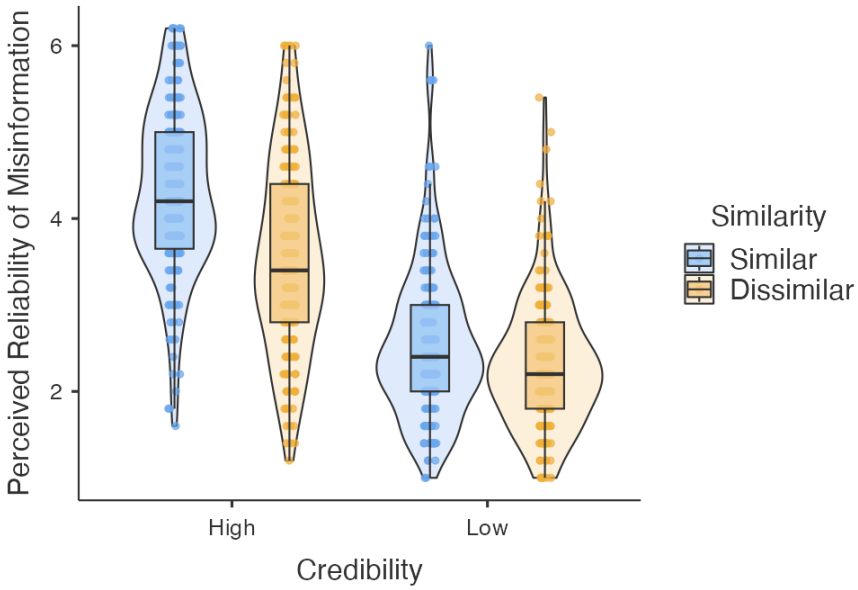
H5. There will be an interaction between source similarity and source credibility such that (a) the effect of source credibility on reliability judgements of misinformation will be greater when the source is politically dissimilar compared to when it is similar and (b) the effect of source similarity on reliability judgements of misinformation will be greater when source credibility is low compared to when source credibility is high.

A 2x2 factorial ANOVA was run with source similarity and source credibility and their interaction as independent variables and perceived reliability of misinformation as the dependent variable. Results revealed both a significant main effect of source credibility ($F(1,786) = 508.25$, $p < 0.001$, $\eta_p^2 = 0.39$), and source similarity ($F(1,786) = 54.08$, $p < 0.001$, $\eta_p^2 = 0.06$). The interaction between source credibility and similarity was also significant ($F(1,786) = 15.32$, $p < 0.001$, $\eta_p^2 = 0.02$).

Examining H5a, post-hoc comparisons showed that when the source was similar, credibility had a larger effect ($M_{diff} = 1.76$, $p < 0.001$, $d = 1.88$) compared to when the source was dissimilar ($M_{diff} = 1.24$, $p < 0.001$, $d = 1.33$), but credibility was significant across both levels of similarity. As such, H5a is rejected, as credibility had a larger effect when the source was similar. Examining H5b, post-hoc comparisons showed that when source credibility was low, source similarity did not have a significant effect on perceived reliability of false headlines ($M_{diff} = -0.23$, $p = 0.07$, $d = -0.24$), whereas when source credibility was high, source similarity had a significant effect ($M_{diff} = -0.75$, $p < 0.001$, $d = -0.80$). As such H5b is not supported, as when source credibility was low, there was no effect of similarity. Results from H5 are visualised in Figure 18.

Figure 18

Average Perceived Reliability of Misinformation by Similarity (Dissimilar vs Similar) and Credibility (High vs Low)



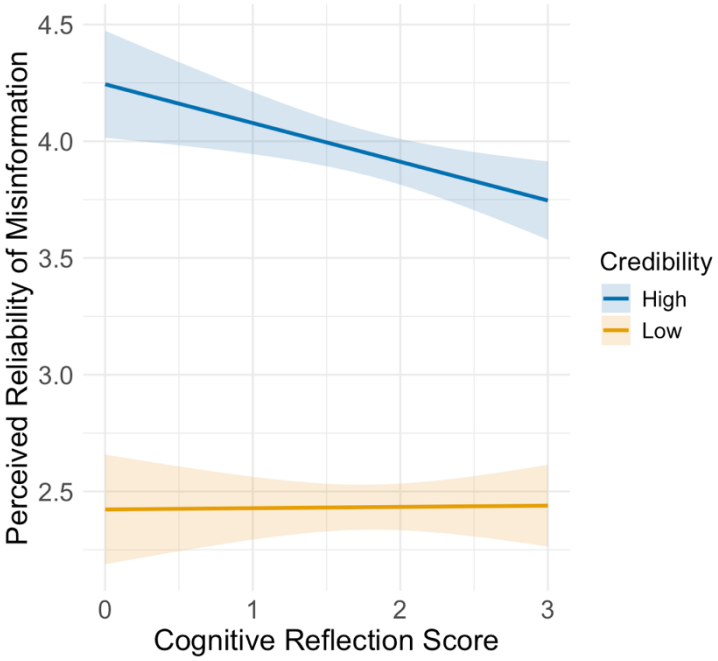
H6. Cognitive Reflection Test scores will moderate the effect of (a) source credibility and (b) source similarity on reliability judgements of misinformation.

To test H6a, a linear regression analysis was conducted. Source credibility, cognitive reflection score, and the interaction term between source credibility and cognitive reflection score were included as predictors, with perceived reliability of misinformation as the dependent variable. The overall model was signifiant: $R^2 = 0.38$, $F(3,786) = 160.75$, $p < 0.001$. The analysis revealed a significant main effect of cognitive reflection score, $b = -0.17$, $SE = 0.06$, $t(786) = -2.83$ $p < 0.01$, indicating that higher cognitive reflection scores were associated with lower perceived reliability of misinformation. A significant main effect of source credibility was also found ($b = -1.82$, $SE = 0.17$, $t(786) = -10.90$, $p < 0.001$), with misinformation from high-credibility sources perceived as more reliable than low-credibility sources. Importantly, there was a significant interaction between source credibility and cognitive reflection score, $b = 0.17$, $SE = 0.08$, $t(786) = 2.02$, $p = 0.04$, supporting H6a. This interaction indicates that cognitive reflection scores moderated the effect of source credibility on perceived reliability. Specifically, individuals with lower cognitive reflection scores showed a greater difference in reliability judgements of misinformation between high- and low-credibility sources, suggesting a stronger reliance on credibility cues. In contrast, individuals with higher cognitive reflection scores demonstrated less reliance on source

credibility, showing a more consistent level of scepticism regardless of the source's credibility. This result is visualised in Figure 19.

Figure 19

Average Perceived Reliability of Misinformation as a Function of Cognitive Reflection and Source Credibility (High vs. Low)



A linear regression analysis was conducted to test H6b, which predicted that cognitive reflection scores would moderate the effect of source similarity on reliability judgements of misinformation. Source similarity, cognitive reflection score, and the interaction term between source similarity and cognitive reflection score were included as predictors, with the perceived reliability of misinformation as the dependent variable. The overall model accounted for a small portion of the variance in reliability judgements, $R^2 = 0.05$, $F(3,786) = 13.83$, $p < 0.001$. The main effect of cognitive reflection score was not statistically significant, $b = -0.10$, $SE = 0.07$, $t(786) = -1.33$, $p = 0.183$, suggesting that cognitive reflection scores alone did not significantly influence perceived reliability in this context. However, source similarity had a significant main effect, $b = -0.67$, $SE = 0.21$, $t(786) = -3.24$, $p < 0.01$, indicating that information from similar sources was perceived as less reliable than information from dissimilar sources.

The interaction between source similarity and cognitive reflection score was not significant, $b = 0.08$, $SE = 0.11$, t(786) = 0.77, $p = 0.44$. This indicates that cognitive reflection scores did not moderate the effect of source similarity on perceived reliability. Thus, H6b is not supported, as cognitive reflection scores did not alter the influence of source similarity on reliability judgements of misinformation. These results suggest that, unlike source credibility, the effect of source similarity on reliability judgements is not influenced by cognitive reflection scores. This implies that individuals' cognitive reflection levels do not impact their reliance on similarity cues when evaluating the reliability of misinformation.

H7. Need for Cognition will moderate the effect of (a) source credibility and (b) source similarity on reliability judgements of misinformation such that when Need for Cognition is low, there will be a larger effect of both source manipulations on reliability judgements of misinformation.

To test H7a, which predicted that Need for Cognition (NFC) would moderate the effect of source credibility on reliability judgements of misinformation, a linear regression analysis was conducted. The model included source credibility, Need for Cognition (NFC), and the interaction term between source credibility and NFC as predictors, with perceived reliability of misinformation as the dependent variable. The overall model explained a significant portion of variance in reliability judgements, $R^2 = 0.38$, $F(3,786) = 160.75$, $p < 0.001$. The analysis revealed a significant main effect of source credibility, $b = -1.11$, $SE = 0.29$, $t(786) = -3.76$, $p < 0.0001$, indicating that high-credibility sources were perceived as more reliable than low-credibility sources. However, the main effect of Need for Cognition (NFC) was not significant, $b = 0.02$, $SE = 0.01$, t(786)=1.64, $p = 0.102$, suggesting that NFC alone did not significantly influence perceived reliability in this model. The interaction between source credibility and NFC was also not significant, $b = -0.02$, $SE = 0.01$, t(786) = −1.43, $p = 0.154$. This indicates that NFC did not moderate the effect of source credibility on reliability judgements of misinformation. Thus, H7a is not supported. These results suggest that, contrary to the hypothesis, individuals' Need for Cognition did not influence the extent to which they relied on source credibility when evaluating the reliability of misinformation.

To test H7b, which predicted that Need for Cognition (NFC) would moderate the effect of source similarity on reliability judgements of misinformation, a linear regression analysis was conducted. The model included source similarity, Need for Cognition (NFC), and the interaction term between source similarity and NFC as predictors, with perceived reliability of misinformation as the dependent variable. The overall model explained a small

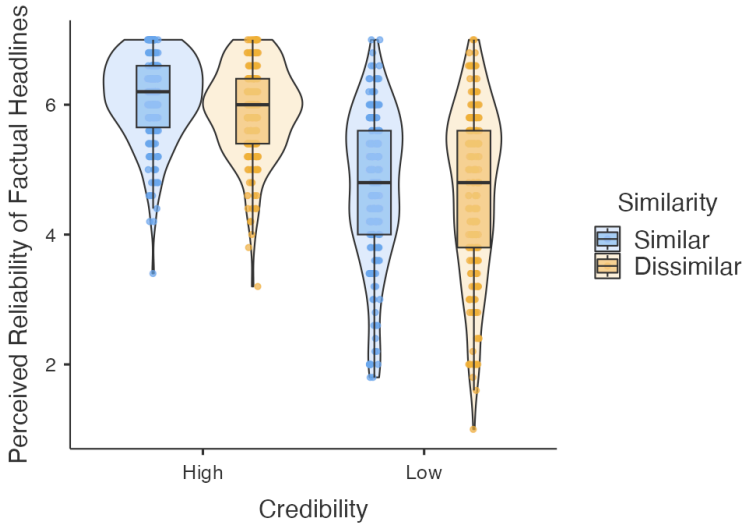
portion of variance in reliability judgements, $R^2 = 0.05$, $F(3,786) = 13.83$, $p < 0.001$. The main effect of Need for Cognition (NFC) was not significant, $b = 0.00$, $SE = 0.01$, $t(786) = 0.35$, $p = 0.723$, indicating that NFC alone did not influence perceived reliability in this model. Additionally, the main effect of source similarity was not significant, $b = -0.48$, $SE=0.36$, $t(786) = -1.32$, $p = 0.186$, suggesting that similarity alone did not significantly affect reliability judgements in this context. The interaction between source similarity and NFC was also not significant, $b = -0.00$, $SE = 0.02$, $t(786) = -0.14$, $p = 0.89$. Thus, H7b is not supported. These results suggest that contrary to the hypothesis, individuals' Need for Cognition did not influence the extent to which they relied on source similarity when evaluating the reliability of misinformation.

The Impact of Sources on Perceived Reliability of Factual Headlines

An exploratory factorial ANOVA with credibility and similarity as the IVs and perceived reliability of factual headlines as the DV showed no significant effect of similarity ($p = 0.09$), but an effect of credibility ($F(1,786) = 328.33$, $p < 0.001$, $\eta_p^2 = 0.29$), with perceived reliability being higher for news outlets described as high in credibility compared to low ($M_{diff} = 1.29$, $SE = 0.07$, $p < 0.001$, $d = 1.29$). There was no interaction between credibility and similarity ($p = 0.31$) (see Figure 20).

Figure 20

Average Perceived Reliability of Factual Headlines by Similarity (Similar vs Dissimilar) and Credibility (High vs Low)



Exploratory Analyses

As data was collected on response times, an exploratory analysis was run to examine whether a) individuals differed in their response times of reliability judgements for misinformation vs factual information and b) whether there was any difference in response time with regards to misinformation and factual information from credible vs non-credible and similar vs dissimilar sources. First, a t-test showed that participants did not differ in their response times for misinformation vs factual information ($t(789) = 1.78$, $M_{diff} = 0.5$, $p = 0.075$, $d = 0.06$).

Response times differ by source credibility, not similarity

A two-way ANOVA was then conducted to examine the effects of source credibility and similarity on response time for misinformation items. The analysis revealed a significant main effect of credibility on response time, $F(1,786) = 15.07$, $p < 0.001$, $\eta_p^2 = 0.02$, $p = 0.02$. Participants took significantly longer to respond when the source was high in credibility compared to low credibility ($M_{diff} = 1.95$, $SE = 0.50$, $t(786) = 3.88$, $p < 0.001$, $d = 0.28$. There was no significant main effect of similarity on response time, $F(1,786) = 0.17$, $p = 0.679$, $\eta_p^2 = 0.00$ and the interaction between credibility and similarity was not significant, $F(1,786) = 0.60$, $p = 0.44$, $\eta_p^2 = 0.00$. These results suggest that source credibility significantly impacts response time, with participants taking longer to evaluate misinformation from sources perceived as highly credible. For factual information, there was no effect of source on response time (all p's > 0.05).

5. Discussion

These findings contribute to the literature on the role of source similarity, social identity and source credibility in misinformation susceptibility. First, the results in this chapter show that individuals' political congruence with news sources significantly increases the perceived reliability of misinformation from those sources, aligning with studies that highlight the role of social identity in shaping trust and acceptance of information (Tanis & Postmes, 2005; Traberg & van der Linden, 2022; Van Bavel et al., 2024). Importantly, the results indicate that judgements of misinformation's reliability are not solely driven by familiarity or bias associated with already known sources, as this study replicated these effects with fictitious sources. This suggests that even in the absence of prior exposure, shared identity cues can bias reliability judgements, underscoring the impact of ideological

alignment as a heuristic cue for information judgements. Second, this work also demonstrated that credibility cues, represented by vignette descriptions and credibility metres, also impacted misinformation susceptibility. This aligns with prior studies showing that people utilise credibility cues as indicators of trustworthiness (Petty & Cacioppo, 1986; Flanagin & Metzger, 2007).

Third, an interaction effect emerged, showing that when sources were depicted as lacking journalistic credibility, participants did not exhibit increased susceptibility to misinformation from politically similar sources. This finding suggests that while group identity can enhance trust, it may not be sufficient to override cues indicating low credibility (Birnbaum & Stegner, 1979; Metzger et al., 2003). It further highlights a potential boundary condition in the influence of social identity, where credibility serves as a threshold, beyond which similarity does not enhance perceived reliability. These findings are somewhat encouraging, as they suggest that audiences are attentive to credibility cues, prioritising adherence to journalistic standards over ideological alignment when evaluating information. However, there are also more concerning elements to the findings of this study, which will be discussed below.

Similarity Reinforces Credibility Rather than Compensating for It

The finding that credibility had a stronger effect in politically congruent conditions suggests that similarity might amplify the effects of credibility rather than compensate for its absence. This finding could reflect a form of group validation, where shared identity bolsters trust in a credible source. The alignment of both credibility and similarity may create a stronger perception of reliability, potentially because individuals interpret congruence in identity as an additional layer of trustworthiness (Tormala & Petty, 2004). This indicates that ideological alignment may deepen the impact of sources perceived to be credible.

The results also show that when sources lacked credibility, similarity did not significantly impact reliability judgements. This finding suggests that audiences prioritise credibility over identity when a source is known to be unreliable, highlighting credibility as a fundamental criterion in information evaluation. In other words, similarity alone does not appear to be persuasive enough to override cues indicating low credibility. This aligns with the notion that credibility acts as a baseline filter for trust (Flanagin & Metzger, 2007), and that without it, the persuasive potential of similarity is limited. This finding is particularly relevant in the context of misinformation, as it suggests that establishing clear credibility signals could mitigate the influence of ideologically similar but unreliable sources. However,

applying this in practice is challenging, as there is no universally accepted, politically neutral standard for signalling a source's bias and credibility.

The findings also suggest that when credibility cues align with individuals' political beliefs (a politically similar, credible source), these sources may exert an amplified influence. This "double validation" effect - where similarity potentially reinforces credibility - indicates that even slight biases in perceived credibility can heighten vulnerability to misinformation. While it is not inherently problematic that credible ingroup sources benefit from increased persuasiveness, the real-world implications of this phenomenon warrant concern. First, objective credibility labels remain relatively uncommon in practice, leaving individuals to rely on their own credibility assessments, which, as demonstrated in Chapter 2, are influenced by political ideology. Second, even when credibility labels are present, some research has shown them to have minimal impact on improving the quality of news consumption and that they fail to significantly reduce misperceptions (Aslett et al., 2022). Indeed, Pew Research Center showed that institutions such as fact-checkers - often perceived as a form of credibility arbiter - are viewed by Republicans as liberally biased (Gottfried, 2019). On a more encouraging note, however, recent work has found that credibility measures like fact-checker warning labels can be effective even for individuals who distrust fact-checkers (Martel & Rand, 2024).

Liberals Are Not More Influenced by Social Identity Cues

Findings from the previous chapter suggested that liberals were more influenced by source similarity (political congruence) than conservatives. This chapter aimed to explore whether a similar asymmetry would emerge with fictitious sources, examining whether liberals actually exhibit a greater "bias" in their evaluations. However, the results revealed no interaction effect between political identity and the impact of credibility or similarity. This suggests that neither liberals nor conservatives in this sample relied more on source cues, challenging the notion of a "liberal bias." These findings stand in contrast to Chapter 2, indicating that the asymmetry observed previously may stem from prior familiarity with real-world sources rather than an ideological bias among liberals. Indeed, it is possible that liberals' greater scepticism toward real Republican-slanted news sources, such as Fox News, which has been documented to spread misinformation (Motta et al., 2020), may reflect a justifiable critical stance rather than undue bias.

Additional Insights into Cognitive Reflection, Response Time, and Credibility

The findings on cognitive reflection (CRT) and response time provide additional insights into how individuals process misinformation and engage with source cues. Participants with lower CRT scores were more reliant on source credibility to judge the reliability of headlines, whereas those with higher CRT scores exhibited less reliance on credibility cues. This aligns with the idea that higher cognitive reflection enables individuals to process information more systematically and deliberately (Chaiken & Maheswaran, 1994; Petty & Cacioppo, 1986), thereby reducing the influence of heuristic cues like source credibility.

Interestingly, participants took longer to evaluate misinformation from credible sources. This finding suggests that misinformation from credible sources may induce a form of "cognitive dissonance" or confusion as participants attempt to reconcile the perceived trustworthiness of the source with the misleading nature of the content. The additional processing time may reflect attempts to resolve this tension. These results align with prior research suggesting that credibility cues are often used for quick judgements but may require more cognitive effort to parse when conflicting with content reliability (Petty & Cacioppo, 1986).

Limitations and Future Directions

While this study provides valuable insights into the interplay between credibility and similarity in misinformation susceptibility, several limitations warrant discussion. First, the experiment was conducted in a controlled, simulated environment, which does not capture the complexities of real-world social media platforms. On these platforms, additional social cues - such as comments, likes, shares, and endorsements may also influence misinformation judgements. Second, this study used perceived reliability as a proxy for susceptibility to misinformation. However, perceived reliability does not necessarily translate into behaviours such as sharing or acting on misinformation. Prior research has demonstrated that even information judged as unreliable can influence attitudes and beliefs through mechanisms like repeated exposure (Pennycook et al., 2018).

Third, although the sample was balanced on several demographic variables, it was not fully representative, particularly in terms of geographic and cultural diversity. This limits the generalisability of the findings. Cross-cultural research could provide valuable insights into whether these effects hold across different sociopolitical contexts, where factors like media

literacy, trust in institutions, and exposure to misinformation may vary significantly. Examining variations in cognitive reflection and political ideology across cultural contexts could enrich our understanding of global misinformation dynamics.

6. Conclusion

In conclusion, this study enhances our understanding of how source credibility and similarity interact to influence misinformation susceptibility. While both factors independently shape reliability judgements, credibility appears to play a dominant role. These findings suggest that credibility cues can help counteract ideological biases, offering a potential pathway to reduce misinformation susceptibility, although the practical challenges of implementing this in practice are acknowledged. Importantly, this work highlights the heightened risk of falling for misinformation from ingroup sources, as political similarity can amplify the perceived reliability of such information. These findings underscore the need for interventions explicitly tested under conditions of ideological alignment, where the challenge of combating misinformation is arguably most acute. Designing and evaluating interventions in such contexts will ensure their efficacy in addressing real-world misinformation, particularly in polarised information environments. By disentangling the effects of similarity and credibility, this research contributes to a more nuanced understanding of how individuals evaluate information in polarised media ecosystems. It offers critical insights into the mechanisms underlying the acceptance of misinformation and lays the groundwork for developing strategies to foster critical media consumption and resilience against misinformation.

4. SOCIAL PROOF AND PERCEIVED SOCIAL CONSENSUS

This chapter features Study 2 from the comprehensive paper published in *Nature Scientific Reports* in the article below:

Traberg, **C. S**., Harjani, T., Roozenbeek, J., & van der Linden, S. (2024). The persuasive effects of social cues and source effects on misinformation susceptibility. *Nature Scientific Reports, 14*, 4205. https://doi.org/10.1038/s41598-024-54030-y

1. Abstract

Research has demonstrated that exposure to others' judgements can influence perceptions and attitudes. Despite this, much prior work exploring the psychological underpinnings of misinformation susceptibility has been conducted in isolation from the social contexts in which news consumption typically occurs. This chapter extends this work by examining the role of social cues in shaping susceptibility to misinformation. It specifically investigates how implicit and explicit social cues affect the perceived reliability of misinformation and explores the potential mediating role of perceived social consensus. Across five experimental conditions (N = 730), the findings reveal that explicit endorsement cues significantly increased susceptibility to misinformation, whereas discrediting cues had no effect. The results indicate that the impact of social cues on misinformation susceptibility was mediated via perceived social consensus. This work advances our understanding of how individuals assess misinformation veracity in social contexts and the chapter concludes by discussing the broader implications for combating misinformation in social media environments.

2. Introduction

From a social psychological perspective, both theoretical and empirical evidence suggests that our judgements, beliefs, and opinions are fundamentally shaped by social life rather than formed in isolation (Gottfried, 2019). Understanding social influence is particularly relevant in today's digital landscape, where individuals encounter news and information within social contexts on social media platforms. Social media platforms have become central to news dissemination, with over half of U.S. adults obtaining at least some news from these sources (Anderson, 2021; Atske, 2021). This shift has fundamentally changed how individuals encounter information. Unlike traditional media, where audiences consume news independently, social media environments are saturated with social cues, where users interact and share their opinions in easily observable ways (Lee et al., 2021). Metrics such as likes, shares, and comments on news items provide visible and immediate signals of others' reactions, which may influence users' evaluations (Metzger et al., 2010; Tandoc Jr et al., 2017). Therefore, recent years have been marked by an increased focus on examining misinformation on social media platforms that have significantly accelerated its spread (Del Vicario et al., 2016).

Social Proof and Its Impact on Attitudes

Within social influence research, the concept of social proof - the phenomenon of people relying on the judgements of others to inform their own beliefs - is well-documented as a factor in shaping attitudes and behaviours (Cialdini & Goldstein, 2004). On social media, visible engagement metrics may function as a form of 'social proof' (Traberg et al., 2024), potentially influencing users' judgements about the reliability of information. However, research on how social proof impacts the perceived reliability of misinformation remains mixed, with some studies indicating these cues impact the perceived credibility of information (Luo et al., 2020), while others report no significant impact (Mena et al., 2020; Traberg et al., 2024). This chapter addresses the question: Do the implicit and explicit judgements of others impact news consumers' perceptions of misinformation reliability? Additionally, it explores the cognitive mechanisms by which social cues may affect misinformation susceptibility.

Judgements as Social Phenomena

The idea that social influences shape our judgements and beliefs is not new. In fact, it was central to foundational theories on attitude validity in classic psychological research

(Sherif & Hovland, 1961), with research demonstrating that our perceptions of the attitudes of others impact our own attitude certainty (Tormala et al., 2009). Social consensus information - information about other people's beliefs - may validate one's beliefs (Gardikiotis et al., 2005), enhancing confidence in those beliefs as perceived agreement with others increases (Petrocelli, Tormala, & Rucker, 2007). Demonstrating the powerful role of social consensusinformation, Kobayashi (2018) found that social consensus had an independent and equal impact on scientific beliefs as scientific consensus. In other words, the beliefs of our peers may carry as much weight as expert opinions, underscoring the powerful role of social information in shaping beliefs. Social consensus has also been found to mediate attitudes toward various topics, including climate change (Lewandowsky et al., 2019), political issues (Guess et al., 2019) as well as intergroup attitudes (Sechrist & Young, 2011; Stangor et al., 2001; Wittenbrink & Henly, 1996) and behaviours (Sechrist & Stangor, 2001).

Kassin (1979) and Hewstone & Jaspars (1988) have suggested that both *explicit* (actual behaviour of others) and *implicit* (what others might do if present) forms of consensus shape beliefs and decision-making. This consensus may manifest on social media through specific social cues that could signal collective opinion or endorsement. As social media sites represent ecosystems of social information where users can explicitly indicate agreement with or interest in information, express their opinions regarding news, and share information with other users (Hilverda et al., 2018), these forms of social proof (Cialdini, 1993) may signal to other users that interest or support for the given information is high. As such, forming judgements about information while simultaneously being exposed to the opinions of others could lead individuals to follow the crowd due to group influence (Deutsch & Gerard, 1955) either if they assume the group holds better information than themselves (informational influence), or if they conform to the group to gain social approval (normative influence).

Scholars in fields beyond social psychology have noted the impact of online social cues on judgements. Research in e-commerce, for example, has found that peer comments and ratings influence readers' evaluations, with people often relying on others' assessments to guide their own purchasing decisions (Walther et al., 2012). Similarly, individuals tend to follow the crowd when selecting movies to watch based on ratings and popularity (Xu & Fu, 2014). Social influence on digital platforms has been linked to herding effects, where people unconsciously follow group behaviour or conform to avoid disagreement (Muchnik et al., 2013). For instance, Muscanell et al. (2014) found that people were more likely to fall for online scams when scammers manipulated the number of social media likes. Muchnik et al. (2013) found that viewing previous ratings on a social news aggregation website biased

individual ratings, dependent on the topic, and influenced by whether or not the previous ratings were from friends or enemies (Muchnik et al., 2013). Examining the effect of online social proof on perceptions of organic food, Hilverda et al. (2018) found that Facebook 'likes' did not have an effect on benefits perception and motivation to find information. However, likes affected consumers' reactions, specifically their negative emotions and willingness to pay.

Mixed Evidence for Social Cues and Misinformation

Results are mixed in relation to the influence of social cues on misinformation susceptibility. Some research has suggested that quantitative social media cues may play a role in susceptibility to health and science misinformation (Luo et al., 2020), and some find that misinformation accompanied by high social engagement can be judged to be more credible (Butler et al., 2023). However, other studies have shown no impact of these cues on credibility judgements (Ali et al., 2022; Mena et al., 2020; Traberg et al., 2024), raising important questions about how social cues affect judgement formation. In addition, Kluck et al. (2019) found that while negative comments reduced the believability of misinformation, credibility ratings provided prior to exposure had little effect on subsequent perceptions.

On top of these mixed results, research has yet to uncover *why* social cues influence or do not influence misinformation susceptibility. Some researchers have speculated that quantitative cues such as 'likes' may lack a negative social interpretation (Ali et al., 2022) and, therefore, do not serve as any form of meaningful representation of perceived social consensus. Without a mechanism to signal disapproval, cues such as 'likes' may not be perceived as indicative of social consensus (Kang et al., 2011). Empirical findings further suggest this may be the case, as Traberg et al. (2024) showed that high vs low engagement numbers ('likes') did not impact perceptions of social consensus in the reliability of misinformation. Even less is known about which social cues *do,* in fact, influence perceived social consensus and whether social consensus perceptions impact individual susceptibility to misinformation.

This chapter investigates whether individuals are influenced by both *implicit* and *explicit* group consensus information. Implicit cues, such as comments reflecting endorsement or concern about a headline, may subtly signal the content's veracity without explicitly stating it. Explicit cues, such as prior judgements from other users, offer direct information about the opinions of a group (Gardikiotis et al., 2005). Additionally, the chapter

examines whether these localised group cues affect perceptions of broader social consensus and how those perceptions influence individual-level judgements.

The Current Study

This chapter builds on previous research that has examined the effects of social cues on misinformation susceptibility. Earlier work by Traberg (2019) - as reported in Traberg et al. (2024), showed that social cues in the form of 'likes' did not significantly influence the perceived reliability of misinformation or perceived social consensus. In these studies, misinformation headlines were presented with varying levels of 'likes' (high, low, or none) to explore whether engagement metrics alone shaped users' judgements about the reliability of news headlines. The results suggested that mere numerical indicators of popularity, isolated from other contexts or content, may not impact perceptions of misinformation reliability or consensus. However, this stands in contrast to other research, which found that such metrics influence judgements of news headlines (Luo et al., 2020). Building on these mixed findings, the current study extends this line of inquiry by examining more nuanced and contextually rich forms of social cues. Specifically, it explores *implicit* social cues, such as comments expressing concern or endorsing a headline, and *explicit* social consensus cues, where previous users' reliability judgements are directly shown. The primary goal is to determine whether these cues influence susceptibility to misinformation and, if so, whether this influence occurs through shaping perceptions of social consensus.

By exploring these questions, the study aims to clarify how different forms of social information affect individual judgements and to identify the cognitive processes that may underlie these effects. Hypotheses were pre-registered prior to data collection and analysis (https://doi.org/10.17605/OSF.IO/4YN7Z). This work contributes to our understanding of the specific mechanisms through which social cues may influence susceptibility to misinformation, with implications for both theoretical frameworks in social influence and practical strategies to mitigate misinformation on social media platforms. The chapter examines the specific hypotheses outlined below.

First, as a test of the impact of social cues on perceptions of misinformation reliability, the following hypothesis was generated:

H1: Reliability judgements of misinformation will be higher in conditions where implicit or explicit cues endorse validity compared to conditions where they discredit validity.

Second, to explore whether perceptions of information reliability differ from judgements of its truth value, we generated the following hypothesis:

H2: Truth judgements of misinformation will be higher in conditions where implicit or explicit cues endorse validity, compared to conditions where they discredit validity.

Third, to examine the impact of social cues on perceived consensus (here, regarding the perceived reliability of misinformation), we test the following hypothesis:

H3: Perceived public reliability will be higher in implicit and explicit endorsement conditions than in corresponding discrediting conditions.

To examine the impact of social cues on perceived consensus (here, regarding the perceived truth value in misinformation), we test the following hypothesis:

H4: Perceived public truth will be higher in implicit and explicit endorsement conditions than in corresponding discrediting conditions.

To test the correlation between social consensus perceptions and individual judgements we test the following two hypotheses:

H5: Higher perceived public reliability will predict higher individual reliability judgements of misinformation.

H6: Higher perceived public truth will predict higher individual truth judgements of misinformation.

Finally, to test whether the effect of social cues on individual judgements is mediated via perceived consensus, we generated the final hypothesis:

H7: The effect of social cue valence (endorsement vs. discrediting) on reliability judgements of misinformation will be mediated by perceived public reliability.

3. Method

Purpose

The purpose of this study was to assess whether explicit and implicit social cues influence the perceived reliability of misinformation and whether these cues influence perceptions of social consensus regarding the reliability of misinformation headlines.

Participants

Participants were recruited on Prolific and invited to participate in a study on "News Perception". A requirement for participating was being fluent in English, but participants worldwide could partake. For this study, a power analysis indicated $N = 470$ participants would be sufficient to detect a small effect size of $d = 0.40$ with five groups, an alpha level of .05 and 95% power. $N = 730$ participants (46% female, $M_{age} = 27$, 53% university educated, 64% left-leaning, 94% had a Facebook account, 46% got majority of news from social media and 32% from online news sites) were recruited across five conditions: Control condition ($n = 148$), Implicit discrediting ($n = 146$), Implicit endorsement ($n = 146$), Explicit discrediting ($n = 145$) and Explicit endorsement ($n = 145$). As specified in the pre-registration, we analysed results with the full sample and with only those who passed both attention checks ($N = 542$). All participants provided informed consent and were compensated for their time. The study was approved by the Cambridge Psychology Research Ethics Committee (PRE.2020.059). All methods were performed in accordance with the relevant guidelines and regulations.

Design and Materials

Social Cue Manipulation

Social cues were manipulated across the following 5 conditions. First, the social cue format was manipulated: *explicit* vs *implicit* and the direction was manipulated: *endorsement* vs *discrediting*. A control condition with no social cues was also included. To manipulate *explicit* consensus, participants were explicitly informed about the judgements of an unidentified previous group of 'other' participants. This involved showing participants a statement before exposure to each headline that read: "X percent of previous participants judged the following headline to be reliable (*endorsement* condition)/unreliable (*discrediting* condition)". This percentage was always above 65% to signal majority agreement. To

manipulate *implicit* consensus, participants were exposed to five comments underneath each headline, which all either implied the individual commenter believed (*endorsement* condition) or did not believe (*discrediting* condition) the headlines (see Figure 21)

Figure 21

Example of Misinformation Headline in Implicit Endorsement Condition (Left) and Implicit Discrediting Condition (Right)

Headlines

Participants were exposed to three misinformation headlines and one control (factual) headline. The misinformation used relied on deceptive techniques rather than being explicit "fake news". As in previous studies in this thesis, misinformation items were identified using the misinformation platform Hoaxy. We included items that had been published in the last year and made use of the emotional language manipulation technique. This time, we used a Facebook layout instead of the Twitter one used previously. An overview of all items used in this study is found below in Table 4.

Table 4

Overview of Headlines used in the Study

Misinformation vs Facts	Category	Content	Source Name
Misinformation	Emotion	Top headline: "Baby formula linked to horrific disease outbreak." Bottom headline: "Baby formula linked to horrific outbreak of new, terrifying disease among helpless infants"	Hidden
Misinformation	Emotion	Top headline: "Make sure these aren't in your cupboard!" Bottom headline: "Shocking finding: Common vitamin supplements are strongly linked to cancer"	Hidden
Misinformation	Emotion	Top headline: "If you own one, return it IMMEDIATELY. Your safety is at risk." Bottom headline: "WARNING: smartphone (x) causes explosion, leaving elderly pensioner to fight for her life."	Hidden
Factual Information	-	Top headline: "Biden returns the US to the Paris Agreement" Bottom headline: "The US officially rejoins the Paris climate agreement."	Hidden

Measures

Perceived Reliability of Misinformation

Participants were asked to rate each item's reliability on a standard 7-point Likert scale: *"On a scale from 1 to 7, how reliable do you find this headline"* (1 = Very Unreliable, 7 = Very Reliable). Participants judged the reliability of three misinformation headlines, which were averaged ($M = 2.46$, $SD = 1.48$, $\alpha = 0.67$)

Perceived Truthfulness in Misinformation

Participants were asked to rate each item's likelihood of being true on a standard 7-point Likert scale: *"On a scale from 1 to 7, how likely is it that the above information is true?"* (1= Very Unreliable, 7 = Very Reliable). Participants judged the truthfulness of three misinformation headlines which were averaged ($M = 2.70$, $SD = 1.55$, $\alpha = 0.78$).

