Routledge Studies in Epistemology

MISINFORMATION, CONTENT MODERATION, AND EPISTEMOLOGY

PROTECTING KNOWLEDGE

Keith Raymond Harris

Misinformation, Content Moderation, and Epistemology

This book argues that misinformation poses a multifaceted threat to knowledge, while arguing that some forms of content moderation risk exacerbating these threats. It proposes alternative forms of content moderation that aim to address this complexity while enhancing human epistemic agency.

The proliferation of fake news, false conspiracy theories, and other forms of misinformation on the internet and especially social media is widely recognized as a threat to individual knowledge and, consequently, to collective deliberation and democracy itself. This book argues that misinformation presents a three-pronged threat to knowledge. While researchers often focus on the role of misinformation in causing false beliefs, this deceptive potential of misinformation exists alongside the potential to suppress trust and to distort the perception of evidence. Recognizing the multifaceted nature of this threat is essential to the development of effective measures to mitigate the harms associated with misinformation. The book weaves together work in analytic epistemology with emerging empirical work in other disciplines to offer novel insights into the threats posed by misinformation. Additionally, it breaks new ground by systematically assessing different forms of content moderation from the perspective of epistemology.

Misinformation, Content Moderation, and Epistemology will appeal to philosophers working in applied and social epistemology, as well as scholars and advanced students in disciplines such as communication studies, political science, and social psychology who are researching misinformation.

Keith Raymond Harris is a postdoctoral researcher in philosophy at Ruhr-Universität Bochum, where he specializes in social and applied epistemology and the philosophy of cognitive science. His recent publications include "Real Fakes: The Epistemology of Online Misinformation," "Epistemic Domination," and "Beyond Belief: On Disinformation and Manipulation."

Routledge Studies in Epistemology

Illuminating Errors
New Essays on Knowledge from Non-Knowledge
Edited by Rodrigo Borges and Ian Schnee

Digital Knowledge
A Philosophical Investigation
J. Adam Carter

Seemings and the Foundations of Justification
A Defense of Phenomenal Conservatism
Blake McAllister

Trust Responsibly
Non-Evidential Virtue Epistemology
Jakob Ohlhorst

Rationality in Context
Unstable Virtues in an Uncertain World
Steven Bland

Seemings
New Arguments, New Angles
Edited by Kevin McCain, Scott