Perceived Public Consensus in Reliability

Participants were asked to indicate what percentage of the general population they believed would find the headline reliable: *"What percentage of the general population do you think would find this tweet reliable?"* (0-100%). Participants made this judgement for three headlines, which were averaged (M = 52.06, SD = 24.08, α = 0.78).

Perceived Public Consensus in Truth

Participants were asked to indicate what percentage of the general population they believed would find the headline true: *"What percentage of the general population do you think would judge the information in the headline to be true?"* (0-100%). Participants made this judgement for three headlines, which were averaged (M = 53.04, SD = 23.94, α = 0.78).

Intention to Share Misinformation

Participants were asked to indicate how likely they would be to share each headline on social media on a 7-point Likert scale: *"On a scale from 1 to 7, how likely is it that you would share this with your network if it came up on your newsfeed?"* (1 = Very Unlikely, 7 = Very Likely). Participants made this judgement for three headlines, which were averaged (M = 1.70, SD = 1.32, α = 0.79).

Demographics

Socio-demographic variables included gender (male [53.43%], female [45.89%], other [0.69%]), age (M_{age} = 27.29, SD = 7.91), political orientation (measured on a 7-point Likert scale, where 1 = Very Left-wing and 7 = Very Right-wing; M = 3.21, SD = 1.26), highest level of education completed (less than high school degree (4.66%), high school graduate (41.78%), Bachelor's degree (31.10%), Master's degree (19.18%), Doctoral degree (1.78%), or Professional degree (1.51%)), current use of Facebook ("I don't have an account" (5.62%); "I have an account but I hardly ever use it" (21.37%); "I have an account, and I use it occasionally" (30.00%); "I have an account and I use it often" (16.44%); "I have an account and I use it on a daily basis" (26.58%)), and news consumption source ("I don't really follow the news" (4.66%); "Social Media" (46.44%); "TV and radio" (14.52%); "Print Media" (newspapers, magazines) (1.23%); "Word of Mouth" (1.51%); and "Online news sites (excluding social media)" (31.64%)).

Attention check

At the end of the study, all participants were asked to state whether the social cues they were exposed to indicated that other individuals found the information reliable or unreliable: "*Did the information you saw about the responses of other individuals indicate that the majority judged the headlines to be reliable or unreliable?"* with response options being "Reliable", "Unreliable", or "I didn't see any such information". 540 out of 730 participants passed the attention check. However, given the large number of participants failing the attention check, we report results for both whenever they differ.

Procedure

Participants were recruited via Prolific, where the study was advertised as a study on "News Perception". Participants were redirected to the Qualtrics platform, where the study was hosted, and provided informed consent prior to starting. Then, participants answered the demographic questions. Following this, participants were randomly assigned to one of the five conditions. They then proceeded to answer the questions for each headline, which were presented in random order. Finally, participants were debriefed.

4. Results

As pre-registered, analyses were run with both the full sample and excluding participants who did not pass the attention check. In general, the results reported in this section are for the full sample and results for the sample after exclusions were applied can be found in Appendix B. However, wherever there was any discrepancy in statistical significance between the samples, both results are reported directly in this chapter.

H1: Reliability judgements of misinformation will be higher in conditions where implicit or explicit cues endorse validity compared to conditions where they discredit validity.

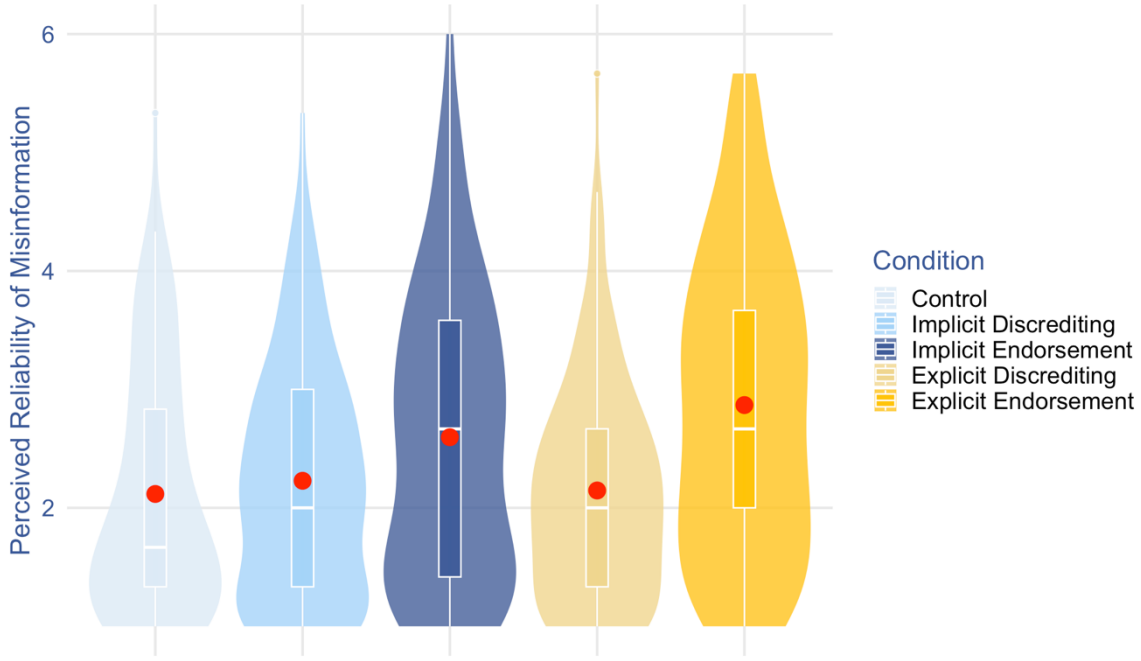
Analyses run with the full sample ($F(4,725) = 6.35$, $p < 0$ .0001, $\eta_p^2 = 0.3$) show a significant main effect of social cues on perceived reliability. Planned contrasts showed that participants judged misleading headlines as more reliable when previous participant judgements endorsed the misinformation ($M = 2.83$, $SD = 1.23$) compared to when they discredited it ($M = 2.23$, $SD = 1.01$, $p < 0.001$, $d = 0.53$) (the 'explicit' conditions). Participants also judged misleading headlines as more reliable when comments underneath endorsed the information ($M = 2.48$, $SD = 1.13$) compared to when they discredited it ($M =$

2.30, *SD* = 1.10). However, this contrast was not significant in analyses run with the full sample ($p = 0.18$, $d = 0.16$), but was significant when excluding those who did not pass the attention check ($M_{endorse} = 2.60$, $SD_{endorse} = 1.22$, $M_{discredit} = 2.23$, $SD_{discredit} = 0.99$, $p = 0.01$, $d = 0.34$).

Post-hoc tests revealed that there were no significant differences between both discrediting and the control conditions (p's > 0.05), no significant differences between the implicit endorsement and control ($p = 0.81$), but a significant difference between the explicit endorsement ($M = 2.83$, $SD = 1.23$) and control condition ($M = 2.34$, $SD = 1.20$, $p < 0.01$, $d = 0.43$). However, after exclusions, there was a significant difference between the implicit endorsement and control ($p = 0.03$, $d = 0.43$). As such, discrediting cues did not affect the perceived reliability of misinformation, whereas endorsement cues increased the perceived reliability of misinformation. Figure 22 shows the influence of social cue conditions. A complete list of results for analyses run after excluding participants who failed the attention check can be found in Appendix B.

Figure 22

Mean Perceived Reliability of Misinformation Headlines by Social Cue Condition



Note: This visualisation was created using the sample excluding attention check fails

H2: Truth judgements of misinformation will be higher in conditions where implicit or explicit cues endorse validity, compared to conditions where they discredit validity.

Results from the ANOVA with the full sample showed a significant effect of social cues on perceived truth in misinformation headlines ($F(4,725) = 5.62$, $p < 0.001$, $\eta_p^2 = 0.3$. Planned contrasts showed a significant difference between the explicit discrediting ($M = 2.44$, $SE = 0.86$) and endorsement conditions ($M = 3.07$, $SE = 0.11$, $p < 0.001$, $d = -0.52$), but no difference between the implicit endorsement ($M = 2.60$, $SE = 0.06$) vs discrediting ($M = 2.72$, $SE = 0.09$, $d = 0.11$, $p = 0.39$). Exploratory post-hoc tests looking at the control condition showed a significant difference between the explicit endorsement and control ($M_{diff} = 0.41$, $SE = 0.14$, $p = 0.02$, $d = 0.35$) and between the explicit endorsement and implicit discrediting ($M_{diff} = -0.47$, $SE = 0.14$, $p < 0.01$, $d = -0.40$), but no further significant differences.

Results from the ANOVA with only participants who passed the attention check showed a significant effect of social cues on perceived truth in misinformation headlines ($F(4,537) = 27.614$, $p < 0.001$, $\eta_p^2 = 0.6$). Planned contrasts showed a significant difference between the explicit discrediting ($M = 2.34$, $SE = 0.09$) and endorsement conditions ($M = 3.12$, $SE = 0.12$, $p < 0.001$, $d = -0.66$), but no significant difference between the implicit endorsement ($M = 2.52$, $SE = 0.10$) vs discrediting ($M = 2.83$, $SE = 0.12$, $d = 0.27$, $p = 0.06$). Turning to the control condition, post-hoc tests showed a significant difference between the explicit endorsement and control ($M_{diff} = -0.62$, $SE = 0.16$, $p < 0.01$, $d = -0.54$) but no significant differences between the other conditions and control (p's > 0.05)

H3: Perceived public reliability will be higher in implicit and explicit endorsement conditions than in corresponding discrediting conditions.

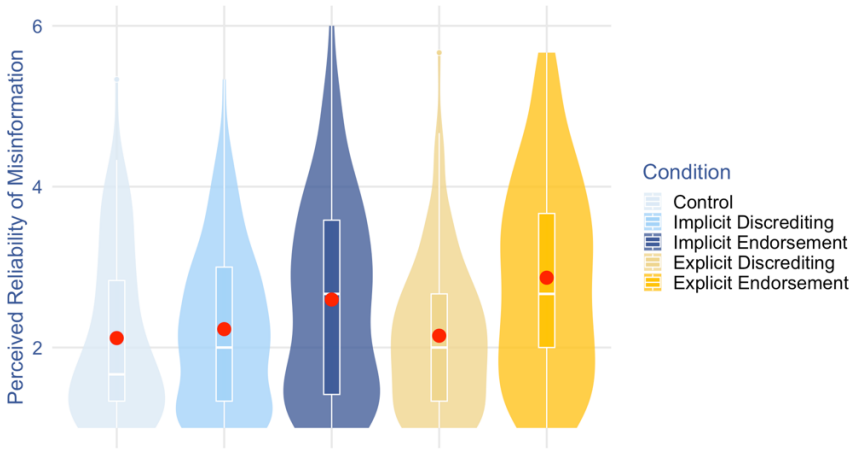
An ANOVA with social cue condition as the independent variable and perceived public reliability was significant ($F(4,725) = 51.84$, $p < 0.001$, $\eta_p^2 = 0.22$). As specified in the pre-registration, we ran two contrasts. We firstly compared the implicit endorsement and discrediting conditions where perceived consensus judgements were significantly higher in the endorsement ($M = 54.82$, $SD = 19.14$) compared to the discrediting conditions ($M = 44.77$, $SD = 19.19$, $p < 0.001$, $d = 0.52$). We subsequently compared the explicit endorsement and discrediting conditions, where perceived consensus was also significantly higher in the endorsement condition ($M = 61.22$, $SD = 16.93$) compared to the discrediting condition ($M = 32.38$, $SD = 14.77$, $p < 0.001$, $d = 1.81$).

Exploratory post-hoc tests show that both explicit discrediting ($M_{diff} = -17.04$, SE = 2.13, $p < 0.001$, $d = -0.93$) and endorsement ($M_{diff} = 11.80$, $SE = 2.13$, $p < 0.001$, $d = 0.65$) were significantly different to the control. However, there were no differences between the implicit conditions and the control (p's > 0.05). However, when excluding participants who

failed the attention check, there was a significant difference between the implicit endorsement condition and the control (M_{diff} = 8.56, p = 0.01, d = -0.43). These results show support for H2. Both implicit and explicit cues influenced perceived consensus in the reliability of misinformation. This result is visualised in Figure 23.

Figure 23

Mean Perceived Public Reliability of Misinformation by Social Cue Condition



Note: This visualisation was created using the sample excluding attention check fails

H4: Perceived public truth will be higher in implicit and explicit endorsement conditions than in corresponding discrediting conditions.

For perceived public truth, the ANOVA showed a significant effect of social cue condition (F(4,725) = 54.31, p < 0.001, η_p^2 = 0.23). Planned contrasts showed a significant difference between the implicit discrediting (M = 45.96, SE = 1.52) and endorsement conditions (M= 56.44, SE= 1.56, p < 0.001, d = -0.56) and between the explicit endorsement (M = 61.62, SE = 1.48) and discrediting conditions (M = 32.52, SE = 1.25, p < 0.001, d = 1.76). In relation to the control, post-hoc tests showed a significant difference between the explicit discrediting and control condition (M_{diff} = -18.17, SE = 2.13, p < 0.001, d = -1.00), between the explicit endorsement and control condition (M_{diff} = 10.17, SE = 2.13, p < 0.001, d = 0.60) and the explicit endorsement and discrediting conditions (M_{diff} = 29.10, SE = 2.14, p < 0.001, d = 1.60) but not between the implicit discrediting and control conditions (p > 0.05).

H5: Higher perceived public reliability will predict higher individual reliability judgements of misinformation.

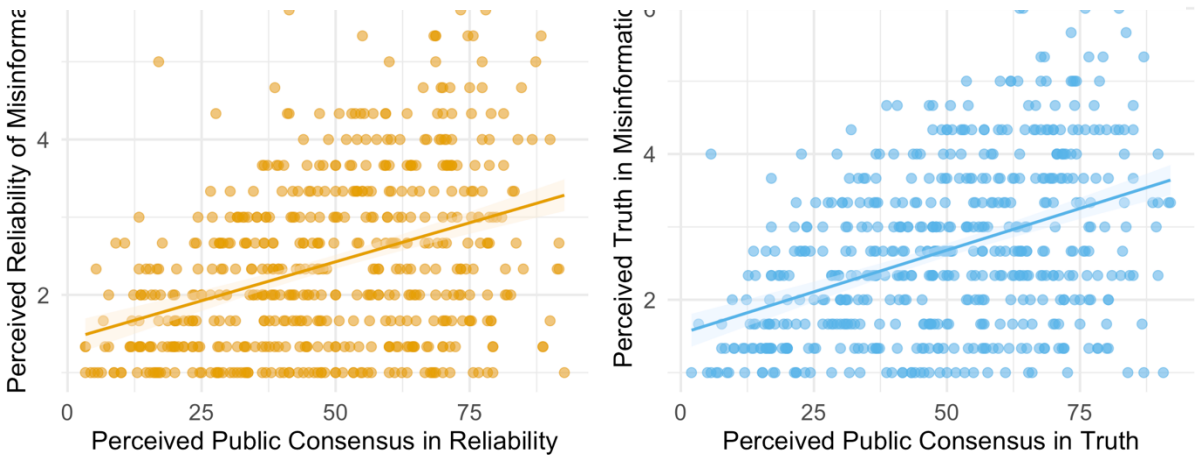
To test whether perceived consensus predicts perceived reliability of misinformation, a linear regression analysis was run, using perceived reliability of misinformation as the DV and perceived consensus as the IV. The regression was significant (Adjusted $R^2 = 0.14$, $F(1,728) = 118.2$, $p < 0.001$), with perceived consensus predicting perceived reliability ($b = 0.02$, $p < 0.001$). Results with exclusions applied are the same (Adjusted $R^2 = 0.14$, $F(1,540) = 89.92$, $p < 0.001$). This suggests that when individuals perceive a majority of others to believe misinformation headlines, they also judge them as reliable. Our results, therefore, support H5.

H6: Higher perceived public truth will predict higher individual truth judgements of misinformation.

A linear regression with perceived truth as the dependent variable and perceived public consensus in the truth of misinformation showed the regression was significant (Adjusted $R^2 = 0.15$, $F(1,728) = 131$, $p < 0.001$), with perceived consensus predicting perceived reliability ($b = 0.02$, $p < 0.001$). Results with exclusions applied are the same (Adjusted $R^2 = 0.16$, $b = 0.02$, $F(1,540) = 107.9$, $p < 0.001$). This suggests that when individuals perceive a majority of others to believe misinformation headlines, they also judge them as reliable. Our results, therefore, support H5. Results for H5 and H6 are visualised in Figure 24.

Figure 24

Perceived Reliability of Misinformation by Perceived Public Consensus (Left) and Perceived Truth in Misinformation by Perceived Public Consensus



H7: The effect of social cue valence (endorsement vs. discrediting) on reliability judgements of misinformation will be mediated by perceived public reliability.

To test perceived consensus as a mediator of the effect of explicit social cues on the perceived reliability of misinformation, a mediation analysis was run with 10,000 bootstrapped confidence intervals in JASP using the SEM module. Here, the direct effect of social cue condition on perceived reliability was not significant with unstandardised regression coefficient b = -0.03, p = 0.85, 95% CI [-0.34, 0.27]. However, the indirect effect of social cue condition via perceived public consensus was significant, as the unstandardised regression coefficient for the mean bootstrapped indirect effect was b = 0.63, p < 0.001, 95% CI [0.43,0.86]. The unstandardised regression coefficient for the total effect of social cue condition on perceived reliability was b = 0.60, p < 0.001, 95% CI [0.33,0.85]. The effect of previous participant judgement social conditions on perceived reliability was thus fully mediated via perceived consensus.

An equivalent mediation analysis was run to test perceived consensus as a mediator of the effect of implicit social cues on the perceived reliability of misinformation. Similar as to with explicit social cues, the direct effect of social cues on the perceived reliability of misinformation was not statistically significant (b = -0.03, p = 0.80, 95% CI[−0.29,0.23], but the indirect effect of the mediator of perceived consensus was (b = 0.21, p < 0.001, 95% CI[0.12,0.34]). This suggests that the observed relationship between social cues and perceived reliability was fully accounted for by the mediator. The total effect, however, was not significant (b = 0.18, p = 0.16, 95% CI[−0.07,0.44]), suggesting a suppressed mediation. In suppressed mediation, the direct impact of social cues on perceived reliability is counteracted or suppressed by the mediating variable, perceived consensus (Jiang et al., 2021). The non-significant total effect thereby stems from the opposing signs of the direct (negative) and indirect (positive) paths, leading to their effects offsetting each other. Importantly, this does not discount the observed significant competitive mediation, emphasising the need for a more nuanced consideration when understanding the impact of implicit social cues on perceived reliability. Analyses conducted with exclusions applied showed the same pattern of results, albeit with a significant total effect (p = 0.02). As such, results support H7.

Together, the mediation models suggest that both implicit cues, such as comments, and explicit cues in the form of previous participant judgements indicating the beliefs of a local group influence perceptions of what the public believes, and this perception influences the perceived reliability of misinformation. We ran correlation analyses to ensure that the full

mediation findings were not statistical artefacts and results of strong correlations between the mediator and IV or DV. These analyses, reported in Appendix C, revealed significant but not excessive correlations between the examined variables, indicating that the mediation findings are likely robust and reflect meaningful psychological processes rather than being confounded by multicollinearity.

The Impact of Social Cues on Perceived Reliability of Factual Headlines

Following the main hypotheses, we further analysed the impact of social cues on the perceived reliability of factual headlines. Results show that social cues play a role to some extent. The social cues model was significant ($F(4,725) = 11.01$, $p < 0.001$, $\eta_p^2 = 0.06$) with significant differences between the implicit discrediting and endorsement ($M_{diff} = -0.98$, $SE = 0.19$, $p < 0.001$, $d = -0.62$) and between the control and implicit discrediting ($M_{diff} = 0.85$, $SE = 0.18$, $p < 0.001$, $d = 0.54$). Interestingly, the contrast between explicit endorsement and explicit discrediting was not significant ($p = 0.08$). Looking at perceived consensus in factual headlines, a similar pattern to misinformation emerged. The primary model was significant ($F(4,725) = 17.3$, $p < 0.001$, $\eta_p^2 = 0.09$), with a significant difference observed between the control and the implicit discrediting conditions ($M_{diff} = 13.23$, $SE = 2.22$, $p < 0.001$, $d = 0.69$), between explicit discrediting and endorsement ($M_{diff} = -9.13$, $SE = 2.24$, $p < 0.001$, $d = -0.48$), and between implicit discrediting and endorsement ($M_{diff} = -13.23$, $SE = 2.23$, $p < 0.001$, $d = -0.69$).

5. Discussion

This work provides novel insight into the link between individual cognition and social context and highlights how socio-cognitive factors can increase individual susceptibility to misinformation. This work highlights that social cues can affect judgements of misinformation and, importantly, identifies a potential mechanism through which it exerts its influence.

The results provide evidence that social cues influence misinformation susceptibility. Results show that explicit cues indicating the judgements of others influenced misinformation susceptibility. Implicit cues endorsing the reliability of the headline also influenced misinformation susceptibility, but only when excluding participants who failed the attention check. It may be that those who failed the attention check were not paying attention to the social cue information, particularly in the implicit conditions that required reading several

comments. This discrepancy, however, further cements the point that social cues may *only* influence misinformation susceptibility if these cues are attended to and influence wider social consensus perceptions. This echoes work by Cialdini & colleagues, which highlights the role of attention and salience in activating social norms, suggesting that social cues—whether explicit or implicit - need to be noticed to influence behaviour, as norms only guide action when they are focused upon and made salient (Cialdini et al., 1991).

It is, in this regard, important to distinguish between social consensus and social norms. Social norms refer to behaviours or beliefs perceived as acceptable or typical within a specific group and often carry an implicit expectation to conform (Bicchieri, 2006; Cialdini & Trost, 1998). In contrast, social consensus reflects a broader perceived agreement across a group or society, focusing less on behavioural conformity and more on informational influence (Prislin & Wood, 2005). While norms often create pressure to align one's actions with group standards, social consensus affects judgements by shaping perceptions of truth and reliability based on the inferred beliefs of others (Hovland & Weiss, 1951). Moreover, social consensus can operate without the direct interpersonal dynamics that characterise social norms. For example, exposure to aggregated judgements (e.g. previous participant judgements) may lead individuals to infer widespread agreement without feeling personally obligated to conform to the majority view (Weimann, 1994).

However, it remains unclear whether participants in this study paid limited attention to the comments beneath the headlines, due to the artificial nature of the experimental setting - perhaps aiming to complete the task quickly - or if this reflects typical engagement with such cues on social media. This ambiguity highlights the need for further research to assess how individuals process social cues in real-world online environments. Nevertheless, the results indicate that for those who do engage with these cues, they exert a notable influence not only on judgements but also on perceptions of broader social consensus.

While the correlational mediation results are suggestive rather than definitive, they raise the possibility that social cues are more likely to influence misinformation susceptibility when they shape individuals' perceptions of broader social consensus. This aligns with prior findings by Traberg (2019) (reported in Traberg et al., 2024), which demonstrated that 'likes' did not affect perceptions of misinformation or social consensus. These findings suggest that exposure to the judgements of others only increases individual susceptibility to misinformation if those judgements signal reliability and influence perceptions of a wider social consensus. The weight that news consumers assign to social cues may depend not only

on the nature of the cues themselves but also on individual differences among news consumers in how they interpret and integrate such cues into their evaluations.

This work also provides a potential explanation for why engagement cues such as *'likes'* may not always impact susceptibility: they may fail to influence perceptions of broader social consensus. The findings indicate that perceived consensus predicts susceptibility to misinformation and mediates the relationship between social cues and individual judgements. Specifically, exposure to social cues that signal the judgements of others led to increased perceptions of wider social consensus, which in turn predicted individual susceptibility to misinformation. While the mediation model does not provide causal evidence for the relationship between perceived consensus and susceptibility (e.g., see Coenen, 2022), it suggests a potential mechanism underlying the persuasive effects of social cues. Exposure to social cues from a limited group (e.g., prior participants in a study) may lead individuals to infer the judgements of a much larger group (e.g., the public), which could then influence their own evaluations. An alternative explanation, however, is that individuals who already perceive misinformation as reliable may, through a false consensus effect, also perceive a greater level of public consensus. Future research could aim to disentangle these two explanations to further clarify the role of perceived consensus in susceptibility to misinformation.

These findings suggest that when social media users are exposed to social cues indicating that others believe misinformation, this leaves them more vulnerable to misjudging the information as reliable. While following the majority can be a reliable heuristic - statistically speaking, the majority is often more likely to be correct (Budescu & Chen, 2015) - this "wisdom of crowd" effect only holds under specific conditions, such as when judgements are made independently, and group members are diverse (Lorenz et al., 2011; Surowiecki, 2004). However, this heuristic on social media may become an unreliable cognitive bias when individuals extrapolate information from a small selection of news consumers. This may occur because such settings violate the key assumptions of "smart crowds," as judgements on social media are neither independent nor likely to be diverse. Given that users primarily interact with and are exposed to information from like-minded others within so-called 'echo chambers' (Cinelli et al., 2021), this could undermine the wisdom-of-the-crowd effect by reducing the diversity of viewpoints and creating "misinformation bubbles" (Rhodes, 2022), leaving some social media users more predisposed to being influenced by misinformation. This further suggests that individuals can contribute

to the spread and influence of misinformation on other news consumers, as their visible judgements may, in turn, shape the judgements of others.

Interestingly, when comparing the results for misinformation judgements to those of factual headlines, results showed that discrediting reduced perceived reliability in relation to the control condition – and only for *implicit* cues rather than *explicit* cues. This suggests that individuals openly questioning facts on the internet may lead other news consumers to question valid news and reporting. In contrast, social cues discrediting *mis*information did not seem to sway judgements. One explanation for this result is that social proof, or the influence of others, may be particularly potent when it challenges preexisting beliefs or when there is a discrepancy between one's own judgement and the perceived judgement of a social group. In other words, news consumers may already have a good sense of what is true and false – until they are exposed to social cues that challenge their own assessments.

6. Limitations

While this chapter explored the broader question of whether and how social cues influence susceptibility to misinformation, future research could build on this work by addressing several limitations and exploring new directions. Firstly, to assess causality in the direction of the relationship between social cues and misinformation susceptibility, this study was conducted using a simulated social media environment and not on a real social media platform where social processes are more complex. The social context of information consumption is multifaceted, and there are many unexplored social psychological variables that may also be at play. For instance, it may be the case that group identity processes interact with social consensus cues, making ingroup majorities more persuasive than outgroup majorities. Future work may seek to explore these potential compounding effects. Indeed, although this work shows that perceived social consensus mediated the relation between social cues and misinformation susceptibility, other mediators may play a role in this process that this chapter did not explore.

This chapter used the perceived reliability of misinformation to study misinformation susceptibility – but the dangers of misinformation do not stop at reliability judgements. It is possible that repeated exposure to misinformation that is judged as unreliable still influences beliefs (e.g., about science, politics, or health) or behaviour. Finally, although the samples were balanced on various demographic variables and were not specific to any population, they were also not *representative* of any country's population. A requirement for participation

was being English-speaking, but as English may not have been the participants' first language, as the comments were written in English, this may have impacted their influence effects. Despite these limitations, this research contributes to the long-standing debate on the influence of social biases. The results provide evidence social context plays a significant role when individuals make judgements about the reliability of misinformation. Given the results, this chapter highlights that it is not sufficient to study cognitive processes involved in news consumption completely removed from the socio-cognitive context in which they may occur.

7. Bridging Theory and Practice: Key Insights from Part I and Implications

Classic persuasion research suggests that individuals unmotivated to engage deeply with a particular message or in environments that do not facilitate deep cognitive processing are more likely to rely on peripheral or heuristic cues when evaluating messages (Chaiken, 1987; Petty & Cacioppo, 1986). These heuristic cues can include not only the source of the message but also social proof, such as the visible judgements of others (Cialdini, 2001). From a theoretical perspective, persuasion research highlights the importance of source similarity (Brock, 1965; Lu, 2013; Pentina et al., 2018; Woodside & Davenport, 1974), while social identity theory emphasises the influence of ingroup membership (Spears, 2021; Van Bavel et al., 2024; Van Bavel & Pereira, 2018). Both perspectives suggest that individuals are more likely to be persuaded or influenced by sources they perceive as similar to themselves (Brock, 1965; Faraji-Rad et al., 2015; Lu, 2013; Ooms et al., 2019; Pentina et al., 2018; Woodside & Davenport, 1974) or as members of their ingroup (Spears, 2021; Turner et al., 1987; Van Bavel & Pereira, 2018). Within the domain of attitudes, source credibility has been identified as a key factor in shaping persuasion (Chaiken & Maheswaran, 1994; Metzger et al., 2003; Pornpitakpan, 2004), with some theorising suggesting that perceived similarity between message source and receiver impacts the perceived credibility of the source (Simons et al., 1970), particularly when they share values and attitudes (Simons et al., 1970). Despite these robust theoretical foundations and their relevance to the misinformation domain, where information processing could occur through similar pathways, applying these concepts to understanding misinformation susceptibility has, prior to the studies presented in Part I of this thesis, remained underexplored. Furthermore, from a social influence perspective, there is evidence that exposure to the judgements of others can shape individual beliefs and attitudes (Cialdini & Goldstein, 2004). However, while limited work has investigated the role of social

cues in misinformation susceptibility (Ali et al., 2022; Butler et al., 2023; Lee et al., 2021), the findings have been inconsistent. Furthermore, it was unclear why exposure to a limited set of social cues from a localised group sometimes influences perceptions of misinformation while, at other times, do not.

The findings presented in Part I of this thesis addressed these gaps, highlighting the role of social and source cues in shaping individual susceptibility to misinformation. These studies underscore the complex interplay between individual cognition and social contexts in misinformation susceptibility. At the core of these findings is the idea that misinformation is rarely evaluated in isolation; instead, judgements are formed within a broader social context. Source effects were particularly salient, as individuals were more likely to believe misinformation from a political ingroup source - an effect likely mediated by perceived credibility. However, when sources were transparently lacking credibility, the influence of partisan biases, ingroup effects, or source similarity diminished, suggesting a boundary condition for these effects. Social proof emerged as another influential heuristic, guiding judgements of misinformation, in part through its impact on perceived social consensus. However, this influence was contingent on individuals paying attention to the cues. When social proof cues were ignored or overlooked, their impact on susceptibility was less pronounced, emphasising the importance of attention and salience in these processes (Cialdini & Trost, 1998). By investigating these socio-cognitive mechanisms, Part I lays a foundation for understanding how misinformation spreads and is perceived as reliable in social environments.

From Mechanisms to Mitigation

Building on these findings, Part II turns to potential interventions aimed at mitigating misinformation susceptibility in contexts where contextual cues identified in Part I exist. Psychological inoculation strategies, grounded in McGuire's (1964) inoculation theory, have shown promise in preemptively bolstering resistance to misinformation by exposing individuals to weakened versions of misleading arguments alongside refutations (Compton et al., 2021; Lewandowsky & van der Linden, 2021; Roozenbeek & van der Linden, 2024; Traberg et al., 2022; van der Linden, 2022). While these interventions have shown effectiveness across various domains (Banas & Rains, 2010; Lu et al., 2023), their real-world applicability remains unclear, as they have yet to be tested in contexts where individuals encounter not only the information itself but also additional factors such as source cues and social proof.

To address this gap, Part II evaluates the robustness of psychological inoculation strategies under conditions where participants encounter misinformation accompanied by persuasive social cues, including exposure to ingroup sources and explicit endorsements. By leveraging the socio-cognitive mechanisms identified in Part I, Part II seeks to determine whether these interventions remain effective in more realistic social settings. Ultimately, this work aims to advance our understanding of psychological interventions' real-world efficacy and inform practical applications for combating misinformation in the digital age.

PART II: IMPLICATIONS FOR PSYCHOLOGICAL INTERVENTIONS

5. INOCULATING AGAINST MISINFORMATION FROM POLITICAL INGROUP NEWS OUTLETS

This work is published in *Harvard Kennedy School Misinformation Review*:

Traberg, C. S., Roozenbeek, J., & van der Linden, S. (2024). Gamified inoculation reduces susceptibility to misinformation from political ingroups. *Harvard Kennedy School (HKS) Misinformation Review*, 5(2). https://doi.org/10.37016/mr-2020-141

1. Abstract

Psychological inoculation interventions, which seek to pre-emptively build resistance against unwanted persuasion attempts, have shown promise in reducing susceptibility to misinformation. However, as many people receive news from popular, mainstream ingroup sources (e.g., a left-wing person consuming left-wing media), which may host misleading or false content, and as research shows individuals are more susceptible to misinformation from political ingroup outlets, the impact of source effects on inoculation interventions demands attention. This chapter uses a between-subjects (3 news source conditions) pre-post (inoculation) design experiment ($N = 657$) to investigate the potential interaction between source effects and inoculation intervention effectiveness against the influence of misinformation. Results demonstrate that although individuals are more susceptible to (non-political) misinformation from political ingroup publishers pre-intervention, inoculation improves both veracity discernment and the perceived reliability of misinformation from both political ingroup and outgroup publishers.

2. Introduction

Research has shown that it is possible to cognitively immunise individuals against misinformation, much like how people can be vaccinated against viral attacks (Compton, van der Linden, Cook & Basol, 2021; Lewandowsky & van der Linden, 2021; McGuire, 1964; Traberg, Roozenbeek & van der Linden, 2022; van der Linden, 2022). Gamified inoculation interventions (Roozenbeek, Linden, et al., 2020b; Roozenbeek & Linden, 2020) – which both forewarn participants about the threat posed by misinformation and cognitively prepare them to refute misinformation that employs deceptive tactics – have shown promise in terms of both their effects and potential for scalability (Roozenbeek, Traberg, et al., 2022; Roozenbeek, van der Linden, et al., 2022). Inoculation interventions focus on reducing susceptibility to misleading content, but as many people receive news from popular political ingroup news sources (Grieco, 2020; Matsa, 2021), which Chapters 2 and 3 demonstrate may increase misinformation susceptibility (Traberg et al., 2024; Traberg & van der Linden, 2022), this prompts the question of whether inoculation interventions can reduce susceptibility to misinformation from these influential sources.

This chapter addresses this question by testing whether a popular gamified inoculation intervention - the *Bad News* game (Roozenbeek & van der Linden, 2019) - can reduce susceptibility to misinformation from popular political ingroup sources. Previous studies have shown that people who play *Bad News* are significantly better at correctly identifying misinformation as being unreliable either post-gameplay (Roozenbeek, Traberg, et al., 2022) or compared to a control condition (Basol et al., 2020).

However, inoculation interventions have largely been tested using misinformation from bogus sources or without source attributions (Traberg et al., 2022). However, as news consumers often rely on and seek out partisan news sources (Grieco, 2020; Matsa, 2021), and as real and popular news sources have also been found to sometimes host misleading content (Motta et al., 2020), it is crucial to understand whether inoculation interventions also protect people even in the presence of potentially powerful source effects. Examining whether interventions can overcome source effects is relevant and important not just for inoculation interventions but any cognitive or psychological intervention aimed at reducing misinformation susceptibility. This is in part because misinformation attributed to popular media sources that news consumers may find particularly credible can be more persuasive than complete falsehoods from bogus sources (Allen et al., 2024; Traberg, 2022; Tsfati et al., 2020; van der Linden & Kyrychenko, 2024).

This chapter employs a between-subjects (3 news source conditions) pre-post (inoculation) design to answer this question. Participants were assigned to view a series of misinformation headlines either from left-leaning US news outlets (e.g., *New York Times, Washington Post*), right-leaning US news outlets (e.g., *Fox News, Wall Street Journal*), or a control condition (with no source information), both before and after playing *Bad News* (total $N = 657$). To examine whether inoculation would be less effective against misinformation from political ingroup publishers than outgroup publishers or control, source conditions were recoded into political ingroups/outgroups based on participants' own indicated political ideology.

It was hypothesised that inoculation would reduce susceptibility to misinformation from all sources (**H1**), regardless of whether the source was politically congruent or incongruent with participants' own ideology. Furthermore, it was hypothesised that inoculation would be effective even against misinformation from political ingroup sources (**H2**), which are typically more persuasive. Moreover, this study explored whether inoculation would mitigate the inherent bias of greater susceptibility to misinformation attributed to ingroup sources compared to outgroup sources post-inoculation (**H3**). Finally, it was hypothesised that inoculation would significantly improve participants' veracity discernment – their ability to distinguish between reliable factual news and misinformation (**H4**).

3. Method

This study aimed to investigate whether source cues impact the efficacy of the inoculation interventions. The study was approved by the University of Cambridge Psychology Research Ethics Committee (PRE.2019.104). All statistical analyses were performed using Jamovi (www.jamovi.org) or R. Visualisations were created using Jamovi and JASP (www.jasp-stats.org). Please see the Harvard Dataverse page for the datasets and analysis scripts.

Participants and Design

Data was collected on the *Bad News* platform (see more information under 'inoculation intervention'). Data collection was set to run for as long as the game platform was available to host the study - between 1st November 2020 until 15th January 2021. The

study and main analyses were preregistered: https://osf.io/uqwd2?view_only=f1030b5dcefe4278b13d092cb0d88676.

A between-subjects design was employed, manipulating source slant across three conditions (left-wing source false/right-wing source facts, n = 374; right-wing source false / left-wing facts, n = 126; and control, n = 157). We re-coded conditions such that we had three group identity categories (ingroup, n = 321, outgroup, n = 179, control, n = 157) based on whether the source slant matched participants' own identified political ideology. The disparity in group sizes resulted from a larger sample of left-leaning participants opting into the study combined with the condition assignment mechanism; this left-wing sampling bias is a known occurrence with samples collected from surveys implemented in the *Bad News* game (e.g., Roozenbeek & van der Linden, 2019). As automatic randomisation was not possible on the game platform, participants were assigned to conditions by selecting a number from 1-3. Participants who selected 1 were assigned to the left-wing source misinformation condition, those who selected 2 were assigned to the right-wing source misinformation condition, and those who selected 3 were assigned to the control condition. Due to more participants selecting 1 and a larger group of left-leaning participants, this resulted in a larger sample in the 'ingroup' condition compared to the two other conditions.

The final sample was N = 657 participants (43% female, 51% male, 5% other; 16% completed high school or less, 40% had some college education, 44% had a higher degree; 71% were between 18-29 years old; 24% between 30-49 years old, 6% over 50 years old; 79% indicated being at least somewhat left-leaning, 20% indicated being at least somewhat right-leaning). As such the sample was unbalanced on these demographic variables.

Materials

Source

The publisher logo was manipulated on news headlines across three conditions: Traditionally left-wing publishers, right-wing publishers, or a control condition where the sources were blurred out. News sources were selected using the same method as Traberg & van der Linden (2022), relying on a previous crowdsourced content analysis of news media (Budak et al., 2016). As in the previous chapters, the left-wing publishers were the *New York Times*, the *Washington Post* and *CNN*, and right-wing publishers were *Fox News*, the *Wall Street Journal* and *Breitbart*. See Figure 25 for an example of the source manipulation across conditions.

Figure 25

Examples of Misinformation Headlines

Note: (A) Left-wing source misinformation, (B) Right-wing source misinformation, (C) Control condition misinformation

Headlines

Participants were exposed to 6 factual headlines and 6 misleading headlines. To select misleading news headlines, the " Hoaxy " platform was developed by researchers at Indiana University, which visualizes the spread of claims and fact-checking. We carried out a search to identify misleading headlines based on their use of one of three misleading strategies used by misinformation producers identified by previous research (Roozenbeek & van der Linden, 2020). These three misinformation tactics included 1) using exaggeratedly emotional language to distort the news story to generate strong emotional responses; 2) creating or inspiring conspiratorial thinking to rationalise current events; 3) discrediting otherwise reputable individuals, institutions, or facts to instil doubt in audiences. The non-misleading headlines did not make use of any manipulation techniques and were based on factual news.

Measures

Perceived Reliability of Misinformation

Participants were asked to rate each item's reliability on a standard 7-point Likert scale: *"How reliable is the above news headline"* (1 = Very Unreliable, 7 = Very Reliable). Participants rated the reliability 6 misinformation headlines for which the perceived reliability was averaged ($M = 3.16$, $SD = 1.16$, $\alpha = 0.70$).

Perceived Reliability of Factual Headlines

To measure misinformation susceptibility, participants were asked to rate each item's reliability on a standard 7-point Likert scale: "*How reliable is the above news headline*?" (1 = Very Unreliable, 7 = Very Reliable). Participants rated the reliability 6 non-misleading headlines for which the perceived reliability was averaged ($M = 4.98$, $SD = 1.14$, $\alpha = 0.69$).

Perceived Source Slant

Participants were asked to rate their perception of the political slant of each of the 6 news sources on a 7-point Likert scale, where 1 was extremely left-wing and 7 was extremely right-wing ($M = 4.26$, $SD = 1.27$).

Source Ingroup/Outgroup

The source group was coded according to whether the participants' reported political ideology matched the political slant of the source. As political affiliation was originally collected on a 7-point scale, this was converted to a binary (left-leaning vs right-leaning) measure and for the analysis, which focused specifically on source similarity, participants who had reported a 'moderate' political stance (that is, a 4 on the 1-7 scale) were excluded. In a secondary analysis and as a manipulation check, we analysed participants' perceptions of the source slant and categorised sources based on this. For example, if a participant identified as being 'left-wing' and rated *CNN* as 'left-wing', this was categorised as an ingroup source.

Inoculation Intervention

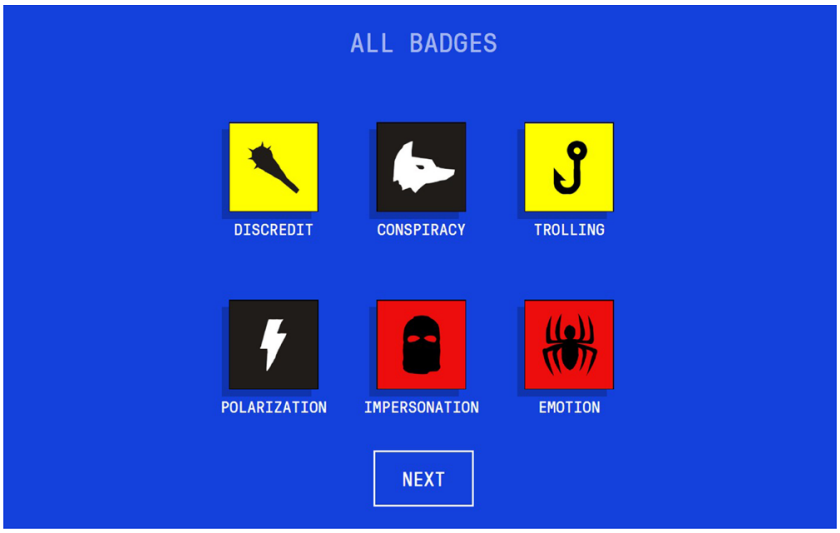
The inoculation intervention used was the *Bad News* game (Basol et al., 2020; Roozenbeek & van der Linden, 2019), an online game in which players learn about six misinformation techniques. Individuals engage in a simulated social media environment, where they are exposed to weakened doses of misinformation strategies. Simultaneously, they are actively prompted to generate their own content. This intervention takes the form of a free social impact game (www.getbadnews.com) created in collaboration with the Dutch media platform DROG (*DROG*, 2023) (see Figure 26).

Figure 26

Screenshots from the Bad News Landing Page (A) and Training Badges (B)

(A)

(B)



Note: Reprinted with permission from Roozenbeek & van der Linden (2019). Training badges represent the deceptive techniques players are taught to use within the game environment.

Demographics

Age was measured categorically (18-29 (71%), 30-49 (24%) and over 50 (6%)). Highest level of completed education was categorised in "High School or Less" (16%), "Some College" (40%), "Higher Degree" (44%), and gender as "Female" (43%), "Male" (51%) and "Other" (16%). Finally, political affiliation was measured on a 7-point scale where 1 = Very Left-wing and 7 = Very Right-Wing (79% indicated being at least somewhat left-leaning, 20% indicated being at least somewhat right-leaning).

Procedure

Participants who visited the website for the *Bad News* game between 1st November 2020 and 15th January 2021 were asked if they would like to take part in a study prior to gameplay. If they agreed, they were asked to provide informed consent. Following this, they answered the socio-demographic information. Participants were assigned to a condition by selecting a number from 1 to 3, which then redirected them to one of the three source conditions. They were then exposed to headlines and told that these were screenshots of real headlines; and asked to report on the above measures (perceived reliability). All participants subsequently played the *Bad News* game after which they were asked to judge the reliability of the same headlines post gameplay. Finally, they answered questions about the various source slants and credibility. After completing the study, participants were debriefed regarding the purpose of the study and informed that the headlines were fictitious.

4. Results

Preliminary Evaluation of Baseline Susceptibility to Ingroup Misinformation

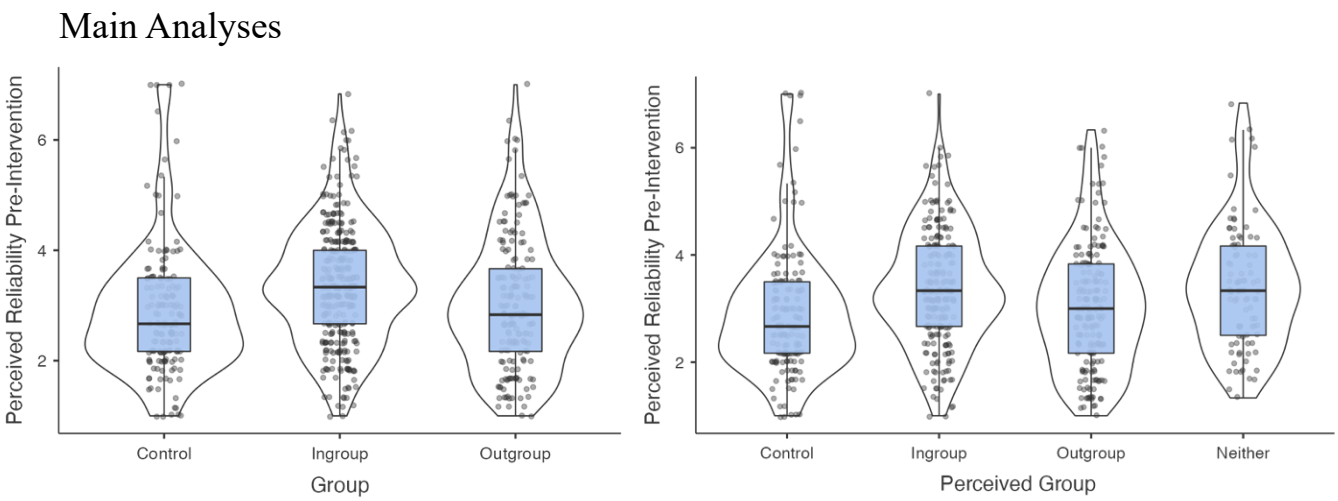
Although not pre-registered, prior to hypothesis testing, it was necessary to assess whether participants would be more susceptible to misinformation from political ingroups pre-inoculation. After all, for it to be helpful to examine whether inoculation interventions can specifically reduce susceptibility to misinformation from ingroup sources, this needs to be a bias observed prior to inoculation interventions in the first place. While the previous chapters (Chapter 2 and Chapter 3) found this to be the case, it was necessary to assess whether this was also the case in the current study.

A general linear model (GLM) was run with perceived reliability of misinformation pre-inoculation as the dependent variable (DV) and group identity (ingroup, outgroup, control) as the independent variable (IV). Results showed an overall model effect ($F(2,654) = 9.97$, $p < 0.001$, $\eta_p^2 = 0.03$). Participants were more likely to judge misinformation headlines as reliable when they were attributed to political ingroup sources ($M = 3.36$, $SE = 0.06$) compared to when attributed to both political outgroups ($M = 3.02$, $SE = 0.09$, $d = 0.29$, $p = 0.005$) and a control condition ($M = 2.91$, $SE = 0.09$, $d = 0.40$, $p < 0.001$). There was no significant difference between the perceived reliability of misinformation attributed to outgroups and the control condition ($p = 0.610$, $d = -0.10$), a finding which suggests that (political) *ingroups* are powerful in making information more persuasive, more so than outgroups making misinformation seem less reliable than if no source attribution is provided.

As a robustness check, we also ran the same analysis with source group level categorised based on participants' own perception of the sources' slant (e.g., if the participant identified as right-wing and they judged the source to be a right-leaning source, this was categorised as a perceived ingroup). In this categorisation, we categorised a source as 'Neither' if participants rated it a 4 (neither left-leaning nor right-leaning on the 7-point scale). A GLM with perceived reliability of misinformation pre-inoculation as DV and perceived group identity as IV was significant *(*$F(3, 644) = 7.15$, $p < 0.001$, $\eta_p^2 = 0.03$), and mirroring the results above, participants were more likely to judge misinformation headlines as reliable when they were attributed to sources they perceived as ingroup sources ($M = 3.36$, $SE = 0.08$) compared to when attributed to both the perceived outgroup ($M = 3.02$, $SE = 0.09$, $d = 0.30$, $p = 0.016$) and the control ($M = 2.91$, $SE = 0.09$, $d = 0.40$, $p < 0.001$). There was also a significant difference between the control ($M = 2.91$, $SE = 0.09$) and the 'neither' conditions ($M = 3.40$, $SE = 0.12$, $d = -0.43$, $p = 0.006$), and between the perceived outgroup ($M = 3.02$, $SE = 0.09$) and neither ($M = 3.40$, $SE = 0.12$, $d = -0.33$, $p = 0.049$) conditions. No other significant differences between conditions were observed. We also ran a Welch's ANOVA accounting for unequal sample sizes ($F(2,341.76) = 9.98$, $p < 0.001$) with Games-Howell post-hoc tests showing significant differences between perceived reliability of misinformation from ingroup sources vs outgroup sources ($M_{diff} = 0.33$, $p = 0.006$) and between the control and the ingroup ($M_{diff} = -0.45$, $p < 0.001$), but no difference between the control and outgroup ($M_{diff} = -0.12$, $p = 0.638$). These findings suggest that sources are powerful persuasive cues when it comes to misinformation and that news consumers are more likely to judge misinformation as reliable if it is attributed to a news publisher perceived to be a political ingroup. Findings are displayed in Figure 27.

Figure 27

Mean Perceived Reliability of Misinformation Pre-Inoculation by Source Group (left) and Perceived Source Group (right)



The results from the main analyses are reported here. Please note that the presentation order differs from that in the pre-registration, and some hypotheses are exploratory (where noted).

H1: Inoculation reduces susceptibility to misinformation from real news publishers.

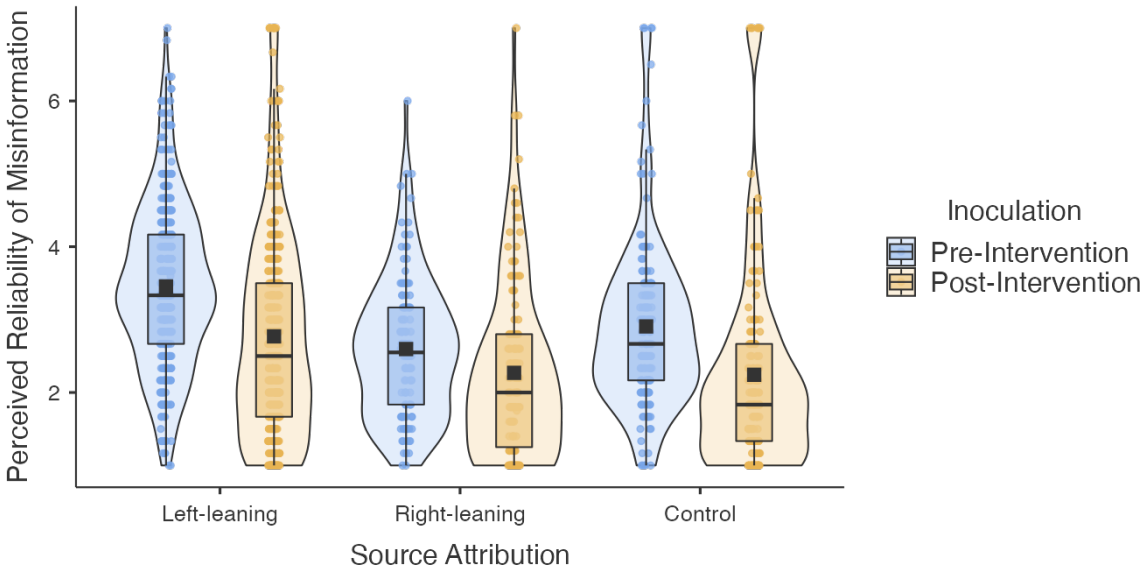
First, results showed a significant overall reduction in the perceived reliability of misinformation after playing *Bad News* ($M = 2.55$, $SE = 0.05$) vs pre-inoculation ($M = 3.16$, $SE = 0.05$, $t(656) = 14.10$, $p < 0.001$, $d = 0.55$).

To further examine the impact of the source conditions compared to the control condition in which no sources were present, we ran a GLM with perceived reliability of misinformation as the DV and inoculation (pre/post intervention) and source condition (left, right, control/no source) as IVs. Based on findings in previous research (Traberg & van der Linden, 2022), we hypothesised that the inoculation effect (the pre/post difference) would be larger for misinformation from left-leaning sources (as our sample skewed left-leaning and as the left-leaning sources included have previously been judged to be more credible, on average). The results showed a main effect of inoculation across conditions ($F(1, 1308) = 53.84$, $p < 0.001$, $\eta_p^2 = 0.04$) such that perceived reliability of misinformation headlines were significantly

lower post-gameplay ($M = 2.43$, $SE = 0.05$) compared to pre-gameplay ($M = 2.99$, $SE = 0.05$, $p < 0.001$ $d = -0.45$). We also find a significant main effect of news source condition ($F(2, 1308) = 38.84$, $p < 0.001$, $\eta_p^2 = 0.06$), such that misinformation headlines from left-leaning sources were judged as significantly more reliable ($M = 3.11$, $SE = 0.05$) compared to from right-leaning sources ($M = 2.43$, $SE = 0.08$, $p < 0.001$, $d = 0.55$) and without source attributions ($M = 2.57$, $SE = 0.07$, $p < 0.001$, $d = 0.43$). No significant differences were observed between reliability judgements of headlines from right-leaning sources and without source attributions ($p = 0.364$, $d = 0.11$). However, we find no significant interaction between the source conditions and the inoculation effect ($F(2, 1308) = 2.07$, $p = 0.13$, $\eta_p^2 < 0.01$), suggesting that there was no significant difference between the inoculation effect based on whether the source was a left-leaning source, a right-leaning source or a control. As such, our hypothesis was not confirmed. These results are displayed in Figure 28.

Figure 28

Mean Perceived Reliability of Misinformation Pre Inoculation vs Post Inoculation Across Source Conditions



H2: Inoculation reduces susceptibility to misinformation from political ingroup sources.

Despite participants being more likely to believe misinformation prior to inoculation, inoculation successfully reduced misinformation susceptibility across all conditions. This

chapter hypothesised that there would be no interaction between political source congruence and the inoculation effect (pre/post difference in perceived reliability of misinformation). Testing null hypotheses can be complicated, as the absence of a significant effect (i.e., $p < 0.05$ under a frequentist framework) does not always mean a *true* absence of effects, as there may be small yet meaningful effects that were not found due to, for instance, a lack of sufficient sample size (Lakens et al., 2020). Therefore, this thesis tested this hypothesis under a Bayesian framework, allowing for more intuitive null hypothesis testing. Specifically, a Bayesian ANOVA was conducted with the perceived reliability of misinformation as the dependent variable, group (ingroup/outgroup/control) and time (pre/post inoculation) as fixed factors, and participant ID as a random effect.

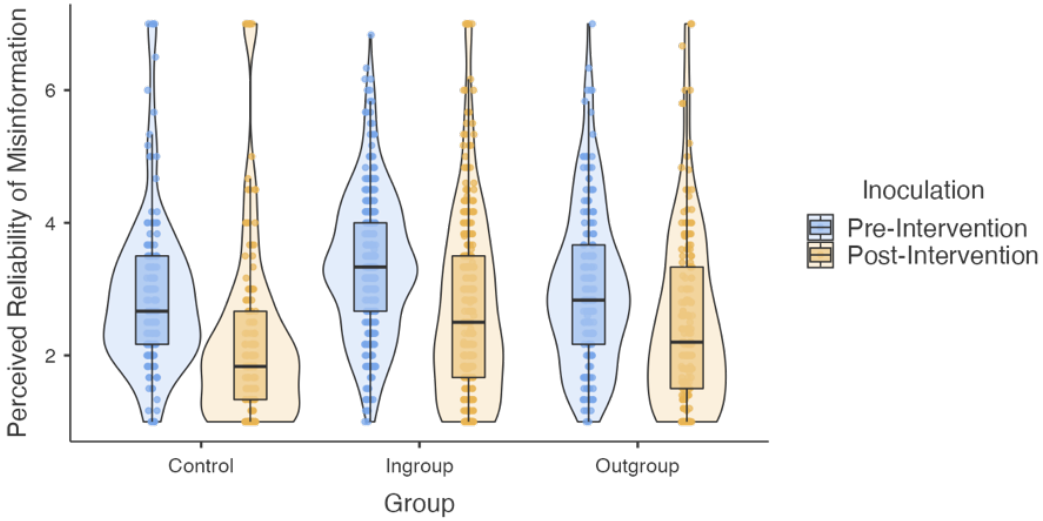
Doing so shows that the interaction group * time (i.e., the interaction between experimental condition and time, i.e., pre vs post) does not add meaningful predictive power under a Bayesian framework ($BF_{Inclusion} = .153$; P(Model | Data) = .037). A Bayes Factor of < .200 is considered sufficient support for the null hypothesis H0, and the equivalent of a Bayes Factor of > 5 is considered sufficient support for the alternative hypothesis H1 (Lakens et al., 2020; van Doorn et al., 2021). In other words, the posterior values of the perceived reliability of misinformation are not meaningfully different if the model controls for this interaction. This means that although both experimental condition (group) and time (pre-post) separately strongly predict the perceived reliability of misinformation ($BF_{10} = 2.67*10^{38}$ for the group + time model), we find support for the null hypothesis that source congruence does not influence the inoculation effect, i.e., the pre-post difference in reliability ratings of misinformation. See Table S1 in Appendix D for a complete overview of this analysis.

In fact, examining the difference in reductions across conditions, there was a greater reduction in the perceived reliability of misinformation in the ingroup-source misinformation condition (M_{diff} = -0.64, SE = 0.10) compared to the outgroup-source misinformation condition (M_{diff} = -0.52, SE = 0.13), although the contrast between the difference scores was not significant (p = 0.51, d = 0.10).

It is important to mention that the samples were unequal and as such, results may be more representative of left-leaning participants, who represent the largest subgroup. However, given the large sample size, the influence of unequal group sizes on overall results is expected to be minimal.

Figure 29

Mean Perceived Reliability of Misinformation Pre-Inoculation and Post-Inoculation Across Source Conditions



H3: News consumers are no longer more susceptible to ingroup (compared to outgroup) attributed misinformation post-gameplay (*please note this hypothesis was exploratory*).

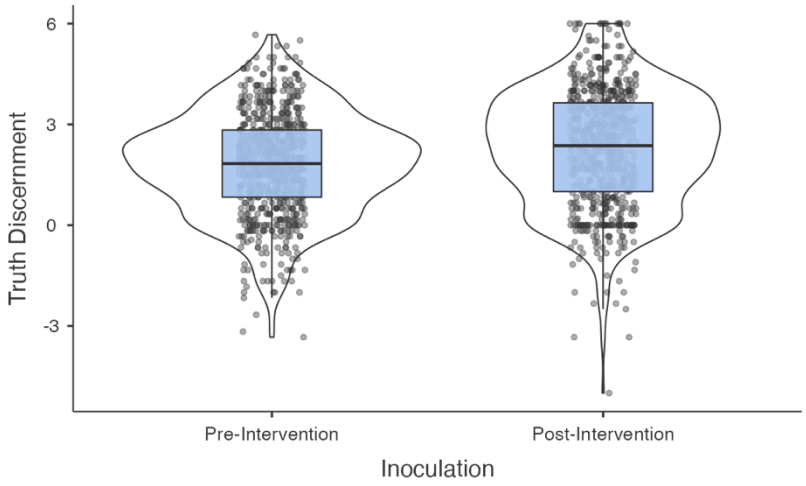
Given that significant differences on perceived reliability of misinformation pre-inoculation were found, an exploratory analysis was found to examine whether these differences were also present post-inoculation. Results from the ANOVA with group identity (ingroup, outgroup, control) and perceived reliability of misinformation post-inoculation showed a significant impact of group ($F(2,654) = 6.55$, $p = 0.002$). However, post hoc comparisons showed the only significant difference was between the ingroup and control ($M_{diff} = 0.48$, $SE = 0.13$ $p = 0.001$, $d = 0.35$), with no significant differences between the ingroup and outgroup ($M_{diff} = 0.22$, $SE = 0.13$ $p = 0.204$, $d = 0.16$) or between the outgroup and control ($M_{diff} = 0.26$, $SE = 0.15$ $p = 0.192$, $d = 0.19$). This suggests that although there was a group-level bias on misinformation judgements prior to the inoculation treatment, this bias was no longer present post-inoculation. It is therefore possible that inoculating individuals against misinformation also helps reduce ingroup-sourced misinformation bias. However, further evidence is necessary to confirm this hypothesis, given the unequal sample sizes. Moreover, source biases may be more pervasive in the real world when it comes to political misinformation. As such, it may still be beneficial for future inoculation interventions to aim to inoculate specifically against source effects on top of inoculating against content-effects.

H4: Inoculation significantly improves veracity discernment.

Given the open question regarding the efficacy of gamified inoculation interventions in improving veracity discernment (van der Linden, 2024), we analysed whether participants' veracity discernment was significantly reduced post-inoculation. Veracity discernment was calculated as the average perceived reliability of factual news minus the average perceived reliability of misinformation. We conducted a paired sample's t-test on veracity discernment pre and post inoculation intervention; showing a significant improvement ($t(651) = -8.48$, $p < 0.001$, $M_{diff} = -0.48$, $SE = 0.06$, $d = -0.33$). That is, participants were significantly better at discerning between the reliability of factual news and misinformation post-inoculation. Figure 30 illustrates this result. We also conducted a Bayesian paired-samples *t*-test on the pre- and post-inoculation reliability judgements of factual news headlines, showing a small reduction in the perceived reliability of factual information ($BF_{10} = 4.36$, $\delta = 0.12$, δ representing the population-level version of Cohen's *d*) and misinformation ($BF_{10} = 1.15*10^{36}$, $\delta = 0.55$), as well as increased discernment ($BF_{10} = 2.49*10^{13}$, $\delta = 0.33$). See the Appendix D for details.

Figure 30

Mean Veracity Discernment (Perceived Reliability of Factual News Minus Perceived Reliability of Misinformation) Pre- and Post-inoculation



5. Discussion

Promisingly, these results show that across conditions, playing *Bad News* reduces misinformation susceptibility; that is, participants are better at correctly identifying misleading headlines as unreliable post-gameplay with no major source effects. This was even though news consumers were more susceptible to misinformation from political ingroup sources before the intervention. These results further show that participants were more likely to perceive misinformation attributed to political ingroup publishers as more reliable pre-inoculation than when it was attributed to an outgroup or with no source information present. This was the case when groups were categorised based on the match or mismatch between a participant's political ideology and a previous crowdsourced content-based analysis of the news source slant (Budak et al., 2016). In addition, similar results emerge when categorising the data based on participants' own perceptions of the source's political slant. At the same time, participants were more likely to reject factual headlines when they were attributed to an outgroup source pre-inoculation.

Encouragingly, across the ingroup, outgroup and control conditions, all participants lowered their reliability judgements of misinformation post-inoculation. We also find that inoculation lowered perceived reliability of factual headlines post-inoculation, but only at an effect size of $d = 0.12$, considered to be a negligible effect size (Sullivan & Feinn, 2012). Furthermore, veracity discernment - participants' ability to discern between factual information and misinformation - significantly improved post-inoculation. Although not entirely possible to rule out, the *Bad News* intervention has previously been found to not suffer from testing effects (where using a pre-test influences the outcome variable or interacts with the intervention), but some evidence of item effects were found (where the intervention itself only improves participants' judgements on the specific test items used) (Basol et al., 2020; Roozenbeek, Maertens, et al., 2021).

This work has a direct impact on our understanding of the impact of source cues on misinformation interventions. The results highlight that sources have a significant impact on misinformation susceptibility but also show that inoculation interventions can significantly improve misinformation susceptibility despite the presence of these effects. In fact, we also find that although group-based source effects are present on pre-inoculation judgements of misinformation, these effects disappear on post-inoculation, suggesting inoculation interventions can contribute to reducing source biases, despite the interventions not being explicitly designed to combat this form of bias itself.

The finding that inoculation reduces susceptibility to misinformation from real news publishers has important implications for our understanding of the efficacy of inoculation interventions in the modern news environment, where even popular and mainstream media sources have been found to *sometimes* publish misleading information (Motta et al., 2020). Misleading information from such publishers may have longer-lasting effects (Tsfati et al., 2020), and these effects may be harder to eliminate, given that mainstream sources may benefit from significantly higher levels of perceived credibility (Traberg, 2022). We further show here that clear source effects appear *before* the intervention. To date, research has only investigated the impact of researcher-categorised in- and outgroups, defining an "ingroup" based on whether a participant's indicated political ideology matches a pre-defined categorisation of source slant – e.g., defining *Fox News* as an ingroup for a participant who identifies as right-wing. We here show the same effects for sources that participants *themselves* believe to share their political ideology.

That source effects appear for both types of group categorisations enhances the real-world applicability of the findings and emphasises the role of source dynamics in shaping individual misinformation susceptibility. Educational interventions may, more generally, benefit from directly tackling the powerful influence of source biases on information processing and empowering learners to understand the persuasive power of source cues. These findings further suggest a call for future research to delve deeper into the complexities of source effects to explore the interplay between shared group identity with sources and misinformation susceptibility.

These results also have a direct impact on our understanding of how we should test and develop interventions tailored to reducing misinformation susceptibility. Although we find that the *Bad News* game reduces susceptibility to misinformation from political ingroup sources, we also find impacts of ingroup/outgroup source effects. This suggests that inoculation interventions might benefit from incorporating training on the impact of contextual cues that may mislead news consumers, just as content-based cues can serve as a means of manipulation.

This work is one of the first tests of a misinformation intervention with contextual cues incorporated. Although we find promising results, the results are based on a limited selection of sources, a limited selection of headlines and, as with most intervention research, limited in the sense that online participants are self-selected and may be those who are already willing to learn about misinformation. Furthermore, there are far more contextual cues and biases in the news environment than sources alone, and publishers are not the only

sources that may play a role in news consumption. The ‘sharer’ or ‘messenger’ of headlines may play an equal if not larger role.

This study advances the growing body of research on inoculation interventions by evaluating their effectiveness in mitigating misinformation susceptibility under more realistic and ecologically valid conditions. As a burgeoning literature is now pointing to a significant impact of sources on misinformation susceptibility (Bryanov & Vziatysheva, 2021; Kim & Dennis, 2018; Traberg & van der Linden, 2022; Westbrook et al., 2023), this work suggests that inoculation interventions may benefit from including training and refutational material that specifically seeks to reduce this bias to achieve effects more relevant to the real world.

6. Limitations

This work is not without limitations. Firstly, while the sources were presented as prominent U.S. media outlets, the global nature of the online recruitment platform likely included participants who were unfamiliar with these outlets or their perceived political biases. This lack of familiarity may have influenced how participants judged the reliability of the headlines and the associated source effects. Secondly, the experimental design does not fully capture the complexity of real-world social contexts. In actual social media environments, factors such as the emotional resonance of a post, group dynamics, and the identity of the person sharing the content (e.g., a trusted friend or influencer) can all significantly affect misinformation susceptibility. For instance, individuals may share misinformation that aligns with their political ingroup, not because they find it reliable but due to identity-affirming motives (Van Bavel et al., 2024). The artificial nature of the study environment limits our ability to fully explore these dynamics. Indeed, as this study used ‘non-political’ news headlines, it remains uncertain whether the inoculation intervention would provide similar protection against other identity-driven biases, such as the tendency to share political rumours (Facciani & Traberg, 2024) and news content (Ecker et al., 2022; Pretus et al., 2023) that align with one’s social or political identity, a pattern observed in previous research. Thirdly, this study did not measure participants’ intent to share misinformation, a key behavioural outcome. Future studies could explore alternative methodologies to better mimic the social motivations and pressures present on real social media platforms, such as simulated sharing tasks or gamified environments.Fourthly, the study utilised a limited selection of misinformation headlines and factual content, which may constrain the generalisability of findings. While the headlines were curated to represent common misinformation tactics, a broader and more diverse set of stimuli could strengthen

evidence for the intervention's effectiveness across different contexts and types of misinformation. Finally, the sample for this study was drawn from a platform where users voluntarily engage with content about misinformation, potentially creating a self-selection bias. These participants may already be predisposed to critical thinking or open to learning about misinformation, which limits the applicability of findings to more skeptical or disengaged populations who may benefit most from such interventions.

Despite these limitations, this chapter provides valuable insights into the interplay between source cues and misinformation susceptibility and the potential of inoculation interventions to mitigate these biases. The following chapter builds on findings in Chapter 4, which demonstrated the influence of social cues on misinformation susceptibility and examines the potential of inoculation interventions in the presence of these persuasive cues.

6. COUNTERACTING SOCIALLY ENDORSEED MISINFORMATION THROUGH INOCULATION

The work presented in this chapter is published in *Advances in Psychology* in the article below:

Traberg, C. S., Morton, T., & van der Linden, S. (2024). Counteracting socially endorsed misinformation through an emotion-fallacy inoculation. *Advances in Psychology*, 2, e765332. https://doi.org/10.56296/aip00017

1. Abstract

This study (N = 755) explores the efficacy of an emotion-fallacy inoculation in reducing susceptibility to emotionally misleading news and investigates the impact of persuasive social cues on its effectiveness. Results show that inoculation significantly reduces the perceived reliability of misinformation (d = 0.23), enhances participants' confidence in their reliability (d = 0.26), and improves veracity discernment (d = 0.23). Findings also reveal that social cues increase the perceived reliability (d = 0.44) and perceived accuracy of misinformation (d = 0.38), even among inoculated individuals. However, the impact of inoculation remains consistent, suggesting that, while social cues enhance the persuasiveness of misinformation, they do not diminish the effectiveness of the inoculation intervention. Finally, participants acknowledge the influence of social cues more on others than on themselves, indicating a third-person consensus effect. The findings highlight the persistent influence of social cues, even in the presence of inoculation, emphasising the need for nuanced interventions to address the complex interplay between emotions, misinformation, and social influence in the digital age.

2. Introduction

One proposed method to combat misinformation is rooted in inoculation theory (McGuire, 1964). Much like how individuals can be immunised against viral contagion, inoculation theory suggests that they can also be pre-emptively vaccinated against undesirable persuasive attacks. However, news increasingly flows through social media sites (Anderson, 2021), which contain a host of other persuasive social cues that complicate the information environment and warrant attention (Traberg et al., 2024). This chapter examines whether it is possible to inoculate news consumers against misinformation, even in the presence of persuasive social cues. As heavy emotional appeals have been identified as one of the 'fingerprints' of misinformation (Carrasco-Farré, 2022; McLoughlin et al., 2024), this chapter focuses on emotionally misleading news and develops an emotion-fallacy inoculation intervention to tackle this. Below, we outline research demonstrating the social context of misinformation susceptibility, followed by a brief review of research on inoculation theory.

The Social Context of Misinformation Susceptibility

Research has demonstrated that 'social proof' - exposure to social cues that indicate the judgements of others - has a significant influence on our own perceptions of information. For example, social interaction can lead to belief acquisition (Moussaïd et al., 2013) and consensus within social groups predicts individual attitudes (Kobayashi, 2018; Lewandowsky et al., 2019). Expressing, sharing, and validating information within groups and communities represents a virtually ineluctable part of the internet (Dvir-Gvirsman, 2019; J. W. Kim, 2018; Lee et al., 2021). A significant proportion of news consumers get at least some of their news from social media sites (Anderson, 2021) and people often discuss socio-cultural issues with their social circles (Sehulster, 2006). As such, news consumers are very likely to encounter the judgements of others in ways that might impact how they assess the veracity of misinformation. For instance, articles are liked and commented on as well as being shared on social media. As seen in Chapter 4 and reported in Traberg et al., 2024, such online commentary has been found to influence perceptions of wider social consensus, with consequences for individual beliefs (Lewandowsky et al., 2019).

Recent research shows that exposure to social proof in the form of fabricated social consensus information (e.g., fabricated percentages of previous participant judgements) im pacts misinformation susceptibility (Traberg et al., 2024). Moreover, exposure to this type of 'local' consensus information also impacts perceptions of wider public consensus (which we

refer to as perceived consensus) – that is, the percentage of the public that news consumers think would find the information reliable.

Although research documents that individual judgements are impacted by social proof, individuals themselves may underestimate their own susceptibility to such cues. Work on social influence shows that this is often undetected by influenced individuals (e.g., Nolan et al. (2008). Similarly, work on media effects demonstrates that people tend to perceive media influences as more pronounced on others than on themselves (i.e., the 'third person media ef- fect'; Sherrick, 2016), which Douglas & Sutton (2004) suggest stems from an underestimation of the persuasibility of the self. This chapter proposes the existence of a similar phenomenon regarding social cues, which we term the "third person consensus effect". If this effect exists, where individuals believe that social (consensus) cues wield a greater influence on others than on themselves, it warrants attention from researchers in the misinformation field. Such a perception could lead individuals to develop a false sense of immunity to social influence, potentially resulting in an underestimation of their vulnerability to misinformation.

Inoculation Theory: A Potential Solution

In response to the spread of misinformation, researchers have investigated whether it is possible to prevent its psychological effects before people are exposed to it (Cook et al., 2017; Lewandowsky & van der Linden, 2021; Traberg, 2024; Traberg et al., 2022, 2023). A problem with misinformation is that even when explicit warnings or retractions are made, misinformation can continue to impact reasoning (Lewandowsky et al., 2012). Inoculation theory (McGuire, 1964) proposes that much like vaccinations against viral contagion, individuals can be psychologically vaccinated against persuasive attacks (Roozenbeek, Traberg, et al., 2022; Traberg et al., 2022; van der Linden, 2022, 2024). In contrast to post-hoc measures such as fact-checks, corrections or retractions, inoculation theory represents a pre-emptive solution to countering misinformation. A growing body of evidence supports the use of inoculation interventions to counter both specific misinformation content (Mason et al., 2023; Spampatti et al., 2023; van der Linden et al., 2017) and the underlying manipulative techniques contained in misinformation (Banas et al., 2023; Cook et al., 2017; Lewandowsky & Yesilada, 2021; Roozenbeek, Traberg, et al., 2022; Roozenbeek, van der Linden, et al., 2022; Roozenbeek & van der Linden, 2024).

According to inoculation theory (McGuire, 1961), inoculation interventions should consist of two components. A threat component forewarns people that a persuasive attack is imminent to cognitively prepare and motivate the mind to be ready to engage with the second component. A refutational pre-emption provides people with the cognitive tools to refute future misinformation, either through exposing specific false arguments (van der Linden et al., 2017), or misleading techniques that underlie misinformation more generally (Cook et al., 2017; Roozenbeek et al., 2022; Roozenbeek, van der Linden, et al., 2022).

Technique-based inoculations have witnessed particular attention due to their potential for scalability (Roozenbeek, van der Linden, et al., 2022). However, while inoculation might help individuals be more discerning in their information environment, information is typically received in the context of other people (Clarkson et al., 2013; Hertz et al., 2017; Tormala et al., 2009). As a result, social processes might also influence the degree to which information is accepted as true. For this reason, researchers have started calling for investigations of how contextual cues in the misinformation environment may impact the efficacy of inoculation interventions (Traberg, 2024).

Emotional Deception & Inoculation

Emotions play a key role in the spread of misinformation (Carrasco-Farré, 2022; McLoughlin et al., 2024). Firstly, people are more susceptible to misinformation in an emotional state (Martel et al., 2020). Secondly, emotions play a role in the spread of misinformation with emotionally deceptive news making up part of the "fingerprints" of misinformation (Carrasco-Farré, 2022). Furthermore, online misinformation has been shown to be significantly more emotional than non-deceptive content (Paschen, 2019; Peng et al., 2023). Although emotions are a valuable part of human communication (Juez & Mackenzie, 2019), when used deceptively to evoke outrage, anger, or other strong emotions (Roozenbeek, van der Linden, et al., 2022), they can hinder our ability to critically assess information and manipulate attention away from the evidence by evoking irrelevant cues (i.e. the appeal-to-emotion-fallacy: Blassnig et al., 2019; Hamlin, 1970). For this reason, technique-based inoculation interventions target emotional deception as a key misinformation strategy (Roozenbeek, van der Linden, et al., 2022).

Despite the demonstrated effects of social cues on individual information judgements, it is unknown whether inoculation interventions can withstand these social influences. Is it possible to inoculate individuals against misinformation even when their social environment appears to endorse misinformation as reliable? In the current work we set out to address this

question. Firstly, this chapter investigates whether an emotion-fallacy inoculation intervention can protect individuals against the influence of misleading emotional news, and secondly, whether this intervention is successful in the presence of strong social influences. Finally, the chapter examines the potential for third person consensus effects.

3. Method

The purpose of this study was to test the efficacy of an inoculation intervention in the presence of persuasive social cues in a 2 (inoculation vs control) by 2 (social cues vs no social cues) randomised control trial design. The study was approved by the Cambridge Research Ethics Committee (PRE.2022.117).

Participants

The sample was UK-based. A power analysis indicated that 580 participants would be required to detect the hypothesised effects based on previously found inoculation (Roozenbeek et al., 2022) and social cue effect sizes (Traberg et al., 2024). After excluding 5 participants who did not pass the attention checks the final sample was $N = 755$ with 48.4% identifying as liberal (i.e., below 4 on a 1-7 Left-Right political orientation scale), 54.6% female, 66% had at least a bachelor's degree and $M_{age} = 43$, $SD_{age} = 13.3$. 96% of participants had social media accounts with 78% reporting they used it at least occasionally and 59% of participants got the majority of their news from online news sites or on social media.

Participants were randomly assigned to one of the four conditions with n=192 in the *Inoculation with social cues* condition, n=187 in the *Inoculation without social cues* condition, n=191 in the *Control with social cues* condition and n=190 in the *Control without social cues* condition.

Procedure

A study on "News Evaluation" was advertised via *Prolific* and the study was run on *Qualtrics*. Upon electronic consent, participants were randomly assigned to one of four conditions: 1) Inoculation with social cues, 2) Inoculation without social cues, 3) Control with social cues, or 4) Control without social cues.

In inoculation conditions, participants were exposed to a two-part text-based inoculation message. They were asked to read the message. The page could not be skipped until at least 30 seconds had passed. In the control conditions, participants were given a word-search task, with a 30 second timer that ensured they stayed on the page for the same

duration. In all conditions, participants were subsequently shown a series of misleading and neutral headlines and asked to make a series of judgements about them (see "Measures"). Participants who were assigned to the conditions with social cues (both the inoculation and control/word search task) were informed that a high percentage of previous participants had judged each headline to be reliable (always above 75 percent). Participants were also informed that their judgements would be used to calculate the average percentage for future participants. In the *no social cues* conditions, participants were not shown this statement. Then, participants answered a series of demographic questions, third person effect questions and provided responses to psychological scales. Finally, participants were debriefed.

Materials

Inoculation Message

The inoculation intervention consisted of a two-part message (see Figure 1). The threat component forewarned participants about the threat of misinformation. The cognitive refutation (pre-bunk) element of the message informed participants about the use of emotionally deceptive language. The message also provided information about the appeal-to-emotion fallacy, which involves manipulating emotions to divert attention or influence perceptions away from the evidence. To avoid fearmongering about the use of emotions in news in general, the message highlighted that emotions are a natural and valuable part of human communication, and that they can enhance the impact of messaging when used honestly and transparently.

Figure 31

Two-Component Inoculation Message: Threat and Pre-Emption

Control Task

In line with previous inoculation research using control conditions (van der Linden et al., 2017) participants in the control condition were given a neutral word-search puzzle.

Headlines

Participants made judgements regarding six headlines which they were informed had been published online between 2021-2022: 3 emotionally misleading headlines and 3 neutral headlines (see Table 5). The emotionally misleading headlines were created for the purpose of the research and used the appeal-to-emotion fallacy: That is, they used emotional language to divert attention from an issue. e.g., *"Horrific nuclear meltdown causes chaos and despair in local town, nuclear energy no solution for climate change residents claim."* In this instance, the mention of 'nuclear meltdown' does not pertain to the potential of nuclear power as a solution to climate change. The neutral headlines were real headlines that had been published

in 2021 and did not make use of emotions, instead stating 'matter of fact' happenings or news.

Table 5

Misleading and Neutral Headlines Included

Headline	Misleading	Neutral
Horrific nuclear meltdown causes chaos and despair in local town, nuclear energy no solution for climate change residents claim.	x	
Heartbreaking story: baby elephant gets horribly hurt after falling off a ledge, mother elephant cries for HOURS!	x	
Students accused of cheating note how outrageous and inhumane housing conditions make it difficult to study.	x	
Physical fitness keeps your brain in good shape.		x
Netflix to include mobile games for subscribers.		x
Apple, Google and Amazon named as most valuable brands in the world.		x

Measures

Participants were asked to make the following judgements about news headlines:

Perceived Reliability of Headlines

After reading each headline, participants were asked: If 1 is "*Not at all reliable*" and 7 is "*Very reliable*", how reliable is the above headline? (1-7 scale). The reliability ratings were averaged separately for misleading headlines ($M = 3.61$, $SD = 1.11$, $\alpha = 0.78$) and neutral headlines ($M = 5.67$, $SD = 1.13$, $\alpha = 0.78$).

Confidence

For each headline, participants were also asked: "If 1 is "*Not at all confident*" and 7 is "*Very confident*", how confident are you in your reliability judgement above?" (1-7 scale). The confidence ratings were averaged separately for misleading headlines ($M = 4.66$, $SD = 1.25$, $\alpha = 0.87$) and neutral headlines ($M = 5.64$, $SD = 0.95$, $\alpha = 0.82$).

Perceived Accuracy

For each headline, participants were also asked: "If 1 is "*Not at all likely*" and 7 is "*Very likely*", how likely is it that the content of the above headline is an accurate description

of reality?" (1-7 scale). The perceived accuracy ratings were averaged separately for misleading headlines (M = 3.56, SD = 1.13, α = 0.78) and neutral headlines (M = 5.9, SD = 0.86, α = 0.75).

Perceived Consensus

For each headline, participants were asked: "What percentage of the general public do you think would believe the above headline?" (*0-100%*). The perceived consensus ratings were averaged separately for misleading headlines (M = 62.38, SD = 16.13, α= 0.80) and neutral headlines (M = 80.62, SD = 10.53, α= 0.75).

The following variables were also computed:

Reliability Discernment

Average perceived reliability of neutral headlines minus average perceived reliability of misleading headlines (M = 2.06, SD = 1.48).

Accuracy Discernment

Average perceived accuracy of non-misleading headlines minus average perceived accuracy of misleading headlines (M = 2.03, SD = 1.27)

In addition to the above main measures, the following data was also collected.

Third Person Consensus Effects

Participants were asked to report on two measures:

Social cue influence on the self: On a scale from 1-7, where 1 is "*Not at all*" and 7 is "*A lot*", please rate how much you think other people's judgements influence your own opinions and behaviour (M = 3.85, SD = 1.44).

Social cue influence on others: On a scale from 1-7, where 1 is "*Not at all*" and 7 is "*A lot*", please rate how much you think other people's judgements influence the general public's opinions and behaviours (M = 5.15, SD = 1.07)

Third person consensus effect: Computed as *social cue influence on others – social cue influence on the self.* A positive score indicates that participants judge others to be more susceptible to social cues than the self.

Attention Checks

Two attention checks were included. In the first, participants were asked to select 'Disagree' for the question: "*Imagine you are planning a trip to the beach. Please select 'Disagree' for this question."* In the second, they were told: *"In the previous questions, you were asked to rate the reliability of various headlines. Please select 'Reliability 6' in this question."*

Demographic Variables

Socio-demographic variables included gender (*male*, *female*, *other*), age (*birth year*), political orientation (measured on a 7-point Likert scale where 1 is *very left-wing* and 7 is *very right-wing*), level of education (*less than high school degree*, *high school graduate*, *bachelor's degree*, *master's degree*, *doctoral degree* or *professional degree*), current use of social media ("*I don't have any accounts*"; "*I have one or more accounts but I hardly ever use them*"; "*I have one or more accounts, and I use them occasionally*"; "*I have one or more accounts and I use them often*"; "*I have one or more accounts and I use them on a daily basis*") and main news consumption source ("*I don't really follow the news*", "*Social Media*", "*TV and radio*"; "*Print Media (newspapers, magazines)";* "*Word of mouth*", "*Online news sites (excluding social media)".*

4. Results

The main hypotheses and analyses were pre-registered: https://osf.io/8mdhp/?view_only=da45188967dd401ba27e0dfb891d3a96. Exploratory analyses are explicitly highlighted as such. Data for this study is available on: https://osf.io/5vzxg/?view_only=400b91a2036847ef9cec70743427ebb4.

Inoculation Effects

Based on previous research, it was hypothesised that inoculated participants would show greater immunity to misinformation compared to those in the control condition on several key outcome variables. It was specifically hypothesised:

H1. Participants who are inoculated will judge misinformation headlines to be less reliable (A) less accurate (B) and (C) be more confident in their reliability judgements than participants in the control condition.

Three independent samples *t*-tests were run with Bonferroni corrections for multiple testing, confirming the three hypotheses (*A*, *B* & *C*) showing that participants who were inoculated judged misinformation headlines to be significantly less reliable ($t(752.91) = 3.10$, $p = 0.002$, $M_{\text{Control}} = 3.73$, 95% CI [3.62, 3.85], $M_{Inoculation} = 3.48$, 95% CI [3.37, 3.59], $d = 0.23$), were more confident in their reliability judgement ($t(751.95) = -3.54$, $p < 0.001$, $M_{\text{Control}} = 4.50$, 95% CI [4.38, 4.62], $M_{Inoculation} = 4.82$, 95% CI [4.69, 4.95], $d = 0.26$) and judged the headlines to be less accurate ($t(750.41) = 2.90$, $p = 0.004$, $M_{\text{Control}} = 3.68$, 95% CI [3.57, 3.80], $M_{Inoculation} = 3.44$, 95% CI [3.33, 3.55], $d = 0.21$). We therefore confirm *H1A, H1B* and *H1C*.

As there has been scholarly discussion regarding the efficacy of inoculation interventions in improving discernment – that is, the ability to discern between true and false – we also put this to the test in the following hypothesis:

H2. Participants who are inoculated will show higher reliability discernment (A) and accuracy discernment (B) than participants in the control condition.

Confirming *H2*, two independent samples t-tests run with Bonferroni corrections for multiple testing showed that participants who were inoculated had higher reliability discernment ($t(746.41) = -3.23$, $p = 0.001$, $M_{\text{Control}} = 1.89$, 95% CI [1.75,2.03], $M_{Inoculation} = 2.23$, 95% CI [2.08, 2.39], $d = 0.24$) and higher accuracy discernment ($t(752.62) = -3.19$, $p = 0.001$, $M_{\text{Control}} = 1.89$, 95% CI [1.76, 2.01], $M_{Inoculation} = 2.18$, 95% CI [2.05, 2.31], $d = 0.23$).That is, inoculated participants were better at discerning between the veracity of misleading and neutral information than those in the control condition.

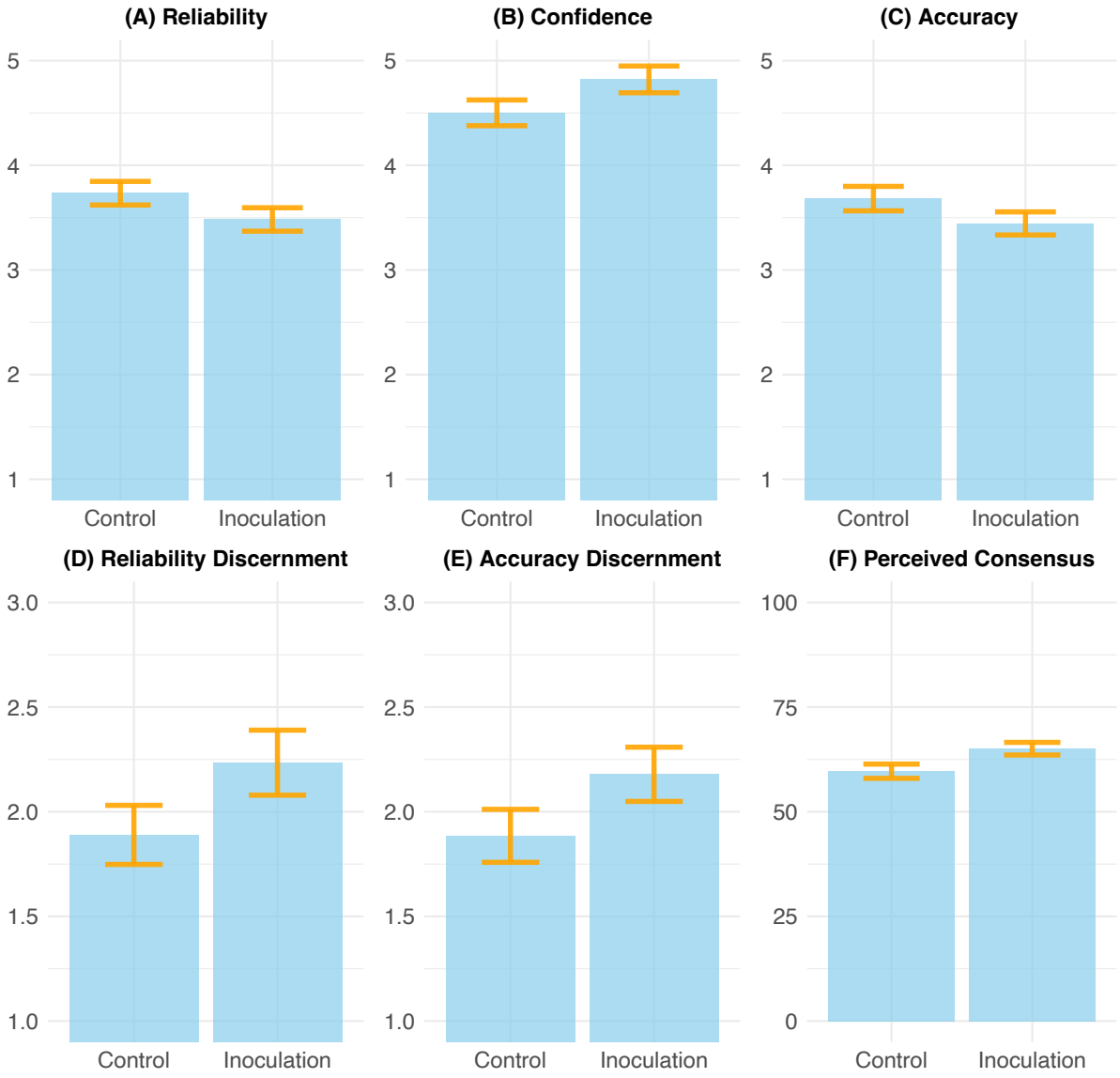
Perceived Consensus

An exploratory analysis showed that participants who were inoculated believed a significantly larger percentage of the general public would judge the headlines to be reliable (which we refer to as '*perceived consensus'*) ($t(742.27) = -4.65$, $p < 0.001$, $M = 65.07$, 95%

CI [63.56, 66.58]) compared to in the control condition (M = 59.68, 95% CI [57.98, 61.38], M_{diff} = -5.49, d = -0.34). Figure 32 illustrates the results of the six comparisons above.

Figure 32

Judgements Compared Across Control vs Inoculation Conditions.



Note: Error bars show 95% confidence intervals.

Social Cue Effects and Interaction with Inoculation

A key question posed in this paper is whether *even* those who are inoculated are impacted by social proof. *H3* tests whether this is the case. *H4* expands on this assumption by testing whether social cues have a greater impact on those who are not inoculated.

H3. Participants in the inoculation with social cues condition will judge misinformation headlines to be more reliable (A) and more accurate (B) than participants in the inoculation without social cues condition.

H4: There will be an interaction between social cue condition and inoculation such that participants in the control condition (who are not inoculated) will be more impacted by social cues than those who are inoculated, showing higher perceived reliability (A) and perceived accuracy (B) of misleading headlines.

To examine *H3* and *H4*, we ran a 2x2 factorial ANOVA to assess whether the difference between the control and inoculation conditions was significantly smaller when social cues were present: That is, did the impact of being inoculated (compared to the control) depend on whether social cues were included or not? The ANOVA model showed a significant main effect of inoculation (F (1,751) = 10.37, $\eta^2 = 0.01$, $p = 0.001$), a main effect of social cues (F(1,751) = 35.75, $\eta^2 = 0.05$, $p < 0.001$) but no interaction between the two (F (1,751) = 0.19, $\eta^2 < 0.01$, $p = 0.659$). That is, inoculated participants judged misleading headlines as less reliable (M = 3.48, 95% CI [3.37, 3.59]) than those in the control (M = 3.73, 95% CI [3.62, 3.84], $p = 0.001$, $d = 0.23$) Furthermore, the presence (M = 3.84, 95% CI [3.73, 3.95]) (versus absence) (M = 3.37, 95% CI [3.26, 3.48]) of social cues increased the perceived reliability of misleading information ($p < 0.001$, $d = -0.44$). Interestingly, although the interaction between social cues and inoculation was not found significant (rejecting *H4A*), the effect size for social cues was larger ($\eta^2 = 0.05$) than for inoculation ($\eta^2 = 0.01$). Results of these analyses are visualised in Figure 3 below.

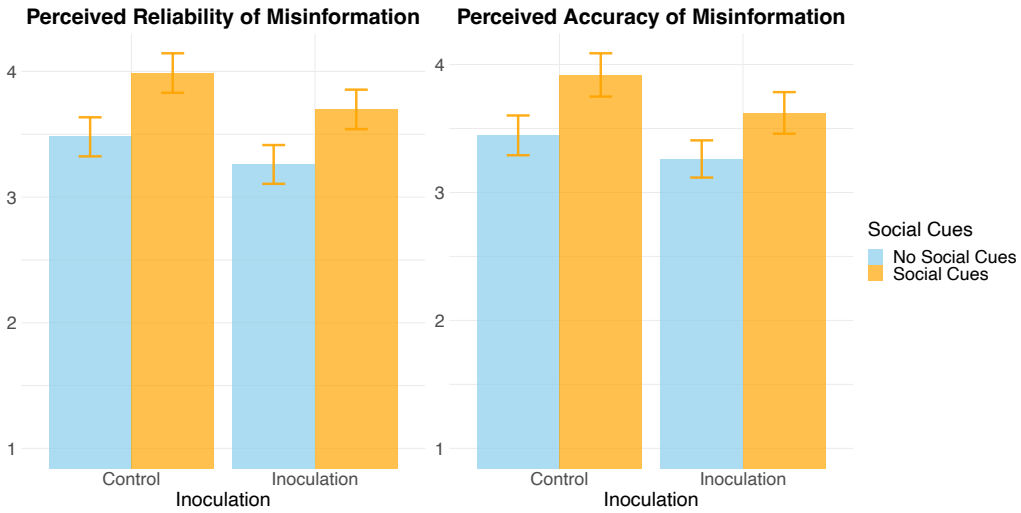
This result suggests that although it is also possible to inoculate individuals in the presence of social cues, social cues make information more persuasive *even* for those who have been inoculated. We therefore confirm *H3A*. Post-hoc tests with planned contrasts between the *inoculation with social cues* condition and the *inoculation without social cues* conditions revealed that inoculated participants were significantly impacted by social cues (t (751) = -4.54, $p < 0.0001$, $d = 0.43$). Inoculated participants who were exposed to social cues judged false headlines to be significantly more reliable (M = 3.70, 95% CI [3.54, 3.85]), than inoculated participants who did not see social cues (M = 3.26, 95% CI [3.10, 3.42]).

To assess the impact of social cues among inoculated individuals on perceived accuracy of false headlines, a 2x2 ANOVA with inoculation and social cues as the IVs and perceived accuracy as the DV was run. We find a significant main effect of inoculation (F (1,751) = 8.71, $p = 0.003$) with perceived accuracy being higher in the control condition (M =

3.68, 95% CI [3.57, 3.79]) compared to inoculation condition (*M* = 3.44, 95% CI [3.33, 3.55], *d* = 0.22), a main effect of social cues (*F* (1,751) = 26.78, *p* < 0.001), with perceived accuracy being higher in the social cues condition (*M* = 3.77, 95% CI [3.66, 3.88) than the no social cues condition (*M* = 3.35, 95% CI [3.24, 3.47], *d* = -0.38), but no interaction between the two (*F* (1,751) = 0.49, *p* = 0.485). Post-hoc tests with planned contrasts between the inoculation with social cues and the inoculation without social cues conditions showed that among inoculated participants, those who saw social cues judged headlines as more accurate (M = 3.62, 95% CI [3.47, 3.78]) than those who did not (*M* = 3.26, 95% CI [3.10, 3.42], *t* (751) = -4.15, *p* = 0.009, *d* = 0.33). As such, we confirm *H3B*. As there was no significant interaction between social cues and inoculation, however, we reject *H4B*. Results are visualised in Figure 33.

Figure 33

Perceived Reliability of Misinformation and Perceived Accuracy of Misinformation Across 4 Conditions



Note: Error bars show 95% confidence intervals.

In line with previous research showing that being exposed to social cues indicating the judgements of others can impact perceptions of wider public consensus (perceived consensus) (Traberg et al., 2024), we test the following hypothesis:

H5: Participants in the social cues condition will show higher perceived general public consensus than those in the no social cues condition.

An independent samples t-test comparing social cue conditions on perceived general public consensus revealed that participants who had been exposed to social cues perceived significantly higher consensus than those who had not been exposed to social cues (t(743.18) = -8.023, $p < 0.0001$, M_{diff} = 9.05, 95% CI [6.84, 11.27], d = 0.58). That is, being exposed to social cues suggesting a local group of previous participants had judged the misleading headlines to be reliable (compared to a control) led participants to judge that a higher percentage of the public would also find the headlines reliable.

Third-Person Consensus Effects

As research has suggested that individuals underestimate the persuasibility of the self (Douglas & Sutton, 2004), it is hypothesised that:

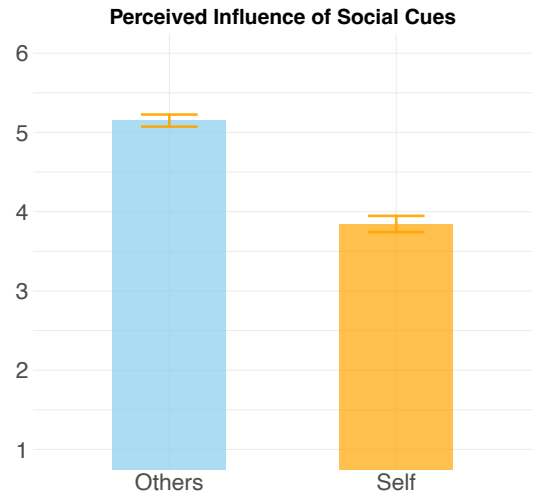
H6: Participants will judge other people as being more likely to be impacted by social cues than themselves.

Confirming H6, a paired samples t-test showed that participants believed other people would be more influenced by social cues (t(754) = -22.60, $p < 0.001$, M_{diff} = 1.30, 95% CI [1.19, 1.42], d = 0.82) than they would themselves.

As data for this variable was collected after the experimental manipulation, as a robustness check we assessed whether the social cue manipulation had a significant impact on the two outcome variables. An independent samples t-test showed that seeing social cues alongside news headlines did not have a significant impact on the perceived influence of social cues on the self (t (752.95) = -0.96, M_{diff} = -0.10, 95% CI [-0.31, 0.11], d = -0.07) nor on others (t (751.18) = -0.48, M_{diff} = -0.04, 95% CI [-0.19, 0.12], d = -0.04). Figure 34 illustrates the respective means for the perceived influence of social cues on self vs others.

Figure 34

Perceived Effects of Social cues on Self vs Others



Note: Error bars show 95% confidence intervals.

Third Person Consensus Effects as a Moderator

As we find a significant difference between judgements of the perceived influence of social cues on others versus the self, we explored the impact of the computed third person consensus effect variable (*social cue influence on others – social cue influence on the self*) on the main outcome measure (*perceived reliability of misinformation*), as well as its potential moderating role.

An exploratory regression analysis was run to assess the moderating role of the third person consensus effect on perceived reliability of misinformation. Results indicated that the main effect of the *third person consensus effect* on perceived reliability of misinformation was not significant ($\beta = 0.01$, $p = 0.729$). In line with previous analyses, the direct impacts of social cues ($\beta = 0.48$, $p < 0.001$) and inoculation ($\beta = -0.34$, $p = 0.001$) were significant. However, the interaction terms for the third person consensus effect with social cues and inoculation were not significant. That is, the interaction between third person consensus and social cues, ($\beta = 0.01$, $p = 0.895$) and the interaction between third person consensus and inoculation ($\beta = -0.06$, $p =. 214$) did not significantly moderate the relationships between these variables and perceived reliability of misinformation.

Inoculation and Non-Misleading Headlines

In exploratory analyses, we evaluate the impact of inoculation and social cues on the evaluation of non-misleading headlines. First, we analyse the impact of inoculation and social cues on perceived reliability of non-misleading headlines. A 2x2 factorial ANOVA showed no significant main effect of inoculation (F (1, 751) = 1.33, p = 0.250), but a significant main effect of social cues (F (1, 751) = 7.48, p = 0.006) and no significant interaction between the two (F (1, 751) = 0.39, p = 0.532). Tukey's post hoc tests revealed that the only significant contrast was between the inoculated participants who saw social cues (M = 5.8, 95% CI [5.64, 5.96]) and the control group who did not see social cues (M = 5.49, 95% CI [5.32, 5.65] p = 0.031). Second, we analyse the impact of inoculation and social cues on perceived accuracy of non-misleading headlines. Here, a 2x2 factorial ANOVA showed no significant effect of inoculation, no significant effect of social cues and no significant interaction effects. As such, neither inoculation nor the social cue manipulation significantly impacted perceptions of non-misleading headlines. Nevertheless, it's worth noting that prior research suggests social cues may only influence perceptions when they oppose one's own views (Traberg et al., 2024). In this case, it is therefore not surprising that social cues indicating previous participants judged non-misleading headlines as *reliable* did not impact judgements of neutral news.

5. Discussion

The findings of this study shed light on the interplay between inoculation, social cues, and susceptibility to emotionally misleading news. The results indicate that an emotion-fallacy inoculation intervention effectively reduces susceptibility to misleading emotional news, supporting the notion that pre-emptive strategies can protect individuals against the persuasive tactics employed in misinformation dissemination. The significant decrease in perceived reliability, increased confidence in reliability judgements, and improved veracity discernment among inoculated participants underscores the robustness of the emotional fallacy inoculation approach. That is, inoculating news consumers against emotionally manipulative news increases their ability to resist judging emotionally deceptive headlines as reliable and accurate and increases their confidence in this assessment.

Of course, emotions are part and parcel of human communication and play a role in all types of discourse (Juez & Mackenzie, 2019), including factual news that adheres to strict journalistic practices. The inoculation message in this study specifically noted this caveat,

and it is therefore positive that inoculated participants were better at discerning between misleading and neutral news compared to the control condition. As such, contrary to some reports that media literacy interventions could make people too sceptical of the news (van der Meer et al., 2023), we find that inoculation messages which help people look out for specific emotional manipulation attempts (rather than all emotional news) improve veracity discernment, consistent with other recent work on logic and fallacy-based inoculation (Banas et al., 2023; Hruschka & Appel, 2023).

Interestingly, the study also reveals a noteworthy influence of social cues on misinformation susceptibility. Even among inoculated individuals, those who were exposed to persuasive social cues judged misleading headlines as more reliable than those who only saw the headlines. This is in line with research showing that when social cues signal a majority judges misinformation to be reliable, this increases misinformation susceptibility (Traberg et al., 2024). This suggests that while inoculation can reduce misinformation susceptibility in general, it does not protect news consumers against the effects of social proof.

It is interesting to note that the social cue effect (increasing misinformation susceptibility) was descriptively larger than the inoculation effect (reducing misinformation susceptibility). This highlights both the potency of social influence in shaping individuals' judgements and underscores the importance of social context in considering what makes people vulnerable to misinformation, including the social dynamics that spread and sustain it. While crowds have been shown to be wise when their judgements are aggregated at scale (Martel et al., 2024), social media users are rarely exposed to the aggregated judgements of sufficiently large, independent and diverse groups – group attributes that seminal research has identified as necessary for the emergence of 'collective intelligence' (Malone & Bernstein, 2022). In fact, due to the existence of online echo-chambers (Törnberg, 2018), is likely that news consumers do not wish to call out misinformation shared within their own networks (Allen et al., 2022).

Given that current interventions against misinformation focus on content rather than context, the need to explore the role of social context more thoroughly in misinformation and interventions against this seems imperative (Traberg, 2024). That said, the lack of a significant interaction between inoculation and social cues suggests that while social influences can be problematic, the protective benefits of inoculation remain consistent in the face of this. This underscores the potential robustness and versatility of inoculation interventions in addressing the challenges posed by social dynamics in the digital age (

Compton et al., 2021). While the current short and text-based intervention may not be as scalable as digital, online-accessible gamified inoculations, the results here speak to the notion that inoculation interventions more generally may be useful for reducing online misinformation susceptibility - even in social contexts.

However, the current research also suggests a necessity for developing inoculation messages that inoculate news consumers against social influence effects, in particular as the study also uncovers what we term a '*third-person consensus effect'*. This is an effect wherein participants acknowledge the influence of social cues more on others than on themselves. This highlights that while individuals are aware of the impact of social influence on others, they do not tend to acknowledge the effect of social influence on themselves. This finding highlights the potential need for further interventions that can address this meta-perception. For example, research has highlighted that meta-cognitive judgements - how citizens reason about their own reasoning - can predict important tendencies such as likelihood of polarising science (Said et al., 2022).

This research is,of course, not without limitations. Firstly, while the study finds significant inoculation effects, the effect size were smaller (d = 0.23) than those found in game-based interventions, which tend to range from d = 0.35 d = 0.60 (Basol et al., 2020, 2021; Roozenbeek, Traberg, et al., 2022; Roozenbeek & van der Linden, 2019) though still in line with the average meta-analytic effect for veracity discernment (Lu et al., 2023). This may be due to the fact that game-based interventions are often longer (e.g., 7–15-minute gameplay) and require more cognitive involvement and reflection than reading text on a screen. However, text-based interventions may be easier to develop, and employ by practitioners or in public communication campaigns. So, despite their more modest effects, these may still be practically useful tools.

Second, the study focused on inoculating individuals against emotionally misleading news. Naturally, there are more deception strategies employed in misinformation (Roozenbeek, van der Linden, et al., 2022). This study cannot speak to whether inoculation against other deception strategies – or against more polarising, extreme, or identity-based misinformation (Pereira et al., 2023; Van Bavel & Pereira, 2018)– is robust in the presence of strongly supportive social cues. However, recent work has shown that inoculation successfully reduces susceptibility to misinformation from political ingroup news outlets (Traberg, Roozenbeek & van der Linden 2024).

In addition, in relation to social cues, the social proof included in the present study was attributed to an unknown and undefined group. But social identity research highlights

that we are more likely to be influenced by others who share with us a salient and meaningful social category – when others are "ingroup members" (Spears, 2021). Attempts at influence by outgroup members are instead likely to fail or may even backfire and contribute to polarisation (e.g., Abrams et al., 1990; Mackie et al., 1990; McGarty et al., 1994). Further research is needed to assess whether inoculation can overcome more specific (in-group) social influence effects. Third, although the protective effects of inoculation have been shown to last up to three months (Maertens et al., 2021), this study cannot speak to whether or not this particular text-based intervention effects stand the test of time. Finally, the study used a U.K.-based sample, limiting the generalisability of the findings to a more diverse global audience.

Despite these limitations, this research contributes to the understanding of the intricate relationship between inoculation, social cues, and misinformation susceptibility. The results suggest that while inoculation provides a valuable defence against emotionally misleading news, the role of social influence cannot be ignored. As misinformation continues to evolve in the digital landscape, the findings underscore the need for multifaceted and adaptive interventions to effectively mitigate its impact on public perception and decision-making. Future interventions may especially need to incorporate strategies to address the influence of social cues to disrupt the real-world dynamics through which misinformation thrives within social networks. For example, injunctive social norms – descriptions of what most people approve or disapprove of – have been shown to increase reportings of fake news (Gimpel et al., 2021) – highlighting that while in the current study social cues play a negative influence role, it may be possible to harness social information to improve the online news environment.

7. DISCUSSION

As misinformation increasingly permeates the modern information landscape (Allen et al., 2024; Carrasco-Farré, 2022; Del Vicario et al., 2016; Ecker et al., 2024; World Economic Forum, 2024), understanding the psychological mechanisms that lead individuals to accept misinformation as true has become a critical area of research. The potential societal consequences - such as eroding trust in democratic institutions or exacerbating public health crises (Ecker et al., 2024; Roozenbeek et al., 2020) - underscore the importance of developing effective strategies to mitigate the influence of misinformation. In today's media ecosystem, individuals increasingly consume information on social media platforms (Anderson, 2021; Atske, 2021), where they are inundated with content (Feng et al., 2015; Holton & Chyi, 2012; Jaeger & Burnett, 2010; Matthes et al., 2020). Not only are individuals likely to experience "information overload" (Matthes et al., 2020; Rodriguez et al., 2014) that could amplify reliance on cognitive shortcuts (Chaiken, 1987; Petty & Cacioppo, 1986), but social media platforms also immerse individuals in dynamic, socially enriched settings full of social contextual cues (Mena et al., 2020). This combination of an environment that discourages deep content processing and platforms that prominently highlight contextual cues (Sha et al., 2020) necessitates an in-depth examination of the impact of contextual cues on misinformation susceptibility.

When individuals are exposed to misinformation on social media sites, there are two main contextual cues that tend to accompany news headlines when they immediately appear on news feeds: The source of the headline and the surrounding social cues. Within this defined social context, and drawing on insights from persuasion research (Chaiken, 1987; Chaiken & Maheswaran, 1994; Petty & Cacioppo, 1986), social identity theory (Pereira et al., 2021; Tajfel & Turner, 1979; Van Bavel et al., 2024) and social influence research (Cialdini & Goldstein, 2004; Deutsch & Gerard, 1955; Spears, 2021), this thesis has addressed two primary questions: First, what is the impact of consuming (mis)information in a social context? Second, how do these social influences impact the efficacy of inoculation interventions? The following section summarises key findings in this thesis and subsequently integrates them with existing literature and finally, offers avenues for application and further research needed.

1. A Brief Summary

News Sources Influence Misinformation Susceptibility

Chapter 2 explored whether people are more likely to believe misinformation when it originates from sources that share people's political identity and whether perceptions of source credibility influence misinformation susceptibility. The findings demonstrated that individuals were more likely to perceive misinformation (and factual information) as reliable when it supposedly originated from political ingroup (vs outgroup) sources. Furthermore, the relationship between political similarity and misinformation susceptibility was mediated via perceived source credibility. This effect echoes previous research highlighting that perceptions of information reliability are shaped by political identity factors (Pereira et al., 2021; Pretus et al., 2023; Van Bavel et al., 2024) rather than solely by the accuracy of the information. Chapter 2 further showed that politically congruent sources (as opposed to incongruent) were viewed as less biased and more credible, suggesting an "ideological blind spot" among partisans. This phenomenon aligns with findings in the working paper by Gentzkow et al. (n.d.), who demonstrated that small cognitive biases can lead individuals to disproportionately trust ideologically aligned sources, even when such sources are less accurate than neutral alternatives.

Chapter 3 demonstrated that both source similarity (political congruence) and source credibility had direct and independent effects on the perceived reliability of misinformation. Participants were more susceptible to misinformation when it was attributed to a political ingroup source (compared to outgroup) and when the source was credible (compared to not credible). An interaction effect revealed that similarity did not play a significant role when the source lacked credibility. This suggests that while political alignment can enhance the perceived reliability of information, it cannot offset the absence of (known) source credibility. Although this finding is technically encouraging, it raises concerns about the role of credibility in fostering trust in misinformation, which will be discussed further.

In Chapter 3, source credibility significantly influenced the perceived reliability of factual information, while source similarity did not. This suggests that individuals prioritise credibility cues over political alignment when evaluating factual information (which, in these studies, did not use manipulative or deceptive techniques), highlighting a potential distinction in how true and false information is assessed. One explanation for why individuals may rely more on political identity cues when evaluating misinformation (defined here as misleading information employing deceptive techniques) than factual information (simple, matter-of-fact

statements) lies in the inherent complexity of misinformation. Although prior research has suggested that misinformation is easier to process due to its simpler language (Carrasco-Farré, 2022), assessing its reliability may require navigating multiple layers of evaluation. For example, misinformation is often more emotionally charged (Ecker et al., 2022) and frequently invokes group identity cues (Van Bavel et al., 2021), adding psychological complexity despite its linguistic simplicity. For example, determining the reliability of a headline such as *"MRI studies are completely unreliable"* involves understanding the scientific topic, addressing the implied negative sentiment, and evaluating the claim's plausibility. This multi-faceted evaluation process may introduce uncertainty for the reader (Carrasco-Farré, 2024). In contrast, editorially neutral factual headlines, such as *"Netflix to include mobile games for subscribers",* are typically devoid of emotionally charged language or manipulative elements. Assessing these headlines may involve a more straightforward judgement: Did this happen, yes or no? Individuals may rely more on credibility cues to make quick assessments in such cases. However, when confronted with the uncertainty and ambiguity inherent in misinformation, individuals may turn to social identity cues, such as shared political alignment, as heuristics to guide their judgements. This aligns with Spears (2021), who suggests that, in particular, within situations of uncertainty or gaps in our knowledge, we assume that ingroup members share our perspective and can provide informative guidance.

While the findings suggest that individuals prioritise credibility cues over political alignment when evaluating factual information, this conclusion is nuanced by results from Chapter 2, where source similarity also influenced judgements of factual news. However, an explanation for this apparent inconsistency lies in the differing methodological approaches between Chapters 2 and 3. In Chapter 2, real-world sources were used, which inherently blend credibility and identity cues, making it challenging for individuals to disentangle their effects. For example, a headline from an ideologically aligned trusted source may simultaneously activate credibility and identity heuristics, creating overlapping influences on judgements. As such, it is technically impossible to disentangle the two within the methodological constraints of Chapter 2. In contrast, Chapter 3 explicitly disentangled these factors, enabling a more precise assessment of their independent effects. The findings from Chapter 3 suggest that when credibility and similarity are experimentally isolated, individuals may indeed rely more on credibility cues when evaluating factual information. However, the blended nature of these cues in real-world settings, as explored in Chapter 2, demonstrates that both can shape judgements, particularly when identity cues are salient.

These findings highlight the importance of context in understanding how credibility and identity cues interact. In real-world scenarios - where these cues are often intertwined - identity-based heuristics may influence even factual information judgements. This interplay underscores the need for future research to explore how integrating these cues shapes perceptions in naturalistic environments and how they impact the efficacy of misinformation interventions.

Social Proof Impacts Misinformation Susceptibility

Chapter 4 investigated the role of social proof in shaping individuals' susceptibility to misinformation. Building on social influence theories (Cialdini & Goldstein, 2004; Deutsch & Gerard, 1955; Spears, 2021), the chapter examined both implicit social cues (in this case, comments expressing approval or doubt), and explicit social cues, where participants could see the judgements of "previous" participants. Results showed that social proof significantly influenced participants' judgements, with explicit endorsement cues substantially increasing susceptibility to misinformation ($d = 0.53$). However, while endorsement cues led to increased perceived reliability of misinformation, discrediting cues did not reduce the perceived reliability of misinformation. This may reflect participants' initial judgements: if their gut reaction is to view misinformation as unreliable, there is little room for discrediting cues to further sway their perception.

This interpretation aligns with anchoring research in judgement and decision-making (JDM), which refers to the cognitive bias where individuals rely heavily on an initial piece of information - the "anchor" - when making subsequent judgements. Anchors serve as reference points that influence evaluations, often leading individuals to insufficiently adjust their estimates away from the anchor, even when additional information becomes available (Tversky & Kahneman, 1974). Notably, anchoring effects are most pronounced when the anchor significantly deviates from individuals' initial expectations or judgements (Epley & Gilovich, 2001). In cases where participants already distrust misinformation, discrediting cues may fail to shift perceptions because they confirm rather than challenge initial beliefs. In contrast, endorsement cues significantly diverge from participants' expectations, making them more likely to adjust their judgements. This dynamic is supported by the finding that discrediting cues were more influential for factual information, where participants' initial agreement with the headline's reliability left room for visible doubts to reduce perceived reliability. Indeed, endorsement cues had no effect on judgements of factual information, likely reflecting participants' pre-existing agreement with the factual headline's reliability.

Furthermore, the chapter found that the influence of social proof may operate through its effect on perceived social consensus. Participants who saw endorsement cues were more likely to believe that a broader public consensus supported the misinformation, which heightened their own acceptance of it. This mediating effect of perceived consensus highlights a pathway through which social proof impacts individual judgements: by shifting perceptions of widespread belief, social proof indirectly validates the misinformation in question. Overall, the findings in Chapter 4 demonstrate that social cues, particularly those that imply a collective endorsement, can amplify misinformation susceptibility by shaping perceived consensus. This necessitates testing psychological interventions in conditions where social proof endorses misinformation to examine intervention robustness under these persuasive effects.

Inoculation Interventions: Effective but Vulnerable to Social Dynamics

The two chapters on inoculation explored the effectiveness of inoculation strategies against misinformation, first in cases involving political ingroup sources and then in the presence of persuasive social cues. Chapter 5 examined whether inoculation could reduce susceptibility to misinformation from politically aligned news publishers. The study used the *Bad News* game to "inoculate" participants, simulating misinformation tactics to help participants develop resistance. Results showed that the intervention reduced the perceived reliability of misinformation, increased veracity discernment and reduced susceptibility to misinformation from both political ingroup and outgroup sources. The study further showed that this bias was no longer present on the post-test despite individuals being more likely to fall for misinformation from ingroup (than outgroup) publishers pre-intervention.

Chapter 6 examined whether an inoculation intervention targeting emotionally misleading information could maintain its effectiveness in the face of social proof, represented by visible social cues (e.g., high endorsement from other users). This text-based intervention warned participants about misinformation's common emotional manipulation technique (the emotion fallacy). Results indicated that the inoculation reduced misinformation susceptibility, improving participants' confidence in identifying unreliable information and enhancing their ability to discern misinformation from neutral content. However, social cues - consensus signals suggesting high reliability of misinformation - significantly increased participants' susceptibility to misleading information, even among those who had received the inoculation intervention. This influence of social proof indicates that while inoculation is beneficial, its effectiveness can be compromised by strong social

endorsement signals. Together, these findings underscore that inoculation interventions can bolster resistance to misinformation but may, in some cases, be moderated by social context. Building on the summarised findings here, the following section discusses how this work can be integrated with current literature, both empirical and theoretical.

2. Theoretical Implications

The Role of Peripheral Cues in Misinformation Susceptibility

From a theoretical standpoint, a key contribution of this work is its emphasis on the role of contextual cues in shaping evaluations of information veracity. Dual-process models of persuasion and information processing, such as the Elaboration Likelihood Model (ELM) (Petty & Cacioppo, 1986) and the Heuristic-Systematic Model (HSM) (Chaiken et al., 1996), posit that persuasion occurs through two distinct routes. The central (or systematic) route involves the deliberate and analytical processing of message content. In contrast, the peripheral (or heuristic) route relies on external cues, such as source credibility or consensus indicators, particularly when motivation or ability to process information deeply is low (Chaiken et al., 1996; Petty & Cacioppo, 1986; Petty & Wegener, 1998). These models suggest that within contexts where individuals are distracted or unmotivated, they are more likely to rely on readily accessible contextual characteristics of the source, message, or environment, assessed using simplified cognitive strategies (Wood, 2000). Early experimental work demonstrated that when individuals are not highly motivated to be accurate - such as when they do not need to remember information - they are more influenced by message sources than content (Chaiken, 1980) and more likely to rely on others' judgements than their own evaluations in judgement tasks (Baron et al., 1996a). These foundational studies highlighted the potential conditions under which individuals are likely to rely on contextual cues rather than critically evaluating information - a distinction particularly relevant to today's social media environments, where users are inundated with information amidst numerous distractions (Koessmeier & Büttner, 2021; Matthes et al., 2020).

Consistent with these foundational models, and as summarised in the above section, this thesis's findings underscore the role of peripheral or heuristic cues - including source similarity (shared political identity), source credibility, and social proof - in shaping susceptibility to misinformation. Beyond establishing their influence, several additional insights warrant further discussion, as outlined in the following section.

When Peripheral Cues Matter: Context, Attention, and Heuristic Processing

First, the impact of peripheral cues varied depending on the information being evaluated. While source similarity increased susceptibility to misinformation, it did not impact judgements of factual information. This divergence may reflect differences in the ambiguity of the content, as misinformation may require greater effort to scrutinise and assess its plausibility (Roozenbeek & van der Linden, 2019). Under conditions of ambiguity, people may turn to heuristic cues, such as shared political identity, to resolve uncertainty. This aligns with dual-process models, which argue that heuristic processing dominates when content is ambiguous (Petty & Cacioppo, 1986; Chaiken, 1980) and in response to uncertainty, where individuals seek cognitively efficient ways to resolve ambiguity (Tversky & Kahneman, 1974; Hogg, 2000). This interpretation is further supported by the finding that when a source was described as lacking credibility, similarity no longer influenced participants' judgements of misinformation. Therefore, clear credibility signals may reduce uncertainty, mitigating the need for heuristic shortcuts like social identity cues.

While sources and social proof are theoretically classified as "heuristic cues" (Chaiken, 1980), it cannot be assumed with certainty that they are always processed heuristically. However, the finding that participants with lower Cognitive Reflection Test (CRT) scores (Frederick, 2005) - a potential proxy for heuristic processing - were more influenced by peripheral source cues when assessing the reliability of headlines suggests that this may indeed be the case. That is, as individuals who are more prone to relying on heuristic processing are particularly sensitive to contextual cues, reinforces the notion that these cues are likely processed heuristically.

The findings of this thesis further contribute to the previously mixed literature on the influence of social cues (Ali et al., 2022; Kim, 2018; Mena et al., 2020) on misinformation susceptibility by clarifying the conditions under which social cues, such as exposure to others' judgements, shape misinformation assessments. First, the findings suggest that social cues only impact individuals' judgements of misinformation reliability when they are attended to. This conclusion is based on the observation that participants who failed to correctly answer questions about the valence of comments or other participants' judgements (the attention check) indicating they had not attended to these cues - were less influenced by them. Notably, the proportion of participants who failed this attention check was significantly higher in the social cues study (26%) compared to the source cues study (6%), indicating that source cues may be more automatically attended to than social cues. Users may not consistently read comments or consider social network judgements on social media

platforms, where information is processed rapidly. However, they may be more likely to notice and quickly assess the source of a message, aligning with previous research on the automaticity of source evaluation (Pornpitakpan, 2004).

Second, the findings indicate that social cues tend to influence individuals only when they challenge their automatic assumptions about a message's reliability. Specifically, participants were affected by social cues when those cues endorsed misinformation or discredited factual information. This suggests that social cues have a greater impact when they disrupt pre-existing perceptions of credibility, forcing individuals to reconsider their initial judgements. Third, the thesis demonstrates that the impact of social cues on individual judgements may operate through their effect on perceived consensus. Unlike previous findings, where "likes" alone did not significantly alter perceptions of consensus (Traberg, 2019, reported in Traberg et al., 2024), the current findings suggest that social cues are more likely to influence individuals when they shape broader perceptions of social norms. In line with Cialdini's principles of social influence (2001), social cues appear to exert an effect only when they are readily accessible and perceived as indicative of normative beliefs within the individual's social context. This distinction may explain why some types of social cues (e.g., explicit comments or group judgements) impact individual judgements while others (e.g., "likes") do not. In sum, these findings highlight the differentiated ways in which social cues influence judgements, with attention, accessibility, and their alignment with perceived social consensus emerging as factors in determining their impact.

Building on the findings of this thesis, an integrative framework is proposed to advance our understanding of how these heuristic cues within the social context may influence susceptibility to misinformation. This framework seeks to extend current knowledge by highlighting the interplay between contextual factors and cognitive assessments of information. While not intended to be exhaustive, the framework, titled "A Framework for Understanding the Influence of Message Context on Misinformation Susceptibility," focuses on the immediate "message context." This term refers to the surrounding cues present at the time of information consumption, such as source characteristics and social proof. However, it is acknowledged that the broader social media environment includes additional layers of complexity, including individuals' personal networks, relationships within those networks, and their dynamic interactions over time (Facciani & Traberg, 2024; Galesic et al., 2021; Hewson & Fang, 2024; Vlasceanu et al., 2024).

3. A Framework for Understanding the Influence of Message Context

The findings of this thesis not only highlight the significant role of social context in shaping susceptibility to misinformation but also provide deeper insights into the underlying mechanisms through which this influence occurs. Building on these findings, a framework is proposed to explain how source cues and social proof shape perceptions of misinformation reliability. This framework (Figure 35) identifies two dimensions of contextual influence: source cues and social proof. Source cues include several factors. First, perceived source similarity (**A**) contributes to the perceived credibility of news sources (**C**). This perceived credibility can, in turn, influence the likelihood of judging misinformation as reliable. While this thesis has operationalised source similarity as alignment with political group membership, similarity may also be inferred through numerous source characteristics. Previous research in persuasion and advertising has, for instance, found that similarity based on race (Morimoto & La Ferle, 2008), sexual preference (Atkinson et al., 1981), language (Aune & Kikuchi, 1993) and lifestyle characteristics such as appearance, cultural background, and basic values (Phua, 2016) significantly impact the perceived credibility of a spokesperson. Although classified in the current framework as "official" credibility (**B**), this term refers to credibility indicators that are external and independent of personal or group identity factors (e.g. similarity or shared group membership), which may also influence judgements of misinformation. These indicators are typically based on established, objective reliability markers, such as external credentials, professional qualifications, institutional affiliations, or ratings from trusted third-party organisations. In the context of misinformation, source credibility could also be represented by markers such as fact-checking labels (Jia & Lee, 2024), third-party verifications (Dennis et al., 2023), or endorsements from non-partisan organisations with a history of reliability (Graves, 2018). Notably, in this framework, 'official' credibility does not necessarily signify formal or governmental authority; rather, it refers to perceived legitimacy derived from external markers of expertise or trustworthiness (Hocevar et al., 2017) that may not directly relate to group identity, such as professional qualifications, institutional affiliations, or endorsements from reputable organisations.

With regard to source credibility, it is critical to acknowledge two factors. First, it is difficult to fully disentangle similarity and political identity from 'official' credibility markers. Fact-checking organisations or labels, for example, may be perceived as biased or aligned with particular ideological stances, even when grounded in rigorous and objective

analysis (Hocevar et al., 2017). Non-identity-related credibility markers may still be tied to social identities, such as institutional credentials like university affiliations, which can carry implicit social or ideological connotations (Santoro & Sydnor, 2024). For example, recent research shows that conservatives tend to be more sceptical of academic institutions than liberals (Santoro & Sydnor, 2024). Similarly, though ostensibly neutral, markers like academic titles or awards can trigger social identity associations when linked to institutions or fields perceived to align with certain ideological or cultural values, such as environmental science in climate change debates (McCright et al., 2013).

Second, "official" credibility markers can be manipulated, as bad actors may create counterfeit fact-checking organisations or co-opt seemingly neutral entities to lend false legitimacy to misinformation (Jack, 2017). In fact, it is easier than ever to misuse credibility cues by creating official-looking websites, purchasing verified badges from fake accounts, and employing tactics such as domain spoofing. Domain spoofing involves creating sites with URLs that closely resemble those of legitimate organisations, tricking users into believing the source is credible (Jack, 2017). Similarly, it has become easier to obtain 'verified' badges on platforms such as X (Center for Countering Digital Hate, 2023), which can mislead users into perceiving greater credibility than actually exists. Indeed, "verified" credentials can be misused to persuade and mislead, as messages from sources with verified credentials are perceived as more credible, regardless of whether the credential is context-relevant, creating a so-called 'virtual lab coat effect' (Geels et al., 2024). The distinction between "official" credibility and identity-based cues is critical. While political or identity-based cues influence perceived credibility through social or psychological alignment, "official" credibility is grounded in externally validated markers that may transcend group identity. Nevertheless, the perception of such markers can still be influenced by contextual and individual factors, including pre-existing trust in the institution or entity providing the endorsement.

Moving on to social proof, this operates through explicit cues (**D**), such as visible group endorsements or vote counts, and implicit cues, such as supportive comments on social media (**E**). These signals both contribute to meta-judgements about perceived public consensus (**F**), shaping how individuals assess the reliability of misinformation. The influence of social proof lies not only in its direct effects on individual judgements but also in its capacity to shape meta-judgements about perceived social consensus. For instance, supportive comments and explicit consensus indicators may convey a sense of collective validation, reinforcing the perception that the information is widely accepted. These cues can heighten the perceived reliability of misinformation, in particular, when individuals might

otherwise approach the content with scepticism. This process aligns with social influence principles, where perceived social consensus is a powerful driver of individual beliefs and behaviours (Clarkson et al., 2013; Gardikiotis et al., 2005; Lewandowsky et al., 2019). As noted in the framework, based on findings in this thesis, social proof may operate by impacting perceived public consensus. However, it is also possible that there are alternative pathways not identified in the current work through which social proof impacts misinformation judgements. Indeed, it may be that social consensus perceptions are especially powerful if they are based on group identity cues. For example, research on personal network composition has shown that individuals' trust in information and their susceptibility to misinformation is influenced by the homogeneity of their social circles, where alignment in beliefs and group identity reinforces shared norms and amplifies the perceived reliability of the information (Facciani & Traberg, 2024).

This framework emphasises the interplay between individual cognitive assessments and socially constructed influences, offering a more nuanced understanding of how misinformation takes hold in digital environments. However, while the factors identified in this framework contribute towards understanding the immediate *message* context on misinformation susceptibility, it is important to acknowledge that the broader context in which information is consumed extends far beyond these immediate cues, as highlighted by other information models. For example, the Identity-Based Model of Belief (Van Bavel et al., 2024; Van Bavel & Pereira, 2018) highlights how the interplay between identity and accuracy goals can shape individuals' informational environments. Social identity goals drive individuals to seek information that aligns with their group's beliefs and reinforces their identity. In contrast, accuracy goals motivate them to pursue information grounded in evidence and reliability. Importantly, these goals can interact in ways that influence behaviour and decision-making; for instance, certain social identity groups, such as scientists or journalists, may prioritise accuracy goals as part of their group norms. Indeed, the informational context can amplify identity-driven processing, fostering environments where in-group norms strongly influence whether it is deemed appropriate to believe or share misinformation - particularly when the content pertains to group-relevant topics (Rathje et al., 2021; Sternisko et al., 2023). Partisan motivations further intensify this dynamic; empirical evidence indicates that the political alignment of headlines exerts a stronger effect on belief and the intention to share news than their actual accuracy (Schwalbe et al., 2024). The interplay between internal social identity and accuracy cues, alongside the broader information environment within which the message context could be situated, enriches our

understanding of misinformation susceptibility by framing individual evaluations within a larger ecosystem of group influences and cognitive priorities, thereby complementing the message-contextual framework presented in this thesis.

Furthermore, the BENDING model (Vlasceanu et al., 2024) conceptualises belief systems as dynamic networks of interrelated elements (evidence, beliefs, and perceived norms). This model emphasises the multidimensional nature of beliefs, illustrating how they are shaped and reinforced by both cognitive structures and social interactions within communities, often making them resistant to change. Unlike linear frameworks such as the ELM (Petty & Cacioppo, 1986), the BENDING model captures the interdependencies between belief elements, showing how changes to one aspect - such as a piece of evidence or a perceived norm - can ripple across the network, altering related beliefs and behaviours. From this perspective, correcting misinformation can fail if the surrounding norms and related beliefs remain unaddressed. The model also emphasises the role of social norms, showing that perceptions of consensus can significantly influence belief updates, especially when aligned with social identities. This perspective complements the framework proposed in this thesis, extending it by situating individual judgements within broader networks of social and cognitive influences.

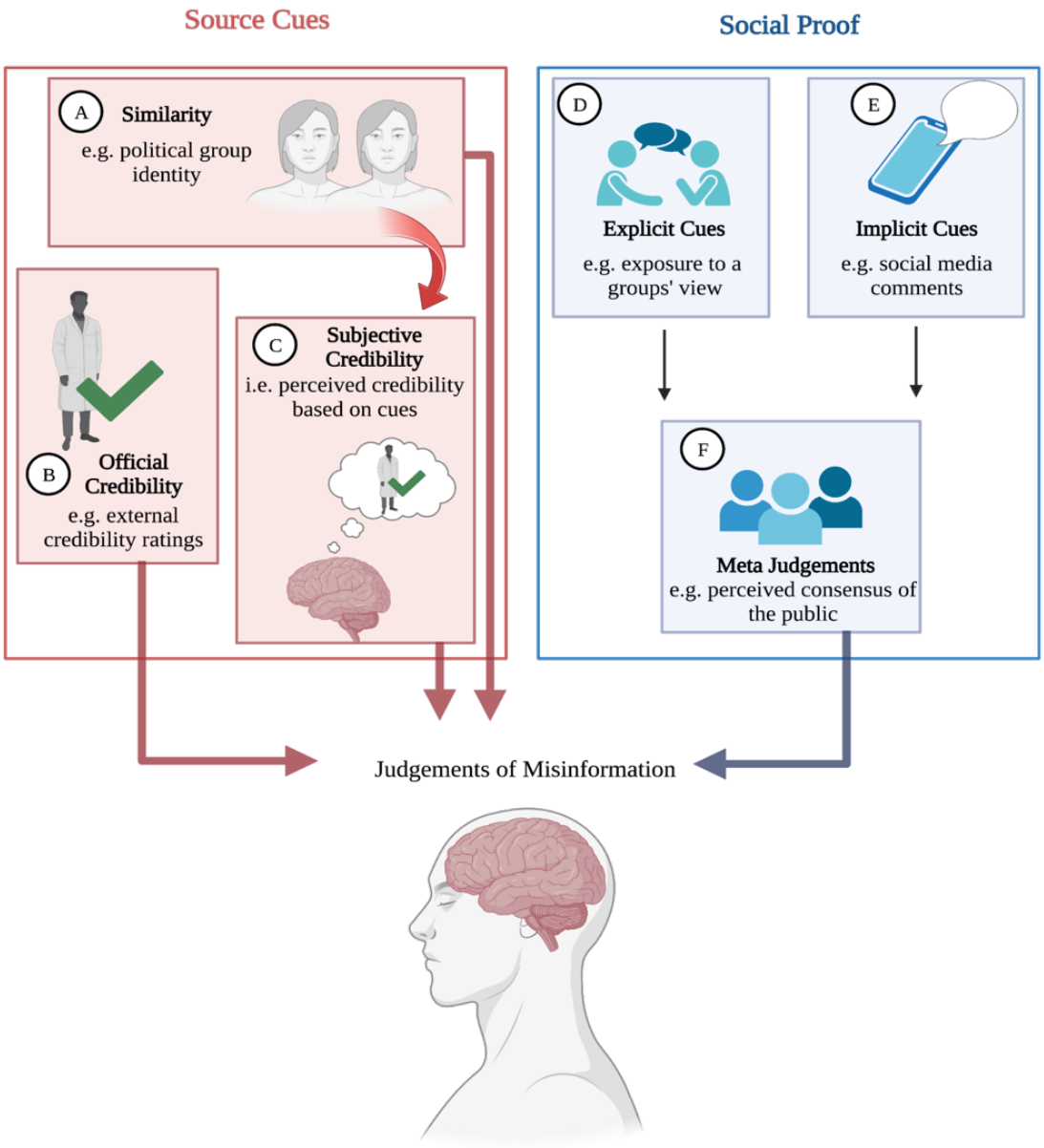
Moreover, the integrative framework proposed by Galesic et al. (2021) provides a quantitative approach that incorporates both cognitive and social network dynamics in belief change processes. This model conceptualises belief dynamics as a system of interconnected individual and social components. Individual beliefs are represented as networks of probabilistic states influenced by personal cognitive structures and perceived social fields, while social networks dynamically adapt to reflect evolving belief systems. For example, individuals might resolve dissonance between personal and social beliefs by modifying their own beliefs or altering their social connections, leading to phenomena such as clustering, polarisation, or consensus. This framework differs from the BENDING model and the framework presented in this thesis by explicitly integrating the structural and process components of belief dynamics. It captures how individual belief changes (e.g., shifts in the degree of confidence or uncertainty assigned to a belief) are influenced by and simultaneously reshape social network structures (e.g., the formation or dissolution of connections based on belief alignment). For instance, it models how individuals might strengthen ties with like-minded peers or disconnect from those with opposing beliefs, creating clusters that amplify misinformation or foster polarisation. This bidirectional

interaction between belief updating and social network adaptation provides a nuanced understanding of how beliefs evolve within interconnected systems.

Together, these models highlight the importance of situating the findings of this thesis within a broader ecosystem of social and cognitive influences, as the dynamics of misinformation extend across multiple layers of context - from immediate message cues to larger-scale social systems and belief networks. While the framework presented in this thesis synthesises theoretical insights and empirical evidence uncovered in this work, it is intended as a conceptual tool for advancing understanding rather than as a fully validated model. Future research is needed to empirically test the proposed pathways, evaluate their robustness across different contexts, and identify alternative mechanisms not captured here. Additionally, it is important to reflect on further insights discovered in this thesis that go beyond the framework's scope, which are discussed in the section below.

Figure 35

A Framework for Understanding the Influence of Message Context on Misinformation Susceptibility



Note: (A) Source similarity influences misinformation susceptibility (Chapters 2 and 3). B) Official credibility of sources has an independent direct effect on misinformation susceptibility (Chapter 3). (C) Source similarity predicts the perceived credibility of sources, which in turn influences misinformation susceptibility (Chapter 2). (D) Explicit social proof influences meta-judgements (perceptions of a more general consensus in the wider public) (Chapter 4). (E) Implicit social proof influences meta-judgements (Chapter 4). (F) Meta-judgements influence individual judgements of misinformation (Chapter 4).

4. The Complexity of Social Cues in Misinformation

While the framework presented above conceptualises the influence of message context on individual judgements of misinformation, additional complexities regarding the potential application of these findings warrant discussion: the nuanced findings in relation to social cues, the limitations of non-political content, and the role of normative influence in online environments.

First, this thesis found that the impact of social cues was pronounced when cues endorsed misinformation as valid or discredited factual information. These findings underscore the risk of social cues in online environments, where misleading social signals may override individuals' initial, more accurate intuitive judgements. However, while social cues were found to exacerbate susceptibility to misinformation in this work, other research suggests they may also represent opportunities for intervention. Pretus et al.(2024) introduced the concept of the "Misleading count," an identity-based intervention that leverages normative social signals to combat misinformation. Their research demonstrated that displaying a "Misleading count" (the number of users tagging content as inaccurate) reduced misinformation sharing by 25%, particularly when the count reflected ingroup norms. This suggests that normative social influence can be a powerful tool to correct inaccuracies, especially in polarised contexts. Indeed, recent scholarship suggests that displaying factual corrections or dissenting voices alongside misinformation could disrupt false perceptions of consensus (Walter et al., 2020). However, the double-edged nature of social cues remains: while they can foster accuracy in ingroup settings, they can also perpetuate inaccuracies when ingroup norms are themselves flawed. For instance, work by Allen et al. (2022) suggests that partisans refrain from 'fact-checking' ingroup members, an important consideration when seeking to apply such measures online.

Second, it is important to consider the limitations of the context in which the effects of these contextual cues were observed. The content examined in this thesis was intentionally politically "neutral", limiting the extent to which content-based partisan biases could come into play. The reason for this was to isolate the effects of contextual cues, as previous research often either conflates partisan sources and partisan content (Pennycook & Rand, 2019; Stein et al., 2024) or only examines content bias (Grady et al., 2021). As such, this thesis can only conclude on the effects of sources and social cues on non-political misinformation. While an important limitation to acknowledge, not all information on the internet is immediately political. For example, individuals frequently engage with

(mis)information about science (Cacciatore, 2021; West & Bergstrom, 2021), health recommendations (Borges do Nascimento et al., 2022; Nan et al., 2022; Peng et al., 2023), consumer products (Fong et al., 2021) and lifestyle advice (Diekman et al., 2023; Topham & Smith, 2023), where the primary context may not invoke partisan divisions. It is, therefore, important to understand whether partisan biases go above and beyond those invoked by the ideological congruence of content (Carnahan et al., 2022). Interestingly, some work has shown that in relation to controversial news (not fake news), people may still pay attention to source cues, even for partisan news. Kim (2016) found that participants rated news as less biased when it was attributed to an ingroup partisan source compared to an outgroup source, regardless of its content valence (congruent vs incongruent with their political identity).

Third, this work can only comment on the role of informational influence – which occurs when individuals conform because they perceive others as having more accurate information (Deutsch & Gerard, 1955). It cannot comment on normative influence, which stems from a desire for social approval or a fear of disapproval, often resulting in conformity regardless of private disagreement (Asch, 1956; Cialdini & Goldstein, 2004). Importantly, therefore, while this work primarily alludes to the impact of social cues on informational influence - given that judgements and responses were anonymous - normative pressures could be even more pronounced in real-world online environments. On social media, where people's actions and endorsements are publicly visible, the social consequences of sharing, liking, or commenting may be far greater. The visibility of these behaviours may intensify normative pressures, pushing individuals to conform to group norms even when they privately disagree (Price et al., 2006).

Despite this work primarily focusing on individual judgements rather than sharing or behaviours related to normative influence, this work carries important implications for social media, where platforms often amplify extreme or emotionally charged content due to engagement-based algorithms, which may expose users disproportionately to polarised, vocal minorities (Rathje et al., 2024; Shin, 2024). When individuals mistake the views of these small but active groups for widespread societal consensus, the consequences of misinformation may be exacerbated. Recent research on echo chambers and filter bubbles (Bakshy et al., 2015; Cinelli et al., 2021) supports the notion that such environments reinforce pre-existing beliefs and limit exposure to diverse perspectives. Although the results of this thesis do not speak directly to this issue, given it did not manipulate the identities of the sources of social cues, the feedback loop between social identity and perceived consensus may become even stronger as individuals primarily interact with and validate the views of

like-minded others. Theoretically, ingroup endorsement should hold greater sway due to shared identity and perceived trustworthiness (Spears, 2021). Indeed, in the work by Pretus et al. (2024), the 'Misleading count' was more effective when it reflected political ingroup norms as opposed to general users in polarised (U.S.) contexts. Interestingly, this was not the case for less politically polarised contexts (the U.K.). This suggests that social cues from ingroups may be more powerful. However, future research is needed to investigate whether such ingroup cues might also exert disproportionate influence when they *support* inaccurate or misleading information, thereby exacerbating the spread of misinformation.

These insights underscore the limitations of interventions that rely solely on improving critical thinking or analytical skills (Pennycook et al., 2021), suggesting that interventions must also address the social dynamics and identity-based heuristics that shape susceptibility. Indeed, norm-based interventions have been shown to be significantly more effective in reducing the sharing of partisan misinformation than identity-neutral accuracy nudges (Pretus et al., 2024). Thus, the thesis calls for a reconceptualisation of misinformation interventions, proposing a shift toward strategies that integrate social context awareness and directly address identity-based cues. Implications for inoculation interventions, which were specifically tested in this thesis, are discussed below.

5. Advancing Inoculation Through Socially Interactive Interventions

Inoculation theory posits that preemptive exposure to weakened forms of misinformation and counterarguments can "inoculate" individuals against future misinformation (McGuire, 1964; Roozenbeek & van der Linden, 2019). Previous empirical research has demonstrated the efficacy of inoculation interventions in increasing resilience against misinformation across various contexts, from climate misinformation (Spampatti et al., 2024; van der Linden et al., 2017) to vaccine misinformation (van der Linden et al., 2021) and pro-Kremlin misinformation (Ziemer et al., 2024) and extremist disinformation (Lewandowsky & Yesilada, 2021), in studies using both fictitious (Basol et al., 2020) and real-world misinformation (Roozenbeek, Traberg, et al., 2022). Inoculation interventions have been shown to be effective across cultural contexts (Basol et al., 2021; Roozenbeek et al., 2020), using pre-post test designs (Roozenbeek & van der Linden, 2019) and randomised-control studies (Basol et al., 2020), using digital games (Roozenbeek & van der Linden, 2020; Roozenbeek & van der Linden, 2019), board games (Roozenbeek & van der Linden,

2018), text-based interventions (Traberg et al., 2024) , and videos (Roozenbeek, van der Linden, et al., 2022), with effects lasting several months (Maertens et al., 2021). While the work in this field is promising and already has important policy implications (Traberg et al., 2023), evaluating the effectiveness of psychological interventions requires testing them in conditions that replicate the persuasive cues individuals are likely to encounter in real-world social settings where misinformation is prevalent. To the best of my knowledge, this thesis is the first to examine the efficacy of inoculation interventions in the context of both source and social cues. The findings from the inoculation chapters suggest that inoculation can indeed shield individuals from misinformation disseminated by ingroup sources and in the face of social proof endorsing its reliability. However, despite the observed effectiveness of inoculation interventions in these controlled settings, these effects should still be interpreted with caution for the reasons outlined below.

First, the news publishers were U.S. news outlets, but the sample was not U.S.-limited. Findings in Chapter 2, which limited participants to U.S. participants, showed that ideological source biases were very potent here. As such, to more robustly justify the conclusion that these interventions work against political ingroup source misinformation, it is necessary to test with a geographically representative sample. As discussed in the previous section, partisan biases in information judgements exist beyond source effects (Van Bavel & Pereira, 2018), and partisan news sources tend to publish content that aligns with the views of their target audience (Budak et al., 2016; Kim, 2016; Larcinese et al., 2011). While some work has shown that inoculation interventions can improve participants' ability to spot manipulation techniques commonly used in political misinformation (Roozenbeek & Linden, 2020), future research is needed to examine inoculation interventions in the face of partisan misinformation from partisan sources.

Second, although this work showed that inoculation remained effective even in the face of social proof, even inoculated individuals were impacted by social proof. In fact, the main effect of social proof was larger (d = 0.44) than the main effect of inoculation (d = 0.23). Importantly, real-world social settings are far more complex, with individuals pursuing multiple, often conflicting goals - such as seeking social validation (Diefenbach & Anders, 2022), maintaining group cohesion (Liu & Hu, 2024), or expressing ideological identities (Rathje et al., 2021). These dynamics create additional challenges that may influence the success of inoculation strategies. To ensure interventions remain effective in such multifaceted environments, future research should explore how inoculation can be adapted to account for these social complexities, ultimately making the interventions more realistic,

scalable, and impactful. Nevertheless, the combined influence of source effects, social proof, and the broader social context on misinformation susceptibility highlights the need to develop more targeted and nuanced inoculation strategies. Building on the findings of this thesis, the following section explores novel avenues to improve the effectiveness of psychological interventions to reduce misinformation susceptibility.

Beyond Content-Focused Inoculation Strategies

The findings of this thesis offer several important insights for designing more effective interventions against misinformation, particularly on digital and social media platforms where social and source cues are omnipresent. The impact of source similarity, credibility, and social proof on misinformation susceptibility suggests that interventions must move beyond content-focused strategies and incorporate social context into their design instead. Current inoculation interventions generally expose individuals to weakened forms of misinformation and equip them with counterarguments (Basol et al., 2020, 2021; Compton & Pfau, 2005; Lewandowsky & van der Linden, 2021; Maertens et al., 2021; Roozenbeek, Traberg, et al., 2022; Spampatti et al., 2024; Traberg et al., 2024). However, given the complex social nature of misinformation consumption, future inoculation strategies could be designed to *specifically* counter the biases introduced by source effects and social proof. This necessity is supported by Chapter 6 identifying a *third-person consensus effect*, where participants were more likely to recognise the influence of social cues on others than on themselves, suggesting a potential blind spot in their awareness of their own susceptibility to social influence.

Building on McGuire's (1961) distinction between *active* refutation (where individuals are engaged in the generation of counterarguments against messages) and *passive* refutation (where individuals are provided with counterarguments against messages), future work could explore a more active involvement of individuals in the social influence process. McGuire's active vs. passive inoculation (McGuire, 1961) distinction emphasises the value of involving individuals in generating their own refutations, suggesting that active participation may have longer-lasting effects. Just as the *Bad News* game places individuals in the shoes of a misinformation producer to more actively teach players about the manipulation strategies used by misinformation producers (Roozenbeek & van der Linden, 2019), future games could place individuals in a scenario in which they may have to avoid influence or actively attempt to influence others. For example, inoculation games or simulations might include scenarios where misinformation is attributed to ideologically similar sources or heavily

endorsed by other actual (or agent-based) participants, allowing users to practice critical thinking within social contexts more resemblant to real communication.

One way to achieve this could be through an inoculation intervention that actively engages individuals in recognising their susceptibility to contextual biases, involving individuals not in creating *misinformation* but in creating or directly experiencing *social influence effects*. For example, Asch's (1956) conformity experiment – which demonstrated how group consensus can lead individuals to override their independent judgement in simple perceptual tasks (Baron et al., 1996b) - could be adapted to expose individuals to their own susceptibility to social influence. Transforming this paradigm into an inoculation tool could involve participants learning about their vulnerabilities to social influence (representing the 'threat' component of inoculation) and actively experiencing how social cues may influence their cognition. This could be combined with an explanation of how this influence works and how to withstand it (representing the 'cognitive' component). For example, participants might take part in an estimation task or an adaptation of the BLAP task (Marks et al., 2019) where a group's consensus may influence their initial judgements, or alternatively, they might play the role of a confederate, helping to sway others' judgements. By engaging individuals in this way, the intervention could make the mechanisms of social influence more tangible, potentially fostering greater self-awareness and evaluation of social cues in the real world, representing a form of experiential learning (Kong, 2021; Morris, 2020). Alternatively, a more *passive* approach could involve simply viewing the original study footage (vs a control video), which demonstrated the paradigm's effects on individual judgements. This type of intervention could further be adapted for digital contexts by using a simulated digital social media environment in which participants could evaluate the reliability of news articles within a mock social media platform, where the visibility of social endorsements varies or where they interact with other real or agent-based participants.

This type of paradigm could be used as both a measurement tool and intervention: By observing how their judgements shift in response to social cues, participants may gain a deeper understanding of how social proof operates in digital environments. To capture the effects of these types of intervention within a context resembling real social life, participants could complete a follow-up diary study where they log and evaluate news articles they encounter on social media. For each article, participants could provide a screenshot of the headline and surrounding social cues (showing likes, comments, or endorsements) and record their perceived reliability of the content. They could also be asked to report their level of 'closeness' to the sharer of the information. Of course, this would directly prompt individuals

to reflect on the relationship between themselves and the sharer prior to judging the article's reliability, so this approach has limitations. However, this diary approach would allow researchers to analyse whether social cues influenced participants' judgements in a more ecologically valid way. Such a diary component may further deepen participants' self-awareness about how socio-contextual biases shape their perceptions over time.

Another way of including more social interactive elements within a form of inoculation intervention is using multi-agent games. Hertz et al. (2025) highlight the potential of multi-agent games in capturing the dynamic interplay between individual cognition and social influence, particularly in contexts where social norms shape decision-making. These types of simulations could enable participants to experience firsthand how their actions and judgements influence - and are influenced by - others within the network. Multi-agent frameworks could be used to examine the iterative nature of social learning, showing how group dynamics evolve through feedback loops of collective behaviour and individual adaptation (Hertz et al., 2025). By engaging participants in environments where they must simultaneously interpret social signals and make strategic decisions, these simulations may more closely replicate the complexities of real-world social networks, such as those found on digital platforms. In this way, multi-agent games could function not only as interventions but also as a tool for understanding how misinformation spreads within social groups.

Another promising avenue to explore is the use of interactive digital games rooted in storytelling, which leverage the persuasive (Bullock et al., 2021) and memorable qualities (Graeber et al., 2024) of narratives to pre-emptively warn players about the threat of misinformation and teach them how to refute it. The first benefit of embedding learning about misinformation within a game rooted in a compelling story is that individuals are significantly more likely to remember them than mere facts or statistics (Graeber et al., 2024), potentially because they engage both cognitive and emotional processing (Green & Brock, 2000; Slater & Rouner, 2002). For example, Graeberg et al. (2024) found that statistical information loses approximately 73% of its influence on beliefs within a day, whereas the influence of a narrative diminishes by only 32% over the same period. Incorporating narratives into inoculation interventions could potentially extend the duration of their effects beyond the current benchmark of up to three months (Maertens et al., 2021).

Second, stories can avoid being overly didactic – a limitation of many misinformation interventions that primarily appeal to those already motivated to learn about misinformation (Basol et al., 2021; Roozenbeek & van der Linden, 2020; Roozenbeek & van der Linden, 2019). Stories can uniquely engage individuals by inviting them to explore complex issues

indirectly (Passon, 2019) and by fostering identification with characters, making the information they convey more personally relevant and impactful. This principle can be integrated into game design, where players become active participants in an unfolding narrative. Instead of presenting overt lessons, games rooted in storytelling can create immersive experiences that allow players to draw their own conclusions, encouraging critical thinking about misinformation and social influence. By weaving lessons into the gameplay itself, stories in games could more subtly challenge biases and assumptions, potentially making the learning process more engaging and less confrontational than traditional inoculation game methods.

An example of such an interactive game rooted in storytelling is *Solomon's Secret* (see Figure 36), a digital game (currently in development) that illustrates the dynamics of misinformation and social influence in an engaging, narrative-driven format (Traberg, 2024). Set in the nostalgic and socially charged atmosphere of a high school reunion turned murder mystery, players are immersed in a series of psychological games and puzzles. The game leverages storytelling to create a compelling context where players learn to identify manipulation tactics and evaluate source credibility through dynamic interactions with a diverse cast of characters. In the game, players observe how source credibility evolves over time based on their choices and interactions. This mechanic not only teaches players about how misinformation can spread but also encourages critical reflection on their own susceptibility to social influence. For example, players navigate group dynamics to sway opinions or avoid suspicion, mirroring real-life scenarios of social manipulation and misinformation spread. As such, the narrative invites players to explore complex social and cognitive dynamics in a way that feels engaging rather than instructive.

Figure 36

Screenshots from Solomon's Secret Game Trailer

Note: Top left: Players must identify the deceptive technique embedded in characters' through gameplay. Top right: Through interactions with characters, players learn the dynamic nature of a source's credibility, made up of expertise and trustworthiness. Bottom: Players collect clues throughout the game, which help them learn about the potential motives and actions of the other characters.

Finally, a way of advancing inoculation interventions is to combine them with other interventions or personalise them to more precisely target individual vulnerabilities. For example, recent work suggests that the most effective countermeasures to misinformation may be those that combine multiple intervention types (Pennycook et al., 2024). For example, across 5 studies assessing the efficacy of misinformation interventions, Pennycook et al. (2024) found that while inoculation significantly improved participants' ability to identify emotional manipulation, truth discernment was improved only when combining the inoculation with an accuracy nudge. In line with the work presented in this thesis, these findings underscore that the most effective countermeasures against misinformation may be those that adopt a multifaceted approach, integrating cognitive, emotional, and social dimensions to address the complexity of misinformation susceptibility. Furthermore, recognising the role of social identity in shaping perceptions of information reliability,

interventions could be tailored to resonate with specific audience identities. Just like precision inoculation has gained traction in biological vaccine research (Jia et al., 2020), misinformation interventions could similarly be tailored to the individual's unique vulnerabilities by considering factors such as cognitive biases, prior knowledge, social identity, and the type of misinformation most likely to influence an individual. This approach may enhance the effectiveness of interventions by aligning strategies with the specific characteristics and needs of the target audience. Taking this notion even further to platform implementation, algorithms could adjust the presentation of prebunking efforts to reflect in-group norms, leveraging insights from Pretus et al. (2024) to enhance message receptivity. By personalising interventions this way, platforms could foster greater trust and effectiveness in countering misinformation.

The perspectives and potential future avenues discussed in this section underscore the importance of adapting inoculation strategies to address the challenges posed by source effects, social proof, and real-world social complexities. While traditional inoculation interventions have demonstrated effectiveness, they may face certain limitations when applied to the nuanced dynamics of digital misinformation. Integrating active engagement through simulations, multi-agent games, or storytelling-based games offers a potential avenue to enhance the effectiveness, scalability, and longevity of psychological interventions aimed at reducing misinformation susceptibility.

6. Limitations and Future Avenues

While this thesis provides insights into the role of social and source cues in misinformation susceptibility and the efficacy of inoculation interventions, several limitations should be acknowledged.

A critical limitation lies in the experimental setting, which, while controlled, cannot capture the complexity and interactivity of real-world social media environments. Social media and face-to-face interactions carry real social consequences, such as reputational risks, social validation, or ostracis - factors absent in these controlled studies. Future research should strive to replicate the ecological validity of social media platforms by conducting studies within actual digital ecosystems or advanced simulated environments. For example, recent work by Vlasceanu & Coman (2021) demonstrates how real communication shapes belief change. Their study on COVID-19 misinformation found that dyadic conversations led individuals to align their beliefs with their conversational partners. Integrating such real-time,

interactive dynamics into future studies could yield richer insights into how social interactions impact misinformation susceptibility and intervention efficacy.

Second, this thesis does not directly address how group membership influences susceptibility to social proof. While the findings highlight the effects of social proof and source similarity on reliability judgements, they do not explore whether individuals are more likely to trust social cues from ingroup members than outgroup members. Theoretically, ingroup endorsement should hold greater sway due to shared identity and perceived trustworthiness (Spears, 2021). However, empirical evidence on how this dynamic operates in misinformation contexts remains sparse. Investigating these dynamics in digital settings, where group identity is often accentuated, could clarify whether ingroup social cues exert a disproportionate influence compared to those from outgroups. This thesis further conceptualised group identity primarily through the lens of political affiliation. However, group identities extend far beyond political boundaries, encompassing various social, cultural, professional, and personal dimensions (Crocetti et al., 2023; Tajfel & Turner, 1979). Future research should investigate whether the source effects observed in this thesis also apply to other types of group identities. Exploring these alternative group contexts could provide a more nuanced understanding of how identity shapes perceptions of credibility and susceptibility to misinformation.

Furthermore, while this thesis focuses on textual misinformation, it does not address the potential influence of multimodal cues such as images, videos, or audio, which are increasingly common in digital misinformation (Hameleers et al., 2020). For instance, emotionally charged imagery or video clips may amplify the persuasive power of misinformation (e.g., through emotional resonance) in ways that text-based cues cannot fully capture. Indeed, deepfakes - artificial but hyper-realistic video, audio, and images created by algorithms - significantly complicate the misinformation landscape (Mustak et al., 2023).

In addition, the thesis does not account for the role of platform algorithms that prioritise engagement and potentially amplify polarising or misleading content. These algorithms may not only shape users' exposure to social and source cues but may also influence the perceived prevalence of misinformation through repeated exposure (Nadarevic et al., 2020). Investigating the interplay between algorithmic mechanisms and human susceptibility to misinformation would provide deeper insights into real-world dynamics.

Relatedly, the studies in this thesis focus on immediate judgements of information but do not examine how the influence of social and source cues evolves over time. Misinformation susceptibility may fluctuate based on repeated exposure, changing social

norms, or evolving trust in sources. Longitudinal studies would be necessary to understand these temporal dynamics.

Furthermore, while the findings are relevant to digital misinformation contexts, they do not fully incorporate specific features of social media platforms, such as trending hashtags, targeted advertisements, or personalised news feeds, which may interact with social cues and influence user judgements. Future work could attempt to simulate these features to better replicate the influence of digital environments. Furthermore, the research primarily examines cognitive and social factors but does not extensively address the role of emotions, such as fear or anger, which are often integral to the spread and acceptance of misinformation (Carrasco-Farré, 2022; McLoughlin et al., 2024).

Another limitation of this work is that this thesis assessed misinformation susceptibility using self-reports of perceived reliability, not actual behavioural engagement on social media. Furthermore, in limiting the dependent variables to information reliability measures, it did not look at the actual impact of misinformation or social cues on beliefs. It did not examine the longitudinal effects of social context nor inoculation on any outcome measures. Defining misinformation as misleading information using deceptive techniques further limited the scope of conclusions that can be drawn from this work. In addition, several findings are based on correlational evidence (e.g., the mediation models) and as such, causal evidence is needed to validate these effects.

The diversity of the sample populations used in this research presents another key limitation in terms of generalisability. While the studies drew from a range of backgrounds, they did not fully represent the global population. Cultural and demographic variations likely influence how individuals interpret social and source cues, as well as how they respond to interventions and, indeed, misinformation. Future research should aim to include more diverse samples to ensure that interventions are universally applicable and culturally adaptive.

Addressing these limitations by incorporating real-world interactivity, exploring group dynamics in greater detail, and enhancing sample diversity will strengthen our understanding of the social cognition of misinformation and improve the design of effective intervention. Despite these limitations, this thesis contributes to current literature on misinformation susceptibility in several key ways, as outlined in the conclusion below.

8. Conclusion

This thesis explored the cognitive and social mechanisms underpinning susceptibility to misinformation, emphasising the role of source and social cues and their implications for designing psychological interventions. A series of experimental studies demonstrated the influence of politically aligned (vs misaligned) sources and social proof on perceptions of misinformation reliability and highlighted the potential limitations of current inoculation strategies when faced with complex social environments. Supporting perspectives outlined by social psychologists (Van Bavel et al., 2024; Van Bavel & Pereira, 2018; Vlasceanu et al., 2024), this work highlights that purely cognitive accounts of misinformation susceptibility (Pennycook & Rand, 2019) are insufficient to explain why people judge misinformation as reliable. Therefore, echoing recent intervention evidence (Pretus et al., 2024), purely cognitive interventions such as accuracy nudges (Pennycook et al., 2021) will be insufficient to address the social nature of information consumption.

A contribution and potential application of this work lies in advancing inoculation theory by suggesting the incorporation of socially interactive elements. Building on existing frameworks, this thesis argued for extending inoculation strategies beyond content-focused approaches to include simulations, interactive digital environments, and storytelling-based games. Such interventions could not only engage individuals more actively but also replicate the nuanced social dynamics of misinformation spread, fostering greater resistance to contextual biases. For example, interactive games like *Solomon's Secret* demonstrate how storytelling can bridge cognitive and social engagement, potentially making interventions memorable and scalable to populations not necessarily interested in learning about misinformation.

The implications of this research are both theoretical and practical. Theoretically, the thesis bridges cognitive and social psychological perspectives, presenting an integrative framework for understanding how context, including source similarity, credibility, and social proof shape misinformation susceptibility. Practically, it offers potential directions for researchers, platform designers and intervention or game developers to develop tools and countermeasures that account for both the individual and the social context.

In conclusion, this thesis calls for a paradigm shift towards viewing misinformation susceptibility as a socially developed process requiring socially interactive, contextually situated and personalised approaches that address the complexities of digital environments. Looking forward, future research should seek to develop and empirically test such interactive

strategies across diverse populations and examine their long-term efficacy in mitigating misinformation. Additionally, further exploration of the interplay between normative and informational social influences could refine our understanding of how to counteract the spread of misinformation in polarised or ideologically charged contexts.

In an era where misinformation undermines democratic processes, public health, and societal trust, this thesis contributes to a growing body of research aimed at fostering resilience against its pervasive influence. By advancing the theoretical foundations and applications of inoculation theory, it further provides a roadmap for combating misinformation in an increasingly complex and socially interconnected media landscape.

9. References

Abrams, D., Wetherell, M., Cochrane, S., Hogg, M. A., & Turner, J. C. (1990). Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization. *British Journal of Social Psychology*, *29*(2), 97–119. https://doi.org/10.1111/j.2044-8309.1990.tb00892.x

Acerbi, A., Altay, S., & Mercier, H. (2022). Research note: Fighting misinformation or fighting for information? *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-87

Albarracín, D., Kumkale, G. T., & Vento, P. P.-D. (2017). How people can become persuaded by weak messages presented by credible communicators: Not all sleeper effects are created equal. *Journal of Experimental Social Psychology*, *68*, 171–180. https://doi.org/10.1016/j.jesp.2016.06.009

Ali, K., Li, C., Zain-ul-abdin, K., & Zaffar, M. A. (2022). Fake news on Facebook: Examining the impact of heuristic cues on perceived credibility and sharing intention. *Internet Research*, *32*(1), 379–397. https://doi.org/10.1108/INTR-10-2019-0442

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *The Journal of Economic Perspectives*, *31*(2), 211–235.

Allen, J., Martel, C., & Rand, D. G. (2022). Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–19. https://doi.org/10.1145/3491102.3502040

Allen, J., Watts, D. J., & Rand, D. G. (2024). Quantifying the impact of misinformation and vaccine-skeptical content on Facebook. *Science*, *384*(6699), eadk3451. https://doi.org/10.1126/science.adk3451

Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on Misinformation: Conceptual and Methodological Challenges. *Social Media + Society*, *9*(1). https://doi.org/10.1177/20563051221150412

Altay, S., Berriche, M., Heuer, H., Farkas, J., & Rathje, S. (2023). A survey of expert views on misinformation: Definitions, determinants, solutions, and future of the field. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-119

Anderson, B. A. and M. (2021, April 7). Social Media Use in 2021. *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgements. In *Groups, leadership and men; research in human relations* (pp. 177–190). Carnegie Press.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://doi.org/10.1037/h0093718

Aslett, K., Guess, A. M., Bonneau, R., Nagler, J., & Tucker, J. A. (2022). News credibility labels have limited average effects on news diet quality and fail to reduce misperceptions. *Science Advances*, *8*(18), eabl3844. https://doi.org/10.1126/sciadv.abl3844

Atske, S. (2021). News Consumption Across Social Media in 2021. *Pew Research Center's Journalism Project*. https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/

Aune, R. K., & Kikuchi, T. (1993). Effects of Language Intensity Similarity on Perceptions of Credibility Relational Attributions, and Persuasion. *Journal of Language and Social Psychology*, *12*(3), 224–238. https://doi.org/10.1177/0261927X93123004

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*(8), 1608–1613. https://doi.org/10.1037/xge0000729

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*. https://doi.org/10.1126/science.aaa1160

Banas, J. A., Bessarabova, E., Penkauskas, M. C., & Talbert, N. (2023). Inoculating Against Anti-Vaccination Conspiracies. *Health Communication*, 1–9. https://doi.org/10.1080/10410236.2023.2235733

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

Baron, R. S., Vandello, J. A., & Brunsman, B. (1996a). The forgotten variable in conformity research: Impact of task importance on social influence. *Journal of Personality and Social Psychology*, *71*(5), 915–927. https://doi.org/10.1037/0022-3514.71.5.915

Baron, R. S., Vandello, J. A., & Brunsman, B. (1996b). The forgotten variable in conformity research: Impact of task importance on social influence. *Journal of Personality and Social Psychology*, *71*(5), 915–927. https://doi.org/10.1037/0022-3514.71.5.915

Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. van der. (2021). Towards psychological herd immunity: Cross-cultural evidence for two

prebunking interventions against COVID-19 misinformation. *Big Data & Society*, *8*(1), 20539517211013868. https://doi.org/10.1177/20539517211013868

Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good News about Bad News: Gamified Inoculation Boosts Confidence and Cognitive Immunity Against Fake News. *Journal of Cognition*, *3*(1), 2. https://doi.org/10.5334/joc.91

Batailler, C., Brannon, S. M., Teas, P. E., & Gawronski, B. (2021). A Signal Detection Approach to Understanding the Identification of Fake News. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691620986135

Bauer, P. C., & Clemm von Hohenberg, B. (2021). Believing and Sharing Information by Fake Sources: An Experiment. *Political Communication*, *38*(6), 647–671. https://doi.org/10.1080/10584609.2020.1840462

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. https://doi.org/10.1093/oso/9780190923624.001.0001

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms* (pp. xvi, 260). Cambridge University Press.

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgement: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, *37*(1), 48–74. https://doi.org/10.1037/0022-3514.37.1.48

Bisgaard, M. (2019). How Getting the Facts Right Can Fuel Partisan-Motivated Reasoning. *American Journal of Political Science*, *63*(4), 824–839. https://doi.org/10.1111/ajps.12432

Blassnig, S., Büchel, F., Ernst, N., & Engesser, S. (2019). Populism and Informal Fallacies: An Analysis of Right-Wing Populist Rhetoric in Election Campaigns. *Argumentation*, *33*(1), 107–136. https://doi.org/10.1007/s10503-018-9461-2

Blunt, D. (2023). *COVID Misinformation on TikTok*.

Borges do Nascimento, I. J., Pizarro, A. B., Almeida, J. M., Azzopardi-Muscat, N., Gonçalves, M. A., Björklund, M., & Novillo-Ortiz, D. (2022). Infodemics and health misinformation: A systematic review of reviews. *Bulletin of the World Health Organization*, *100*(9), 544–561. https://doi.org/10.2471/BLT.21.287654

Braddock, K., & Dillard, J. P. (2016). Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behaviors. *Communication Monographs*. https://www.tandfonline.com/doi/abs/10.1080/03637751.2015.1128555

Briñol, P., & Petty, R. E. (2009). Source factors in persuasion: A self-validation approach. *European Review of Social Psychology*, *20*(1), 49–96. https://doi.org/10.1080/10463280802643640

Brock, T. C. (1965). Communicator-recipient similarity and decision change. *Journal of Personality and Social Psychology*, *1*(6), 650–654. https://doi.org/10.1037/h0022081

Bryanov, K., & Vziatysheva, V. (2021). Determinants of individuals' belief in fake news: A scoping review determinants of belief in fake news. *PLOS ONE*, *16*(6), e0253717. https://doi.org/10.1371/journal.pone.0253717

Budak, C., Goel, S., & Rao, J. M. (2016). Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis. *Public Opinion Quarterly*, *80*(S1), 250–271. https://doi.org/10.1093/poq/nfw007

Budescu, D. V., & Chen, E. (2015). Identifying Expertise to Extract the Wisdom of Crowds. *Management Science*, *61*(2), 267–280.

Bundgaard, H., Bundgaard, J. S., Raaschou-Pedersen, D. E. T., von Buchwald, C., Todsen, T., Norsk, J. B., Pries-Heje, M. M., Vissing, C. R., Nielsen, P. B., Winsløw, U. C., Fogh, K., Hasselbalch, R., Kristensen, J. H., Ringgaard, A., Porsborg Andersen, M., Goecke, N. B., Trebbien, R., Skovgaard, K., Benfield, T., … Iversen, K. (2020). Effectiveness

of Adding a Mask Recommendation to Other Public Health Measures to Prevent SARS-CoV-2 Infection in Danish Mask Wearers. *Annals of Internal Medicine*, M20-6817. https://doi.org/10.7326/M20-6817

Burnstein, E., & Vinokur, A. (1977). Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology*, *13*(4), 315–332. https://doi.org/10.1016/0022-1031(77)90002-6

Butler, L. H., Fay, N., & Ecker, U. K. H. (2023). Social endorsement influences the continued belief in corrected misinformation. *Journal of Applied Research in Memory and Cognition*, *12*(3), 364–375. https://doi.org/10.1037/mac0000080

Cacciatore, M. A. (2021). Misinformation and public opinion of science and health: Approaches, findings, and future directions. *Proceedings of the National Academy of Sciences*, *118*(15), e1912437117. https://doi.org/10.1073/pnas.1912437117

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Carnahan, D., Ulusoy, E., Barry, R., McGraw, J., Virtue, I., & Bergan, D. E. (2022). What should I believe? A conjoint analysis of the influence of message characteristics on belief in, perceived credibility of, and intent to share political posts. *Journal of Communication*, *72*(5), 592–603. https://doi.org/10.1093/joc/jqac023

Carrasco-Farré, C. (2022). The fingerprints of misinformation: How deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions. *Humanities and Social Sciences Communications*, *9*(1), Article 1. https://doi.org/10.1057/s41599-022-01174-9

Carrasco-Farré, C. (2024). Unmasking the Info War: The Communication Dynamics of Reliable and Misinformation Sources During the COVID-19 Pandemic—Carlos

Carrasco-Farré, 2024. *Business & Society*. https://doi.org/10.1177/00076503241255343

Chaiken, S. (1980). Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion. *Journal of Personality and Social Psychology*, *39*(5).

Chaiken, S. (1987). The heuristic model of persuasion. In *Social influence: The Ontario symposium, Vol. 5.* (pp. 3–39). Lawrence Erlbaum Associates, Inc.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In *Unintended thought* (pp. 212–252). The Guilford Press.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgement. *Journal of Personality and Social Psychology*, *66*(3), 460–473. https://doi.org/10.1037/0022-3514.66.3.460

Chaiken, S., Wood, W., & Eagly, A. H. (1996). Principles of persuasion. In *Social psychology: Handbook of basic principles* (pp. 702–742). The Guilford Press.

Chan, M. S., Jones, C. R., Jamieson, K. H., & Albarracín, D. (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychological Science*, *28*(11), 1531–1546.

Charness, G., & Dave, C. (2017). Confirmation bias with motivated beliefs. *Games and Economic Behavior*, *104*, 1–23. https://doi.org/10.1016/j.geb.2017.02.015

Cheung, C. M. K., Chiu, P.-Y., & Lee, M. K. O. (2011). Online social networks: Why do students use facebook? *Computers in Human Behavior*, *27*(4), 1337–1343. https://doi.org/10.1016/j.chb.2010.07.028

Cialdini, R. (1993). *Influence: Science and Practice*.

Cialdini, R. B. (2001). *Influence: Science and Practice*. Allyn and Bacon.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*, 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior. *Advances in Experimental Social Psychology*, *24*, 201–234. https://doi.org/10.1016/S0065-2601(08)60330-5

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In *The handbook of social psychology, Vols. 1-2, 4th ed* (pp. 151–192). McGraw-Hill.

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9), e2023301118. https://doi.org/10.1073/pnas.2023301118

Clarkson, J. J., Tormala, Z. L., Rucker, D. D., & Dugan, R. G. (2013). The malleable influence of social consensus on attitude certainty. *Journal of Experimental Social Psychology*, *49*(6), 1019–1022. https://doi.org/10.1016/j.jesp.2013.07.001

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2020). Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media. *Political Behavior*, *42*(4), 1073–1095. https://doi.org/10.1007/s11109-019-09533-0

Coe, K., Tewksbury, D., Bond, B. J., Drogos, K. L., Porter, R. W., Yahn, A., & Zhang, Y. (2008). Hostile news: Partisan use and perceptions of cable news programming.

Journal of Communication, *58*(2), 201–219. https://doi.org/10.1111/j.1460-2466.2008.00381.x

Coenen, L. (2022). The indirect effect is omitted variable bias. A cautionary note on the theoretical interpretation of products-of-coefficients in mediation analyses. *European Journal of Communication*, *37*(6), 679–688. https://doi.org/10.1177/02673231221082244

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

Colicchio, T. (2023). *Bias in Fact Checking?: An Analysis of Partisan Trends Using PolitiFact Data*. https://hdl.handle.net/10161/29011

Compton, J. A., & Pfau, M. W. (2005). Inoculation Theory of Resistance to Influence at Maturity: Recent Progress in Theory Development and Application and Suggestions for Future Research. In *Communication yearbook 29* (pp. 97–145). Lawrence Erlbaum Associates Publishers.

Compton, J., van der Linden, S., Cook, J., & Basol, M. (2021). Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories. *Social and Personality Psychology Compass*, *15*(6), e12602. https://doi.org/10.1111/spc3.12602

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

D'Alessio, D., & Allen, M. (2000). Media bias in presidential elections: A meta-analysis. *Journal of Communication*, *50*(4), 133–156. https://doi.org/10.1111/j.1460-2466.2000.tb02866.x

De keersmaecker, J., & Roets, A. (2019). Is there an ideological asymmetry in the moral approval of spreading misinformation by politicians? *Personality and Individual Differences*, *143*, 165–169. https://doi.org/10.1016/j.paid.2019.02.003

DeFleur, M. L., & Ball-Rokeach, S. (1989). *Theories of mass communication* (5th ed). Longman.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559. https://doi.org/10.1073/pnas.1517441113

Dennis, A. R., Moravec, P. L., & Kim, A. (2023). Search & Verify: Misinformation and source evaluations in Internet search results. *Decision Support Systems*, *171*, 113976. https://doi.org/10.1016/j.dss.2023.113976

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgement. *The Journal of Abnormal and Social Psychology*, *51*(3), 629–636. https://doi.org/10.1037/h0046408

Dias, N., Pennycook, G., & Rand, D. G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *Harvard Kennedy School Misinformation Review*, *1*(1). https://doi.org/10.37016/mr-2020-001

Diefenbach, S., & Anders, L. (2022). The psychology of likes: Relevance of feedback on Instagram and relationship to self-esteem and social status. *Psychology of Popular Media*, *11*(2), 196–207. https://doi.org/10.1037/ppm0000360

Diekman, C., Ryan, C. D., & Oliver, T. L. (2023). Misinformation and Disinformation in Food Science and Nutrition: Impact on Practice. *The Journal of Nutrition*, *153*(1), 3–9. https://doi.org/10.1016/j.tjnut.2022.10.001

Domenico, G. D., Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, *124*, 329–341. https://doi.org/10.1016/j.jbusres.2020.11.037

Douglas, K. M., & Sutton, R. M. (2004). Right about others, wrong about ourselves? Actual and perceived self-other differences in resistance to persuasion. *British Journal of Social Psychology*, *43*(4), 585–603. https://doi.org/10.1348/0144666042565416

Dvir-Gvirsman, S. (2019). I like what I see: Studying the influence of popularity cues on attention allocation and news selection. *Information, Communication & Society*, *22*(2). https://doi.org/10.1080/1369118X.2017.1379550

Eagly, A. H., Wood, W., & Chaiken, S. (1978). Causal inferences about communicators and their effect on opinion change. *Journal of Personality and Social Psychology*, *36*(4), 424–435. https://doi.org/10.1037/0022-3514.36.4.424

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), Article 1. https://doi.org/10.1038/s44159-021-00006-y

Ecker, U. K. H., Tay, L. Q., Roozenbeek, J., Van Der Linden, S., Cook, J., Oreskes, N., & Lewandowsky, S. (2024). Why misinformation must not be ignored. *American Psychologist*. https://doi.org/10.1037/amp0001448

Ejaz, W., Altay, S., Fletcher, R., & Nielsen, R. K. (2024). Trust is key: Determinants of false beliefs about climate change in eight countries. *New Media & Society*, 14614448241250302. https://doi.org/10.1177/14614448241250302

Ellemers, N., & Haslam, S. A. (2012). Social identity theory. In *Handbook of theories of social psychology, Vol. 2* (pp. 379–398). Sage Publications Ltd. https://doi.org/10.4135/9781446249222.n45

Ellemers, N., Spears, R., & Doosje, B. (1997). Sticking together or falling apart: In-group identification as a psychological determinant of group commitment versus individual mobility. *Journal of Personality and Social Psychology*, *72*(3), 617–626. https://doi.org/10.1037/0022-3514.72.3.617

Ellul, J. (1973). *Propaganda: The Formation of Men's Attitudes* (K. Kellen & J. Lerner, Trans.; First Edition). Vintage.

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, *12*(5), 391–396. https://doi.org/10.1111/1467-9280.00372

Facciani, M., & Steenbuch-Traberg, C. (2024). Personal Network Composition and Cognitive Reflection Predict Susceptibility to Different Types of Misinformation. *Connections*, *0*(0). https://doi.org/10.21307/connections-2019.044

Faraji-Rad, A., Samuelsen, B. M., & Warlop, L. (2015). On the persuasiveness of similar others: The role of mentalizing and the feeling of certainty. *Journal of Consumer Research*, *42*(3), 458–471. https://doi.org/10.1093/jcr/ucv032

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993–1002. https://doi.org/10.1037/xge0000098

Feng, L., Hu, Y., Li, B., Stanley, H. E., Havlin, S., & Braunstein, L. A. (2015). Competing for Attention in Social Media under Information Overload Conditions. *PLOS ONE*, *10*(7), e0126090. https://doi.org/10.1371/journal.pone.0126090

Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based

information. *New Media & Society*, *9*(2), 319–342. https://doi.org/10.1177/1461444807075015

Fong, J., Guo, T., & Rao, A. (2021). *Debunking Misinformation About Consumer Products: Effects on Beliefs and Purchase Behavior*. https://dx.doi.org/10.2139/ssrn.3875665

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Galesic, M., Olsson, H., Dalege, J., van der Does, T., & Stein, D. L. (2021). Integrating social and cognitive aspects of belief dynamics: Towards a unifying framework. *Journal of The Royal Society Interface*, *18*(176), 20200857. https://doi.org/10.1098/rsif.2020.0857

Gardikiotis, A., Martin, R., & Hewstone, M. (2005). Group Consensus in Social Influence: Type of Consensus Information as a Moderator of Majority and Minority Influence. *Personality & Social Psychology Bulletin*, *31*(9). https://journals.sagepub.com/doi/10.1177/0146167205277807

Garrett, R. K., & Bond, R. M. (2021). Conservatives' susceptibility to political misperceptions. *Science Advances*, *7*(23), eabf1234. https://doi.org/10.1126/sciadv.abf1234

Gavin, L., McChesney, J., Tong, A., Sherlock, J., Foster, L., & Tomsa, S. (2022). Fighting the Spread of COVID-19 Misinformation in Kyrgyzstan, India, and the United States: How Replicable Are Accuracy Nudge Interventions? *Technology, Mind, and Behavior*, *3*(3: Autumn). https://doi.org/10.1037/tmb0000086

Gentzkow, M., & Shapiro, J. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, *78*(1), 35–71. https://doi.org/10.3982/ECTA7195

Gentzkow, M., Wong, M. B., & Zhang, A. T. (n.d.). *Ideological Bias and Trust in Information Sources*.

Gerard, H. B., Wilhelmy, R. A., & Conolley, E. S. (1968). Conformity and group size. *Journal of Personality and Social Psychology*, *8*(1, Pt.1), 79–82. https://doi.org/10.1037/h0025325

Gino, F., Shang, J., & Croson, R. (2009). The impact of information from similar or different advisors on judgement. *Organizational Behavior and Human Decision Processes*, *108*(2), 287–302. https://doi.org/10.1016/j.obhdp.2008.08.002

Gottfried, M. W. and J. (2019, June 27). Republicans far more likely than Democrats to say fact-checkers tend to favor one side. *Pew Research Center*. https://www.pewresearch.org/short-reads/2019/06/27/republicans-far-more-likely-than-democrats-to-say-fact-checkers-tend-to-favor-one-side/

Graves, L. (2018). Boundaries Not Drawn: Mapping the institutional roots of the global fact-checking movement. *Journalism Studies*, *19*(5), 613–631. https://doi.org/10.1080/1461670X.2016.1196602

Green, J., Hobbs, W., McCabe, S., & Lazer, D. (2022). Online engagement with 2020 election misinformation and turnout in the 2021 Georgia runoff election. *Proceedings of the National Academy of Sciences*, *119*(34), e2115900119. https://doi.org/10.1073/pnas.2115900119

Greenaway, K. H., Wright, R. G., Willingham, J., Reynolds, K. J., & Haslam, S. A. (2015). Shared identity is key to effective communication. *Personality and Social Psychology Bulletin*, *41*(2), 171–182. https://doi.org/10.1177/0146167214559709

Grieco, J. G., Galen Stocking and Elizabeth, Stocking, G., & Grieco, E. (2018). *Partisans Remain Sharply Divided in Their Attitudes About the News Media*. Pew Research Center. https://www.pewresearch.org/journalism/2018/09/25/partisans-remain-sharply-divided-in-their-attitudes-about-the-news-media/

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, *363*(6425), 374–378. https://doi.org/10.1126/science.aau2706

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), eaau4586. https://doi.org/10.1126/sciadv.aau4586

Hamlin, C. L. (1970). *Fallacies*. Methuen. https://www.abebooks.com/Fallacies-Hamblin-C.L-London-Methuen-Ltd/31169487055/bd

Harper, C. A., & Baguley, T. (2019). *"You are Fake News": Ideological (A)symmetries in Perceptions of Media Legitimacy*. OSF. https://doi.org/10.31234/osf.io/ym6t5

Hertz, U., Palminteri, S., Brunetti, S., Olesen, C., Frith, C. D., & Bahrami, B. (2017). Neural computations underpinning the strategic management of influence in advice giving. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/s41467-017-02314-5

Hewson, J. T. S., & Fang, K. (2024). *Evaluating internal and external dissonance of belief dynamics in social systems* (arXiv:2410.07240). arXiv. https://doi.org/10.48550/arXiv.2410.07240

Hewstone, M., & Jaspars, J. (1988). Implicit and explicit consensus as determinants of causal attribution: Two experimental investigations. *European Journal of Social Psychology*, *18*(1), 93–98. https://doi.org/10.1002/ejsp.2420180108

Hilverda, F., Kuttschreuter, M., & Giebels, E. (2018). The Effect of Online Social Proof Regarding Organic Food: Comments and Likes on Facebook. *Frontiers in Communication*, *3*, 30. https://doi.org/10.3389/fcomm.2018.00030

Hogg, M. A., & Terry, D. J. (2000). Social Identity and Self-Categorization Processes in Organizational Contexts. *The Academy of Management Review*, *25*(1), 121–140. https://doi.org/10.2307/259266

Holton, A. E., & Chyi, H. I. (2012). News and the overloaded consumer: Factors influencing information overload among news consumers. *Cyberpsychology, Behavior and Social Networking*, *15*(11), 619–624. https://doi.org/10.1089/cyber.2011.0610

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion; psychological studies of opinion change* (pp. xii, 315). Yale University Press.

Hovland, C. I., & Weiss, W. (1951). The Influence of Source Credibility on Communication Effectiveness. *Public Opinion Quarterly*, *15*(4), 635–650. https://doi.org/10.1086/266350

Hruschka, T. M. J., & Appel, M. (2023). Learning about informal fallacies and the detection of fake news: An experimental intervention. *PLOS ONE*, *18*(3), e0283238. https://doi.org/10.1371/journal.pone.0283238

Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, *59*(1), 19–39. https://doi.org/10.1111/j.1460-2466.2008.01402.x

Iyengar, S., & Massey, D. S. (2019). Scientific communication in a post-truth society. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(16), 7656–7661. https://doi.org/10.1073/pnas.1805868115

Iyengar, S., & Westwood, S. J. (2015). Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, *59*(3), 690–707. https://doi.org/10.1111/ajps.12152

Jack, C. (2017). *Lexicon of Lies: Terms for Problematic Information*. Data and Society Research Institute.

Jaeger, P. T., & Burnett, G. (2010). *Information Worlds: Behavior, Technology, and Social Context in the Age of the Internet* (1st ed.). Taylor & Francis Group. http://ebookcentral.proquest.com/lib/cam/detail.action?docID=496296

Jiang, Y., Zhao, X., Zhu, L., Liu, J. S., & Deng, K. (2021). Total-Effect Test Is Superfluous for Establishing Complementary Mediation. *Statistica Sinica*, *31*(4), 1961–1983.

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1420–1436. https://doi.org/10.1037/0278-7393.20.6.1420

Johnson, T. J., & Kaye, B. K. (1998). Cruising is Believing?: Comparing Internet and Traditional Sources on Media Credibility Measures. *Journalism & Mass Communication Quarterly*, *75*(2), 325–340. https://doi.org/10.1177/107769909807500208

Jolley, D., & Paterson, J. L. (2020). Pylons ablaze: Examining the role of 5G COVID-19 conspiracy beliefs and support for violence. *The British Journal of Social Psychology*, *59*(3), 628–640. https://doi.org/10.1111/bjso.12394

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, *129*(3), 339–375. https://doi.org/10.1037/0033-2909.129.3.339

Jost, J. T., van der Linden, S., Panagopoulos, C., & Hardin, C. D. (2018). Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation. *Current Opinion in Psychology*, *23*, 77–83. https://doi.org/10.1016/j.copsyc.2018.01.003

Jowett, G. S., & O'Donnell, V. (2018). *Propaganda and Persuasion* (7th ed.). SAGE Publications, Inc.

Juez, L. A., & Mackenzie, J. L. (2019). Emotion, lies, and "bullshit" in journalistic discourse: *Ibérica*, *38*, Article 38.

Jurkowitz, M., Mitchell, A., Shearer, E., & Walker, M. (2020). *U.S. Media Polarization and the 2020 Election: A Nation Divided*. Pew Research Center. https://www.pewresearch.org/journalism/2020/01/24/u-s-media-polarization-and-the-2020-election-a-nation-divided/

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgement and Decision Making*, *8*(4), 407–424. https://doi.org/10.1017/S1930297500005271

Kahan, D. M., Landrum, A., Carpenter, K., Helft, L., & Hall Jamieson, K. (2017). Science Curiosity and Political Information Processing. *Political Psychology*, *38*(S1), 179–199. https://doi.org/10.1111/pops.12396

Kang, H., Bae, K., Zhang, S., & Sundar, S. S. (2011). Source cues in online news: Is the proximate source more powerful than distal sources? *Journalism and Mass Communication Quarterly*, *88*(4), 719–736. https://doi.org/10.1177/107769901108800403

Kassin, S. M. (1979). Consensus information, prediction, and causal attribution: A review of the literature and issues. *Journal of Personality and Social Psychology*, *37*(11), 1966–1981. https://doi.org/10.1037/0022-3514.37.11.1966

Kim, A., & Dennis, A. R. (2018). *Says Who? The Effects of Presentation Format and Source Rating on Fake News in Social Media*. https://doi.org/10.2139/ssrn.2987866

Kim, J. W. (2018). They liked and shared: Effects of social media virality metrics on perceptions of message influence and behavioral intentions. *Computers in Human Behavior*, *84*, 153–161. https://doi.org/10.1016/j.chb.2018.01.030

Kim, M. (2016). The Role of Partisan Sources and Audiences' Involvement in Bias Perceptions of Controversial News. *Media Psychology*, *19*(2), 203–223. https://doi.org/10.1080/15213269.2014.1002941

Kluck, J. P., Schaewitz, L., & Kraemer, N. C. (2019). Doubters are more convincing than advocates The impact of user comments and ratings on credibility perceptions of false news stories on social media. *Scm Studies in Communication and Media*, *8*(4), 446–470. https://doi.org/10.5771/2192-4007-2019-4-446

Kobayashi, K. (2018). The Impact of Perceived Scientific and Social Consensus on Scientific Beliefs. *Science Communication*, *40*(1). https://doi.org/10.1080/15534510.2019.1650105

Koessmeier, C., & Büttner, O. B. (2021). Why Are We Distracted by Social Media? Distraction Situations and Strategies, Reasons for Distraction, and Individual Differences. *Frontiers in Psychology*, *12*, 711416. https://doi.org/10.3389/fpsyg.2021.711416

Kumkale, G. T., Albarracín, D., & Seignourel, P. J. (2010). The effects of source credibility in the presence or absence of prior attitudes: Implications for the design of persuasive communication campaigns. *Journal of Applied Social Psychology*, *40*(6), 1325–1356. https://doi.org/10.1111/j.1559-1816.2010.00620.x

Lakens, D., McLatchie, N., Isager, P., Scheel, A., & Dienes, Z. (2020). Improving Inferences About Null Effects With Bayes Factors and Equivalence Tests. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *75*(1). https://doi.org/10.1093/geronb/gby065

Larcinese, V., Puglisi, R., & Snyder Jr., J. M. (2011). Partisan bias in economic news: Evidence on the agenda-setting behavior of U.S. newspapers. *Journal of Public Economics*, *95*, 1178–1189. https://doi.org/doi:10.1016/j.jpubeco.2011.04.006

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science

of fake news. *Science (New York, N.Y.)*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lee, S. S., Liang, F., Hahn, L., Lane, D. S., Weeks, B. E., & Kwak, N. (2021). The Impact of Social Endorsement Cues and Manipulability Concerns on Perceptions of News Credibility. *Cyberpsychology, Behavior and Social Networking*, *24*(6), 384–389. https://doi.org/10.1089/cyber.2020.0566

Lewandowsky, S., Cook, J., Fay, N., & Gignac, G. E. (2019). Science by social media: Attitudes towards climate change are mediated by perceived social consensus. *Memory & Cognition*, *47*(8), 1445–1456. https://doi.org/10.3758/s13421-019-00948-y

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest, Supplement*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lewandowsky, S., & van der Linden, S. (2021). Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology*, *32*(2), 348–384. https://doi.org/10.1080/10463283.2021.1876983

Lewandowsky, S., & Yesilada, M. (2021). Inoculating against the spread of Islamophobic and radical-Islamist disinformation. *Cognitive Research: Principles and Implications*, *6*(1), 57. https://doi.org/10.1186/s41235-021-00323-z

Lins de Holanda Coelho, G., H. P. Hanel, P., & J. Wolf, L. (2020). The Very Efficient Assessment of Need for Cognition: Developing a Six-Item Version. *Assessment*, *27*(8), 1870–1885. https://doi.org/10.1177/1073191118793208

Liu, Y., & Hu, K. (2024). How social media exposure constructs social confidence: An empirical study on impact, mechanisms, and multilateral relationships. *PLOS ONE*, *19*(9), e0308745. https://doi.org/10.1371/journal.pone.0308745

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, *5*(3), 337–348. https://doi.org/10.1038/s41562-021-01056-1

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, *108*(22), 9020–9025. https://doi.org/10.1073/pnas.1008636108

Lu, A. S. (2013). An Experimental Test of the Persuasive Effect of Source Similarity in Narrative and Nonnarrative Health Blogs. *Journal of Medical Internet Research*, *15*(7), e142. https://doi.org/10.2196/jmir.2386

Lu, C., Hu, B., Li, Q., Bi, C., & Ju, X.-D. (2023). Psychological Inoculation for Credibility Assessment, Sharing Intention, and Discernment of Misinformation: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, *25*, e49255. https://doi.org/10.2196/49255

Lu, J., & Xiao, Y. (2024). Heuristic Information Processing as a Mediating Factor in the Process of Exposure to COVID-19 Vaccine Information and Misinformation Sharing on Social Media. *Health Communication*, *39*(12), 2779–2792. https://doi.org/10.1080/10410236.2023.2288373

Ludwig, J., & Sommer, J. (2024). Mindsets and politically motivated reasoning about fake news. *Motivation and Emotion*, *48*(3), 249–263. https://doi.org/10.1007/s11031-024-10067-0

Luo, M., Hancock, J. T., & Markowitz, D. M. (2020). Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues. *Communication Research*. https://doi.org/10.1177/0093650220921321

Mackie, D. M., Gastardo-Conaco, M. C., & Skelly, J. J. (1992). Knowledge of the advocated position and the processing of in-group and out-group persuasive messages. *Personality and Social Psychology Bulletin*, *18*(2), 145–151. https://doi.org/10.1177/0146167292182005

Mackie, D. M., Worth, L. T., & Asuncion, A. G. (1990). Processing of persuasive in-group messages. *Journal of Personality and Social Psychology*, *58*(5), 812–822. https://doi.org/10.1037/0022-3514.58.5.812

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

Maio, G., & Haddock, G. (2010). *The Psychology of Attitudes and Attitude Change*. https://doi.org/10.4135/9781446214299

Malone, T. W., & Bernstein, M. S. (2022). *Handbook of Collective Intelligence*. MIT Press.

Mang, V., Fennis, B. M., & Epstude, K. (2024). Source credibility effects in misinformation research: A review and primer. *Advances.in/Psychology*, *aip00028*. https://doi.org/10.56296/aip00028

Marks, J., Copland, E., Loh, E., Sunstein, C. R., & Sharot, T. (2019). Epistemic spillovers: Learning others' political views reduces the ability to assess and use their expertise in nonpolitical domains. *Cognition*, *188*, 74–84. https://doi.org/10.1016/j.cognition.2018.10.003

Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications*, *5*(1), 47. https://doi.org/10.1186/s41235-020-00252-3

Mason, A. M., Compton, J., Tice, E., Peterson, B., Lewis, I., Glenn, T., & Combs, T. (2023). Analyzing the Prophylactic and Therapeutic Role of Inoculation to Facilitate Resistance to Conspiracy Theory Beliefs. *Communication Reports*, *0*(0), 1–15. https://doi.org/10.1080/08934215.2023.2256803

Matsa, M. W. and K. E. (2021, September 20). News Consumption Across Social Media in 2021. *Pew Research Center's Journalism Project*. https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/

Matthes, J., Karsay, K., Schmuck, D., & Stevic, A. (2020). "Too much to handle": Impact of mobile social networking sites on information overload, depressive symptoms, and well-being. *Computers in Human Behavior*, *105*, 106217. https://doi.org/10.1016/j.chb.2019.106217

McGarty, C., Haslam, S. A., Hutchinson, K. J., & Turner, J. C. (1994). The effects of salient group memberships on persuasion. *Small Group Research*, *25*(2), 267–293. https://doi.org/10.1177/1046496494252007

McGuire, W. J. (1961). Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *The Journal of Abnormal and Social Psychology*, *63*(2), 326–332. https://doi.org/10.1037/h0048344

McGuire, W. J. (1964). Inducing resistance to persuasion: Some contemporary approaches. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 1, pp. 191–229). Academic Press. 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

McGuire, W. J., & Papageorgis, D. (1962). Effectiveness of Forewarning in Developing Resistance to Persuasion. *Public Opinion Quarterly*, *26*(1), 24–34. https://doi.org/10.1086/267068

McLoughlin, K. L., Brady, W. J., Goolsbee, A., Kaiser, B., Klonick, K., & Crockett, M. J. (2024). *Misinformation exploits outrage to spread online*.

Mena, P., Barbe, D., & Chan-Olmsted, S. (2020a). Misinformation on Instagram: The Impact of Trusted Endorsements on Message Credibility. *Social Media & Society*, *6*(2), 2056305120935102. https://doi.org/10.1177/2056305120935102

Mena, P., Barbe, D., & Chan-Olmsted, S. (2020b). Misinformation on Instagram: The Impact of Trusted Endorsements on Message Credibility. *Social Media & Society*, *6*(2), 2056305120935102. https://doi.org/10.1177/2056305120935102

Messing, S., & Westwood, S. J. (2014). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. *Communication Research*, *41*(8), 1042–1063. https://doi.org/10.1177/0093650212466406

Metzger, M., Flanagin, A., Eyal, K., Lemus, D., & Mccann, R. (2003). Credibility for the 21st Century: Integrating Perspectives on Source, Message, and Media Credibility in the Contemporary Media Environment. *Annals of the International Communication Association*, *27*, 293–335. https://doi.org/10.1080/23808985.2003.11679029

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and Heuristic Approaches to Credibility Evaluation Online. *Journal of Communication*, *60*(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Morstein-Marx, R. (2004). *Mass Oratory and Political Power in the Late Roman Republic*. Cambridge University Press. https://doi.org/10.1017/CBO9780511482878

Mosleh, M., Pennycook, G., & Rand, D. G. (2020). Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. *PLOS ONE*, *15*(2), e0228882. https://doi.org/10.1371/journal.pone.0228882

Motta, M., Stecula, D., & Farhart, C. (2020). How Right-Leaning Media Coverage of COVID-19 Facilitated the Spread of Misinformation in the Early Stages of the Pandemic in the U.S. *Canadian Journal of Political Science*, *53*(2), 335–342. https://doi.org/10.1017/S0008423920000396

Moussaïd, M., Kämmer, J. E., Analytis, P. P., & Neth, H. (2013). Social Influence and the Collective Dynamics of Opinion Formation. *PLOS ONE*, *8*(11), e78433. https://doi.org/10.1371/journal.pone.0078433

Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social Influence Bias: A Randomized Experiment. *Science*. https://doi.org/10.1126/science.1240466

Mulcahy, R., Barnes, R., de Villiers Scheepers, R., Kay, S., & List, E. (2024). Going Viral: Sharing of Misinformation by Social Media Influencers. *Australasian Marketing Journal*, 14413582241273987. https://doi.org/10.1177/14413582241273987

Muscanell, N. L., Guadagno, R. E., & Murphy, S. (2014). Weapons of influence misused: A social influence analysis of why people fall prey to internet scams. *Social and Personality Psychology Compass*, *8*(7), 388–396. https://doi.org/10.1111/spc3.12115

Nadarevic, L., Reber, R., Helmecke, A. J., & Köse, D. (2020). Perceived truth of statements and simulated social media postings: An experimental investigation of source credibility, repeated exposure, and presentation format. *Cognitive Research: Principles and Implications*, *5*(1), 56. https://doi.org/10.1186/s41235-020-00251-4

Naeem & Ozuem. (2022). Understanding misinformation and rumors that generated panic buying as a social practice during COVID-19 pandemic: Evidence from Twitter, YouTube and focus group interviews. *Information Technology & People*, *35*(7), 2140–2166. https://doi.org/10.1108/ITP-01-2021-0061

Nan, X., Wang, Y., & Thier, K. (2022). Why do people believe health misinformation and who is at risk? A systematic review of individual differences in susceptibility to health misinformation. *Social Science & Medicine*, *314*, 115398. https://doi.org/10.1016/j.socscimed.2022.115398

Nelson, T., Clawson, R., & Oxley, Z. (1997). Media Framing of A Civil Liberties Conflict and Its Effect on Tolerance. *The American Political Science Review*, *91*, 567–583. https://doi.org/10.2307/2952075

New York Times. (2020). *'PizzaGate' Conspiracy Theory Thrives Anew in the TikTok Era*. https://www.nytimes.com/2020/06/27/technology/pizzagate-justin-bieber-qanon-tiktok.html

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative Social Influence is Underdetected. *Personality and Social Psychology Bulletin*, *34*(7), 913–923. https://doi.org/10.1177/0146167208316691

O'Keefe, D. J. (2016). *Persuasion: Theory and Research*. SAGE Publications, Inc.

Ooms, J., Hoeks, J., & Jansen, C. (2019). "Hey, that could be me": The role of similarity in narrative persuasion. *PLOS ONE*, *14*(4), e0215359. https://doi.org/10.1371/journal.pone.0215359

Paschen, J. (2019). Investigating the emotional appeal of fake news using artificial intelligence and human contributions. *Journal of Product & Brand Management*, *29*(2), 223–233. https://doi.org/10.1108/JPBM-12-2018-2179

Pehlivanoglu, D., Lin, T., Deceus, F., Heemskerk, A., Ebner, N. C., & Cahill, B. S. (2021). The role of analytical reasoning and source credibility on the evaluation of real and fake full-length news articles. *Cognitive Research: Principles and Implications*, *6*(1), 24. https://doi.org/10.1186/s41235-021-00292-3

Peng, W., Lim, S., & Meng, J. (2023). Persuasive strategies in online health misinformation: A systematic review. *Information, Communication & Society*, *26*(11), 2131–2148. https://doi.org/10.1080/1369118X.2022.2085615

Pennycook, G., Berinsky, A. J., Bhargava, P., Lin, H., Cole, R., Goldberg, B., Lewandowsky, S., & Rand, D. G. (2024). Inoculation and accuracy prompting increase accuracy discernment in combination but not alone. *Nature Human Behaviour*, *8*(12), 2330–2341. https://doi.org/10.1038/s41562-024-02023-2

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology. General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), Article 7855. https://doi.org/10.1038/s41586-021-03344-2

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. (2019a). Fighting misinformation on social media using crowdsourced judgements of news source quality. *Proceedings of the National Academy of Sciences*, *116*(7), 2521–2526. https://doi.org/10.1073/pnas.1806781116

Pennycook, G., & Rand, D. G. (2019b). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, *88*(2), 185–200. https://doi.org/10.1111/jopy.12476

Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Pentina, I., Bailey, A. A., & Zhang, L. (2018). Exploring effects of source similarity, message valence, and receiver regulatory focus on yelp review persuasiveness and purchase intentions. *Journal of Marketing Communications*, *24*(2), 125–145. https://doi.org/10.1080/13527266.2015.1005115

Pereira, A., Harris, E., & Van Bavel, J. J. (2021). Identity concerns drive belief: The impact of partisan identity on the belief and dissemination of true and false news. *Group Processes & Intergroup Relations*, *26*(1), 24–47. https://doi.org/10.1177/13684302211030004

Petty, R. E., & Cacioppo, J. T. (1984). Source Factors and the Elaboration Likelihood Model of Persuasion. *Advances in Consumer Research*, *11*(1), 668–672.

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 123–205). Academic Press. https://doi.org/10.1016/S0065-2601(08)60214-2

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In *The handbook of social psychology, Vols. 1-2, 4th ed* (pp. 323–390). McGraw-Hill.

Phua, J. (2016). The effects of similarity, parasocial identification, and source credibility in

obesity public service announcements on diet and exercise self-efficacy. *Journal of Health Psychology,* 21(5), 699-708. https://doi.org/10.1177/1359105314536452

Pornpitakpan, C. (2004). The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Pretus, C., Servin-Barthet, C., Harris, E. A., Brady, W. J., Vilarroya, O., & Van Bavel, J. J. (2023). The role of political devotion in sharing partisan misinformation and resistance to fact-checking. *Journal of Experimental Psychology. General*, *152*(11), 3116–3134. https://doi.org/10.1037/xge0001436

Price, V., Nir, L., & Cappella, J. N. (2006). Normative and Informational Influences in Online Political Discussions. *Communication Theory*, *16*(1), 47–74. https://doi.org/10.1111/j.1468-2885.2006.00005.x

Prislin, R., & Wood, W. (2005). Social Influence in Attitudes and Attitude Change. In *The handbook of attitudes* (pp. 671–705). Lawrence Erlbaum Associates Publishers.

Pronin, E. (2007). Perception and misperception of bias in human judgement. *Trends in Cognitive Sciences*, *11*(1), 37–43. https://doi.org/10.1016/j.tics.2006.11.001

Rathje, S., Robertson, C., Brady, W. J., & Van Bavel, J. J. (2024). People Think That Social Media Platforms Do (but Should Not) Amplify Divisive Content. *Perspectives on Psychological Science*, *19*(5), 781–795. https://doi.org/10.1177/17456916231190392

Rathje, S., Roozenbeek, J., Traberg, C., Van Bavel, J., & van der Linden, S. (2022). Letter to the Editors of Psychological Science: Meta-Analysis Reveals that Accuracy Nudges Have Little to No Effect for U.S. Conservatives: Regarding Pennycook et al. (2020). *Psychological Science*. https://doi.org/10.25384/SAGE.12594110.v2

Rathje, S., Roozenbeek, J., Van Bavel, J. J., & van der Linden, S. (2023). Accuracy and social motivations shape judgements of (mis)information. *Nature Human Behaviour*, *7*(6), 892–903. https://doi.org/10.1038/s41562-023-01540-w

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, *118*(26), e2024292118. https://doi.org/10.1073/pnas.2024292118

Reuters. (2020). *Fact check: Video shows ballot box stuffing in Russia, not Michigan*. https://www.reuters.com/article/uk-factcheck-election-fraud-russia-michi-idUSKBN27Q21W

Rhodes, S. C. (2022). Filter Bubbles, Echo Chambers, and Fake News: How Social Media Conditions Individuals to Be Less Critical of Political Misinformation. *Political Communication*, *39*(1), 1–22. https://doi.org/10.1080/10584609.2021.1910887

Rodriguez, M. G., Gummadi, K., & Schoelkopf, B. (2014). Quantifying Information Overload in Social Media and Its Impact on Social Contagions. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), Article 1. https://doi.org/10.1609/icwsm.v8i1.14549

Roozenbeek, J., Freeman, A. L. J., & Linden, S. van der. (2021). How Accurate Are Accuracy-Nudge Interventions? A Preregistered Direct Replication of Pennycook et al. (2020): *Psychological Science*. https://doi.org/10.1177/09567976211024535

Roozenbeek, J., Linden, S. V. D., & Nygren, T. (2020a). Prebunking interventions based on 'inoculation' theory can reduce susceptibility to misinformation across cultures. *The Harvard Kennedy School of Misinformation Review*. www.getbadnews.com

Roozenbeek, J., & Linden, S. van der. (2018). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research*. https://www.tandfonline.com/doi/abs/10.1080/13669877.2018.1443491

Roozenbeek, J., & Linden, S. van der. (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., Linden, S. van der, & Nygren, T. (2020b). Prebunking interventions based on "inoculation" theory can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review*, *1*(2). https://doi.org/10.37016//mr-2020-008

Roozenbeek, J., Maertens, R., Herzog, S. M., Geers, M., Kurvers, R., Sultan, M., & Linden, S. van der. (2022). Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. *Judgement and Decision Making*, *17*(3), 547–573. https://doi.org/10.1017/S1930297500003570

Roozenbeek, J., Maertens, R., McClanahan, W., & van der Linden, S. (2021). Disentangling Item and Testing Effects in Inoculation Research on Online Misinformation: Solomon Revisited. *Educational and Psychological Measurement*, *81*(2), 340–362. https://doi.org/10.1177/0013164420940378

Roozenbeek, J., Schneider, C., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, *7*(10). https://royalsocietypublishing.org/doi/10.1098/rsos.201199

Roozenbeek, J., Traberg, C. S., & van der Linden, S. (2022). Technique-based inoculation against real-world misinformation. *Royal Society Open Science*, *9*(5), 211719. https://doi.org/10.1098/rsos.211719

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

Roozenbeek, J., & van der Linden, S. (2024). *The Psychology of Misinformation*. Cambridge University Press. https://doi.org/10.1017/9781009214414

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

Rosenbaum, L. (2020, November 18). Lead Researcher Behind Controversial Danish Study Says You Should Still Wear A Mask. *Forbes*. https://www.forbes.com/sites/leahrosenbaum/2020/11/18/lead-researcher-behind-controversial-danish-study-says-you-should-still-wear-a-mask/

Rysgaard Møller, K. (2020, November 12). DR beklager fejl i artikel om mundbind. *DR*. https://www.dr.dk/etik-og-rettelser/fejl-og-fakta/dr-beklager-fejl-i-artikel-om-mundbind

Said, N., Fischer, H., & Anders, G. (2022). Contested science: Individuals with higher metacognitive insight into interpretation of evidence are less likely to polarize. *Psychonomic Bulletin & Review*, *29*(2), 668–680. https://doi.org/10.3758/s13423-021-01993-y

Schaffner, B. F., & Roche, C. (2017). Misinformation and motivated reasoning: Responses to economic news in a politicized environment. *Public Opinion Quarterly*, *81*(1), 86–110.

Schwalbe, M. C., Joseff, K., Woolley, S., & Cohen, G. L. (2024). When politics trumps truth: Political concordance versus veracity as a determinant of believing, sharing, and

recalling the news. *Journal of Experimental Psychology: General*, *153*(10), 2524–2551. https://doi.org/10.1037/xge0001650

Sechrist, G. B., & Stangor, C. (2001). Perceived consensus influences intergroup behavior and stereotype accessibility. *Journal of Personality and Social Psychology*, *80*(4), 645–654. https://doi.org/10.1037/0022-3514.80.4.645

Sechrist, G. B., & Young, A. F. (2011). The Influence of Social Consensus Information on Intergroup Attitudes: The Moderating Effects of Ingroup Identification. *The Journal of Social Psychology*, *151*(6), 674–695. https://doi.org/10.1080/00224545.2010.522615

Sehulster, J. R. (2006). Things We Talk about, How Frequently, and to Whom: Frequency of Topics in Everyday Conversation as a Function of Gender, Age, and Marital Status. *The American Journal of Psychology*, *119*(3), 407–432. https://doi.org/10.2307/20445351

Sha, H., Dolan, R., & Rashidirad, M. (2020). The Role of Social Media Content Format and Platform in Users' Engagement Behavior. *Journal of Interactive Marketing*, *53*. https://doi.org/10.1016/j.intmar.2020.05.001

Shao, C., Ciampaglia, G., Flammini, A., & Menczer, F. (2016). Hoaxy: A Platform for Tracking Online Misinformation. *WWW '16 Companion: Proceedings of the 25th International Conference Companion on World Wide Web*. https://doi.org/10.1145/2872518.2890098

Sharockman, A. (2014). *PolitiFact—Fact-checking Fox, MSNBC and CNN: PunditFact's network scorecards*. Politifact. https://www.politifact.com/article/2014/sep/16/fact-checking-fox-msnbc-and-cnn-punditfacts-networ/

Sherif, M., & Hovland, C. I. (1961). *Social judgement: Assimilation and contrast effects in communication and attitude change* (pp. xii, 218). Yale Univer. Press.

Sherrick, B. (2016). The Effects of Media Effects: Third-Person Effects, the Influence of Presumed Media Influence, and Evaluations of Media Companies. *Journalism & Mass Communication Quarterly*, *93*(4), 906–922. https://doi.org/10.1177/1077699016637108

Shin, D. (2024). Misinformation, Extremism, and Conspiracies: Amplification and Polarization by Algorithms. In D. Shin (Ed.), *Artificial Misinformation: Exploring Human-Algorithm Interaction Online* (pp. 49–78). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-52569-8_3

Simons, H. W., Berkowitz, N. N., & Moyer, R. J. (1970). Similarity, credibility, and attitude change: A review and a theory. *Psychological Bulletin*, *73*(1), 1–16. https://doi.org/10.1037/h0028429

Sindermann, C., Cooper, A., & Montag, C. (2020). A short review on susceptibility to falling for fake political news. *Current Opinion in Psychology*, *36*, 44–48. https://doi.org/10.1016/j.copsyc.2020.03.014

Snopes. (2022). *Is Disney World Lowering the Drinking Age to 18?*

Spampatti, T., Hahnel, U. J. J., Trutnevyte, E., & Brosch, T. (2023). A Trust Inoculation to Protect Public Support of Governmentally Mandated Actions to Mitigate Climate Change. *Journal of Experimental Social Psychology*, 115. https://doi.org/10.1016/j.jesp.2024.104656

Spears, R. (2021a). Social Influence and Group Identity. *Annual Review of Psychology*, *72*(1), 367–390. https://doi.org/10.1146/annurev-psych-070620-111818

Spears, R. (2021b). Social Influence and Group Identity. *Annual Review of Psychology*, *72*(1), 367–390. https://doi.org/10.1146/annurev-psych-070620-111818

Stangor, C., Sechrist, G. B., & Jost, J. T. (2001). Changing racial beliefs by providing consensus information. *Personality and Social Psychology Bulletin*, *27*(4), 486–496. https://doi.org/10.1177/0146167201274009

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–665. https://doi.org/10.1017/S0140525X00003435

Sternthal, B., Dholakia, R., & Leavitt, C. (1978). The Persuasive Effect of Source Credibility: Tests of Cognitive Response. *Journal of Consumer Research*, *4*(4), 252–260. https://doi.org/10.1086/208704

Sterrett, D., Malato, D., Benz, J., Kantor, L., Tompson, T., Rosenstiel, T., Sonderman, J., & Loker, K. (2019). Who Shared It?: Deciding What News to Trust on Social Media. *Digital Journalism*, *7*(6), 783–801. https://doi.org/10.1080/21670811.2019.1623702

Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior*, *30*(3), 341–366. https://doi.org/10.1007/s11109-007-9050-9

Sullivan, G. M., & Feinn, R. (2012). Using Effect Size—Or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Sundar, S. S., Snyder, E. C., Liao, M., Yin, J., Wang, J., & Chi, G. (2024). Sharing without clicking on news in social media. *Nature Human Behaviour*, 1–13. https://doi.org/10.1038/s41562-024-02067-4

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations* (pp. xxi, 296). Doubleday & Co.

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society Open Science*, *4*(3), 160802. https://doi.org/10.1098/rsos.160802

Swire-Thompson, B., Ecker, U. K. H., Lewandowsky, S., & Berinsky, A. J. (2020). They might be a liar but they're my liar: Source evaluation and the prevalence of misinformation. *Political Psychology*, *41*(1), 21–34. https://doi.org/10.1111/pops.12586

Tajfel, H. (1978). *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations*. European Association of Experimental Social Psychology by Academic Press.

Tajfel, H., & Turner, J. C. (1979). An Integrative Theory of Intergroup Conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33–47). Brooks/Cole.

Talhelm, T., Haidt, J., Oishi, S., Zhang, X., Miao, F. F., & Chen, S. (2015). Liberals think more analytically (more 'WEIRD') than conservatives. *Personality & Social Psychology Bulletin*, *41*(2), 250–267. https://doi.org/10.1177/0146167214563672

Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2017). Defining 'Fake News': A typology of scholarly definitions. *Digital Journalism*. https://doi.org/10.1080/21670811.2017.1360143

Tanis, M., & Postmes, T. (2005). A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour. *European Journal of Social Psychology*, *35*(3), 413–424. https://doi.org/10.1002/ejsp.256

Topham, J., & Smith, N. (2023). One day of eating: Tracing misinformation in 'What I Eat In A Day' videos—Justine Topham, Naomi Smith, 2023. *Journal of Sociology*, *59*(3). https://doi.org/10.1177/144078332311613

Tormala, Z. L., DeSensi, V. L., Clarkson, J. J., & Rucker, D. D. (2009). Beyond attitude consensus: The social context of persuasion and resistance. *Journal of Experimental Social Psychology*, *45*(1), 149–154. https://doi.org/10.1016/j.jesp.2008.07.020

Tormala, Z. L., & Petty, R. E. (2004). Source Credibility and Attitude Certainty: A Metacognitive Analysis of Resistance to Persuasion. *Journal of Consumer Psychology*, *14*(4), 427–442. https://doi.org/10.1207/s15327663jcp1404_11

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLOS ONE*, *13*(9), e0203958. https://doi.org/10.1371/journal.pone.0203958

Traberg, C., Morton, T., & van der Linden, S. (2024). Counteracting socially endorsed misinformation through an emotion-fallacy inoculation. *Advances.in/Psychology*, *2*, e765332. https://doi.org/10.56296/aip00017

Traberg, C. S. (2019). *What's in a 'Like'? Examining the effects of social proof on perceptions of and resistance to persuasion by misinformation*. University of Cambridge.

Traberg, C. S. (2022). Misinformation: Broaden definition to curb its societal influence. *Nature*, *606*(7915), 653. https://doi.org/10.1038/d41586-022-01700-4

Traberg, C. S. (2024). Coercion by Misinformation: Challenges and Solutions. In *Coercion and Trust: A Multi-Disciplinary Dialogue* (Vol. 1). Routledge.

Traberg, C. S., Harjani, T., Basol, M., Biddlestone, M., Maertens, R., Roozenbeek, J., & van der Linden, S. (2023). Prebunking Against Misinformation in the Modern Digital Age. In T. D. Purnat, T. Nguyen, & S. Briand (Eds.), *Managing Infodemics in the 21st Century: Addressing New Public Health Challenges in the Information Ecosystem*. Springer International Publishing. https://doi.org/10.1007/978-3-031-27789-4

Traberg, C. S., Harjani, T., Roozenbeek, J., & van der Linden, S. (2024). The persuasive effects of social cues and source effects on misinformation susceptibility. *Scientific Reports*, *14*(1), Article 1. https://doi.org/10.1038/s41598-024-54030-y

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). https://doi.org/10.1177/00027162221087936

Traberg, C. S., & van der Linden, S. (2022). Birds of a feather are persuaded together: Perceived source credibility mediates the effect of political bias on misinformation susceptibility. *Personality and Individual Differences*, *185*(111269).

Tsang, S. J. (2020). Motivated Fake News Perception: The Impact of News Sources and Policy Support on Audiences' Assessment of News Fakeness. *Journalism & Mass Communication Quarterly*. https://doi.org/10.1177/1077699020952129

Tsfati, Y., Boomgaarden, H. G., Strömbäck, J., Vliegenthart, R., Damstra, A., & Lindgren, E. (2020). Causes and consequences of mainstream media dissemination of fake news: Literature review and synthesis. *Annals of the International Communication Association*, *44*(2), 157–173. https://doi.org/10.1080/23808985.2020.1759443

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory* (pp. x, 239). Basil Blackwell.

Turner, J. C., & Oakes, P. J. (1989). Self-categorization theory and social influence. In *Psychology of group influence, 2nd ed* (pp. 233–275). Lawrence Erlbaum Associates, Inc.

Tversky, A., & Kahneman, D. (1974). *Judgement under Uncertainty: Heuristics and Biases*. *185*.

Tyler, S. W., Hertel, P. T., McCallum, M. C., & Ellis, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(6), 607–617. https://doi.org/10.1037/0278-7393.5.6.607

Van Bavel, J. J., Hackel, L. M., & Xiao, Y. J. (2014). The Group Mind: The Pervasive Influence of Social Identity on Cognition. In J. Decety & Y. Christen (Eds.), *New Frontiers in Social Neuroscience* (pp. 41–56). Springer International Publishing. https://doi.org/10.1007/978-3-319-02904-7_4

Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., & Tucker, J. A. (2021). Political Psychology in the Digital (mis)Information age: A Model of News Belief and Sharing. *Social Issues and Policy Review*, *15*(1), 84–113. https://doi.org/10.1111/sipr.12077

Van Bavel, J. J., & Pereira, A. (2018). The Partisan Brain: An Identity-Based Model of Political Belief. *Trends in Cognitive Sciences*, *22*(3), 213–224. https://doi.org/10.1016/j.tics.2018.01.004

Van Bavel, J. J., Rathje, S., Vlasceanu, M., & Pretus, C. (2024). Updating the identity-based model of belief: From false belief to the spread of misinformation. *Current Opinion in Psychology*, *56*, 101787. https://doi.org/10.1016/j.copsyc.2023.101787

van der Linden, S. (2015). The conspiracy-effect: Exposure to conspiracy theories (about global warming) decreases pro-social behavior and science acceptance. *Personality and Individual Differences*, *87*, 171–173. https://doi.org/10.1016/j.paid.2015.07.045

van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, *28*(3), Article 3. https://doi.org/10.1038/s41591-022-01713-6

van der Linden, S. (2024). Countering misinformation through psychological inoculation. *Advances in Experimental Social Psychology*. https://doi.org/10.1016/bs.aesp.2023.11.001

van der Linden, S., Albarracín, D., Fazio, L. K., Freelon, D., Roozenbeek, J., Swire-Thompson, B., & Van Bavel, J. (2023). *Using Psychological Science to Understand and Fight Health Misinformation: An APA Consensus Statement* (An APA Consensus Statement). American Psychological Association.

van der Linden, S., Dixon, G., Clarke, C., & Cook, J. (2021). Inoculating against COVID-19 vaccine misinformation. *EClinicalMedicine*, *33*, 100772. https://doi.org/10.1016/j.eclinm.2021.100772

van der Linden, S., & Kyrychenko, Y. (2024). A broader view of misinformation reveals potential for intervention. *Science (New York, N.Y.)*, *384*(6699), 959–960. https://doi.org/10.1126/science.adp9117

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

van der Linden, S., Panagopoulos, C., & Roozenbeek, J. (2020). You are fake news: Political bias in perceptions of fake news. *Media, Culture & Society*, *42*(3). https://doi.org/10.1177/0163443720906992

van der Linden, S., & Roozenbeek, J. (2020). Psychological Inoculation Against Fake News. In *The Psychology of Fake News*. Routledge.

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 Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*(3), 813–826. https://doi.org/10.3758/s13423-020-01798-5

van Knippenberg, D., Lossie, N., & Wilke, H. (1994). In-group prototypicality and persuasion: Determinants of heuristic and systematic message processing. *British Journal of Social Psychology*, *33*(3), 289–300. https://doi.org/10.1111/j.2044-8309.1994.tb01026.x

Vlasceanu, M., Dyckovsky, A. M., & Coman, A. (2024). A Network Approach to Investigate the Dynamics of Individual and Collective Beliefs: Advances and Applications of the BENDING Model. *Perspectives on Psychological Science*, *19*(2), 444–453. https://doi.org/10.1177/17456916231185776

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Vraga, E. K., & Bode, L. (2020). Defining Misinformation and Understanding its Bounded Nature: Using Expertise and Evidence for Describing Misinformation. *Political Communication*, *37*(1), 136–144. https://doi.org/10.1080/10584609.2020.1716500

Wallace, L. E., Wegener, D. T., & Petty, R. E. (2020). When Sources Honestly Provide Their Biased Opinion: Bias as a Distinct Source Perception With Independent Effects on Credibility and Persuasion. *Personality and Social Psychology Bulletin*, *46*(3), 439–453. https://doi.org/10.1177/0146167219858654

Walther, J. B., Liang, Y. (Jake), Ganster, T., Wohn, D. Y., & Emington, J. (2012). Online Reviews, Helpfulness Ratings, and Consumer Attitudes: An Extension of Congruity

Theory to Multiple Sources in Web 2.0. *Journal of Computer-Mediated Communication*, *18*(1), 97–112. https://doi.org/10.1111/j.1083-6101.2012.01595.x

Watts, D. J., Rothschild, D. M., & Mobius, M. (2021). Measuring the news and its impact on democracy. *Proceedings of the National Academy of Sciences*, *118*(15), e1912443118. https://doi.org/10.1073/pnas.1912443118

Weimann, G. (1994). *The Influentials: People Who Influence People*. SUNY Press.

West, J. D., & Bergstrom, C. T. (2021). Misinformation in and about science. *Proceedings of the National Academy of Sciences*, *118*(15), e1912444117. https://doi.org/10.1073/pnas.1912444117

Westbrook, V., Wegener, D. T., & Susmann, M. W. (2023). Mechanisms in continued influence: The impact of misinformation corrections on source perceptions. *Memory & Cognition*, *51*(6), 1317–1330. https://doi.org/10.3758/s13421-023-01402-w

Williams, B. A., & Carpini, M. X. D. (2011). *After Broadcast News: Media Regimes, Democracy, and the New Information Environment*. Cambridge University Press.

Wittenbrink, B., & Henly, J. R. (1996). Creating social reality: Informational social influence and the content of stereotypic beliefs. *Personality and Social Psychology Bulletin*, *22*(6), 598–610. https://doi.org/10.1177/0146167296226005

Wood, M. L. M. (2007). Rethinking the inoculation analogy: Effects on subjects with differing preexisting attitudes. *Human Communication Research*, *33*(3), 357–378. https://doi.org/10.1111/j.1468-2958.2007.00303.x

Wood, W. (2000). Attitude Change: Persuasion and Social Influence. *Annual Review of Psychology*, *51*(1), 539–570. https://doi.org/10.1146/annurev.psych.51.1.539

Woodside, A. G., & Davenport, J. W. (1974). The Effect of Salesman Similarity and Expertise on Consumer Purchasing Behavior. *Journal of Marketing Research*, *11*(2), 198–202. https://doi.org/10.2307/3150562

World Economic Forum. (2018). *The Global Risks Report 2018, 13th Edition*. https://reliefweb.int/report/world/global-risks-report-2018-13th-edition

World Economic Forum. (2024). *Global Risks Report*.

Xu, X., & Fu, W. W. (2014). Aggregate Bandwagon Effects of Popularity Information on Audiences' Movie Selections. *Journal of Media Economics*, *27*(4), 215–233. https://doi.org/10.1080/08997764.2014.963229

Yanagizawa-Drott, D. (2014). Propaganda and Conflict: Evidence from the Rwandan Genocide. *The Quarterly Journal of Economics*, *129*(4), 1947–1994. https://doi.org/10.1093/qje/qju020

Zhou, D. X., Resnick, P., & Mei, Q. (2011). Classifying the Political Leaning of News Articles and Users from User Votes. *Proceedings of the International AAAI Conference on Web and Social Media*, *5*(1), Article 1. https://doi.org/10.1609/icwsm.v5i1.14108

Ziemer, C.-T., Schmid, P., Betsch, C., & Rothmund, T. (2024). Identity is key, but Inoculation helps – how to empower Germans of Russian descent against pro-Kremlin disinformation. *Advances.in/Psychology*, *2*, e628359. https://doi.org/10.56296/aip00015

10. APPENDIX

1. Appendix A

Pilot Study Methods

For the pilot study, $N = 656$ people were recruited (36% female, 55.8% male, 7.8% other, 51.1% ages 18-29, 35.2% ages 30-49, 13.7% ages 50+, 75.6% liberals, 56.9% higher educated) via an online platform following a university-wide press release in relation to a separate study. As such, anyone with an internet connection could participate in the pilot study. Upon providing informed consent, participants were asked to select one of three random images, following which they were automatically redirected to their respective conditions. Using a between-subjects design, participants were assigned to one of three source ideology conditions: Liberal Source, Conservative Source or a control in which the source of the news article was hidden. Participants then proceeded to rate the reliability of a series of news headlines, which were counterbalanced across conditions with source being the manipulated variable. Results of the exploratory analyses are outlined below. Due to the assignment of conditions being determined by the selected image of the participants, groups were unbalanced (Liberal Source $n = 211$, Conservative Source $n = 281$, Control $n = 164$). Furthermore, due to the nature of the open-access survey platform, the sample was also highly skewed (towards Liberals) with Liberal $n=496$ and Conservative $n = 160$.

Pilot Study Results

H_A Political ideology will moderate the effect of source ideology on reliability judgements of misinformation such that individuals identifying as conservative will be more likely than liberals to judge misinformation from liberal sources as being unreliable, and individuals identifying as liberals will be more likely than conservatives to judge misinformation from conservative sources as being unreliable.

To test H_A, a factorial ANOVA was run with accuracy judgements of misinformation as DV, and source ideology and political ideology as IVs, yielding a main effect for source ideology, $F(1,484) = 26.58$, $p < 0.001$, partial $\eta^2 = 0.05$, a main effect for political ideology, $F(1,484) = 8.99$, $p < 0.01$, partial $\eta^2 = 0.02$. The interaction effect between source ideology and political ideology was significant, $F(4,484) = 7.64$, $p < 0.01$, partial $\eta^2 = 0.02$. A simple slopes analysis further revealed that political ideology was only significant for

misinformation from conservative sources: Accuracy judgements of misinformation from conservative sources was significantly higher for conservative participants (M = 3.57, 95% CI [3.29,3.86]) than for liberal participants, (M = 2.81, 95% CI [2.63,2.98]), $p < 0.001$, d = 0.47, whereas accuracy judgements of misinformation from liberal sources did not differ significantly between conservative (M = 3.58, 95% CI [3.18,3.98]), and liberal participants (M = 3.6, 95% CI [3.40,3.79]), p = 0.94, $d < 0.01$. Among conservatives, source did not have a significant impact on accuracy judgements, with conservative judgements not differing between misinformation from conservative sources (M = 3.57, 95% CI [3.29, 3.86]) compared to misinformation from liberal sources (M = 3.58, 95% CI [3.18, 3.98]), p = 0.97, d = 0.03. Liberals participants, however, judged misinformation from liberal sources to be significantly more accurate (M = 3.6, 95% CI [3.40, 3.79]) than that from conservative sources (M = 2.81, 95% CI [2.63,2.98]), $p < 0.001$, d = 0.30.

Post-hoc Tests to Rule Out Content Effects

Although H_A in the pilot study largely found similar results to the main study, indicating that the effects observed in the main study were driven by *source* and not *content* of the news items (as the pilot study employed a between-subjects design with counterbalancing and a control condition) we ran further systematic tests to more confidently rule out this possibility.

Post-hoc Test 1

In the first post-hoc test, we examined whether liberal participants and conservative participants in the control (no source) condition differed in their judgements of items used in the conservative misinformation condition and liberal misinformation condition in the main study. If indeed there were systematic differences in that conservatives judged headlines which were subsequently used in the main study as conservative misinformation as significantly more reliable than headlines subsequently used as liberal misinformation (and vice versa) this would indicate that the headline content was politically biased and would suggest the effects observed in the main study may not exclusively be driven by source effects.

However, this is not what we observe. Two independent samples t-tests showed that while compared to conservative participants (M = 3.55, SD = 1.61), liberal participants *(*M = 2.80, SD = 1.20) rated misinformation subsequently used in the conservative source condition as significantly less reliable, $t(45.5) = -2.58$, $p < 0.05$, d = 0.57, they also rated

misinformation subsequently used in the liberal source condition (M = 2.56, SD = 1.11) as significantly less reliable than conservative participants (M = 3.17, SD=1.69), t(43.20) = -2.05, $p < 0.05$, d = 0.48, indicating that no systematic political bias in the items themselves was present.

Post-hoc Test 2

In the second pilot post-hoc test, we examined whether both liberal participants reliability judgements and conservative participants reliability judgements of the items subsequently used as conservative vs liberal source misinformation differed significantly in the control condition, where no source was included. If we in fact did see that liberal participants judged headlines subsequently used for the liberal source ideology condition to be more reliable than the headlines subsequently used for the conservative source ideology condition, this would indicate a potential issue. Likewise, if we found that conservative participants judged headlines subsequently used for the conservative source ideology condition to be more reliable than the headlines subsequently used for the liberal source ideology condition, this may indicate that findings in the main study were driven by content, not source. However, this is not what we find. Paired samples t-tests showed that that conservative reliability judgements of items subsequently used for the conservative source ideology condition (M = 3.55, SD = 1.61) were not significantly different from the items used for the liberal source ideology condition (M = 3.17, SD = 1.69), t(34) = -2.01, p = 0.052, d = 0.25.

Furthermore, while there was a difference observed for liberal participants, the direction of this effect was the opposite of one that may have been concerning: Liberal participants judged misinformation subsequently used for the conservative condition as significantly more reliable (M = 2.77, SD = 1.21) than the items subsequently used for the liberal condition (M = 2.55, SD = 1.11), t(116) = -2.39, $p < 0.05$, d = 0.21; this therefore this could not explain the effects observed in the main study. If anything, this only suggests an even stronger effect, as without a source, liberal participants in fact found misinformation used in the conservative condition in the main study to be more reliable than the misinformation used in the liberal condition.

Post-hoc Test 3

Finally, we wanted to assess whether, on the whole, items subsequently used in the main study for the liberal source ideology condition were judged to be more reliable than the

items subsequently used for the conservative source ideology condition. If indeed it were the case that the finding that all participants in the main study judged misinformation from liberal sources as more accurate was driven by the content, not the source, we would expect to see the same results in the pilot study control condition where no source was present. However, a t-test across all participants in the control condition showed that compared to items subsequently used for the liberal source ideology condition ($M = 2.69$, $SD = 1.29$), the items subsequently used for the conservative source ideology condition ($M = 2.95$, $SD = 1.35$) were in fact judged to be more reliable ($t(151) = -3.10$, $p < 0.01$, $d = 0.22$), further supporting the conclusion that the observed effects in the main study were not driven by item content, but indeed the source.

2. Appendix B

Table A1

Chapter 4 Analyses with Exclusions Applied

Hypothesis (H)	Manipulation	DV	Result
Social cues impact perceived reliability of misinformation	Explicit (previous group judgements) & implicit social cues (comments)	Perceived reliability	$F(4,537) = 10.25$, $p < 0$ .001. Explicit contrast: $p < 0.001$, $d = 0.66$. Implicit contrast: $p = 0.01$, $d = 0.34$).
Social cues impact perceived truth in misinformation	Explicit (previous group judgements) & implicit social cues (comments)	Perceived truth	$F(4,537) = 8.48$ $p <$ 0 .001. Explicit contrast: $p < 0.001$, $d = 0.66$. Implicit contrast: $p = 0.06$, $d = 0.27$).
Social cues impact perceptions of public consensus in reliability of misinformation	Explicit (previous group judgements) & implicit social cues (comments)	Perceived public reliability	$F(4,537) = 56.56$, $p < .001$. Explicit contrast: $p < 0.001$, $d = 1.76$. Implicit contrast: $p < 0.001$, $d =$-0.56.
Social cues impact perceptions of public consensus in truth in misinformation	Explicit (previous group judgements) & implicit social cues (comments)	Perceived public truth	$F(4,537) = 54.97$, $p < .001$. Explicit contrast: $p < 0.001$, $d = 1.91$. Implicit contrast: $p < 0.001$, $d = 0.81$.
Perceived consensus predicts perceived reliability of misinformation	Explicit (previous group judgements) & implicit social cues (comments)	Perceived public consensus	(Adjusted $R^2 = 0.14$, $F(1,540) = 89.92$, $p < 0.001$)
Perceived consensus mediates the effect of social cues on perceived reliability of misinformation	Explicit social cues (previous group judgements)	Perceived public consensus	Indirect effect: $b = 0.267$, $p < 0.001$, 95% CI [0.144,0.429].
Perceived consensus mediates the effect of social cues on perceived reliability of misinformation	Implicit social cues (comments)	Perceived public consensus	Indirect effect: $b = 0.715$, $p < 0.001$, 95% CI [0.468,0.996].

3. Appendix C

Correlation Analyses in Chapter 4

Variables	**df**	**t**	**Pearson correlation**	***p*-value**
Social Cue (Implicit Discrediting vs Endorsement) by Perceived Consensus	290	4.48	*0.25*	*<0.001*
Social Cue (Explicit Discrediting vs Endorsement) by Perceived Consensus	288	15.45	*0.67*	*<0.001*
Perceived Consensus (in Social Cue Conditions Implicit Discrediting vs Endorsement) by Perceived Reliability	290	1.38	*0.08*	*<0.001*
Perceived Consensus (in Social Cue Conditions Explicit Discrediting vs Endorsement) by Perceived Reliability	228	4.49	*0.26*	*<0.001*

4. Appendix D

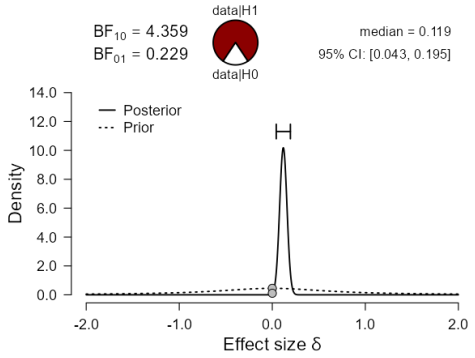
To test **H2**, we conducted a Bayesian paired-samples *t*-test on the pre- and post-inoculation reliability judgements of true news headlines. Doing so shows that the perceived reliability of true news is reduced post-inoculation compared to pre-inoculation (BF_{10} = 4.359, indicating moderate support for the alternative hypothesis). However, we note that this effect is very small (δ = .119, which represents the population-level version of Cohen's *d*). Furthermore, an exploratory Bayesian paired-samples *t*-test on the pre- and post-scores for the difference between average reliability ratings for misinformation and true news (i.e., veracity discernment) yields strong support for the alternative hypothesis ($BF_{10} = 2.49*10^{13}$, δ = .332), meaning that inoculation strongly reduces the perceived reliability of misinformation relative to true news. See Table S2 and Figures S1 and S3 for a full overview, including how the models were specified within Jamovi.

Using a frequentist approach, a paired-samples *t*-test showed that the difference between perceived reliability of factual headlines pre and post intervention, $t(1303) = 4.31$, $M_{diff} = 0.13$, $SE = 0.03$, $p < 0.001$, but only at a small effect size of $d = 0.12$.

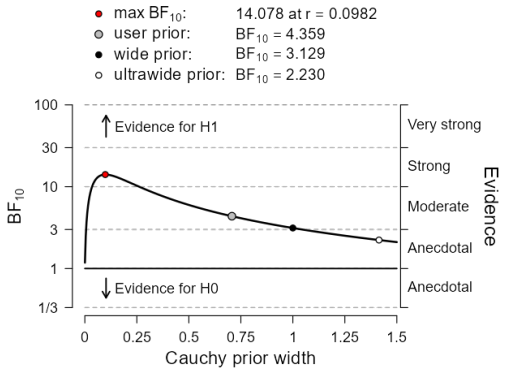
Figure S1. *Bayes Factors and effect sizes (left panels) and robustness checks with various priors and levels of support for the alternative hypothesis H1 (right panels) for true news, misinformation, and veracity discernment. See also Table S2 and Figure S3.*

True News

Prior and Posterior

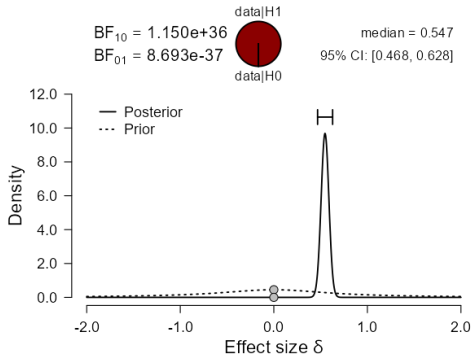


Bayes Factor Robustness Check

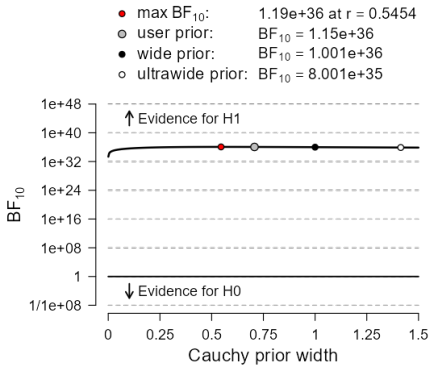


Misinformation

Prior and Posterior

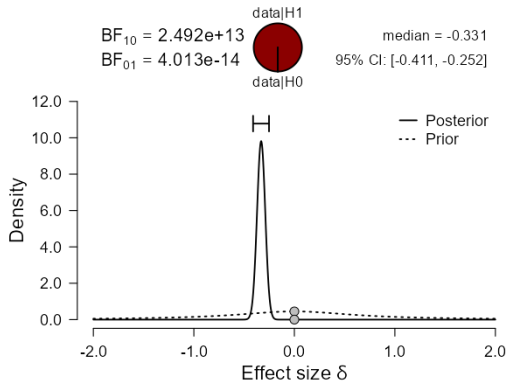


Bayes Factor Robustness Check



Discernment (difference true news – misinformation)

Prior and Posterior



Bayes Factor Robustness Check

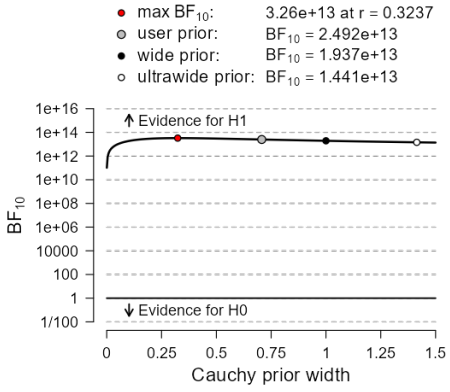


Figure S2. *Screenshot of Bayesian paired-samples t-tests run in Jamovi (www.jamovi.org), for replication purposes. Dataset used is "complete_data.csv" found on the OSF. Analysis is "Bayesian paired-samples t-test", which can be run inside Jamovi after installing the "jsq" package, from the T-Tests tab. See also Table S2 and Figure S1 for plots generated using this specification.*

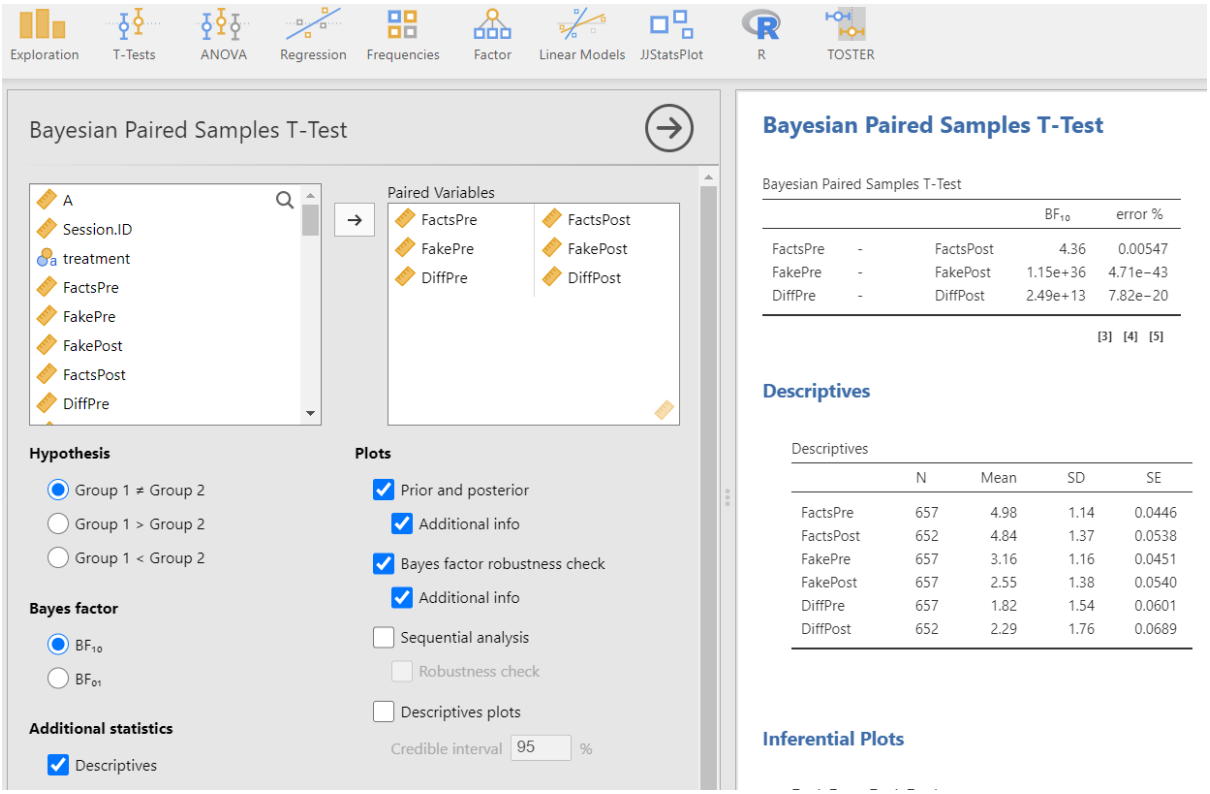
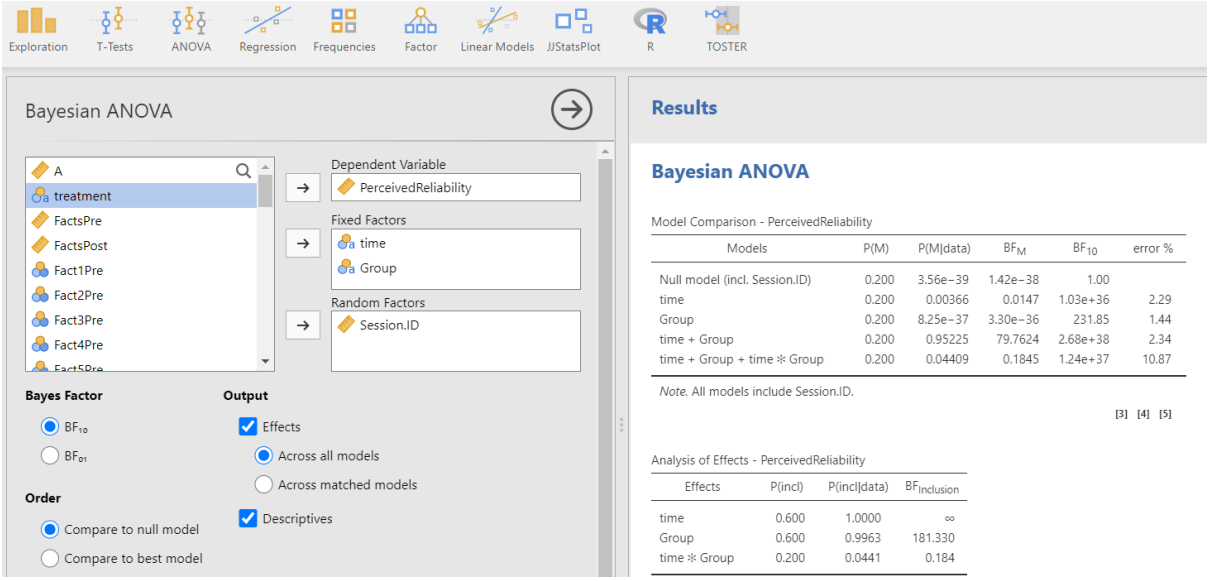


Figure S3. *Screenshot of Bayesian ANOVA analysis run in Jamovi (www.jamovi.org), for replication purposes. Dataset used is "long_dataperceivedgroup.csv" found on the OSF. Analysis is "Bayesian ANOVA", which can be run inside Jamovi after installing the "jsq" package, from the ANOVA tab. See also Table S1; minor variations in probabilities and Bayes factors are due to differences in model convergence.*



\

Table S1.

*Bayesian ANOVA results for the interaction between experimental condition (Group) and time (pre-post intervention). P(M) represents the prior probabilities of each model; since we have 5 models, this is 1/5 = 0.2 for each model. P(M|data) signifies the posterior probability of each model after seeing the data. BF_m is a Bayes Factor that compares each model to the P(M|data) of the other models. BF_{10} gives the probability of the data if H1 is true, P(D|H1). The inclusion Bayes Factor ($BF_{Inclusion}$) represents the evidence in the data for including a predictor (the interaction group * time).*

Model Comparison - PerceivedReliability (Perceived reliability of misinformation)

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. Session.ID)	0.2	3.59E-39	1.44E-38	1	
Group	0.2	8.21E-37	3.28E-36	228.86	1
time	0.2	0.00362	0.0145	1.01E+36	1.16
Group + time	0.2	0.95949	94.7437	2.67E+38	2.85
Group + time + Group ✻ time	0.2	0.03689	0.1532	1.03E+37	1.57

Note. All models include Session.ID.

Analysis of Effects - PerceivedReliability

Effects	P(incl)	P(incl|data)	$BF_{Inclusion}$
Group	0.6	0.9964	183.583
time	0.6	1	6.00E+15
Group ✻ time	0.2	0.0369	**0.153**

Note: Analysis was conducted using the "JSQ" package in Jamovi (https://www.jamovi.org). Specifically, we ran a Bayesian ANOVA with the perceived reliability of misinformation (PerceivedReliability) as the dependenet variable, experimental condition (Group) and time (pre-post) as fixed factors, and Session.ID (i.e., Participant ID) as a random effect. We used the long-format dataset available on the OSF.

Table S2. *Paired-samples Bayesian t-tests for the pre-post difference in the perceived reliability of true news (FactsPre/FactsPost), misinformation (FakePre/FakePost), and the difference between ratings of true news and misinformation, i.e., "veracity discernment" (DiffPre/DiffPost). Bottom table shows descriptive statistics.*

			BF_{10}	error %
FactsPre	-	FactsPost	4.36	0.00547
FakePre	-	FakePost	1.15E+36	4.71E-43
DiffPre	-	DiffPost	2.49E+13	7.82E-20

Descriptives

	N	Mean	SD	SE
FactsPre	657	4.98	1.14	0.0446
FactsPost	652	4.84	1.37	0.0538
FakePre	657	3.16	1.16	0.0451
FakePost	657	2.55	1.38	0.054
DiffPre	657	1.82	1.54	0.0601
DiffPost	652	2.29	1.76	0.0689

Table S3. *Sample composition in percentages of total sample.*

Sample Composition in Percentages of Total Sample

Education	Gender	Age	Political Ideology
High school or less (16.3%)	Male (51.3%)	18-29 (70.6%)	1 (10.2%)
Higher degree (43.8%)	Female (43.2%)	30-49 (23.7%)	2 (32.1%)
Some College (39.9%)	Other (5.5%)	Over 50 (5.6%)	3 (36.8%)
			5 (11.4%)
			6 (3.8%)
			7 (5.6%)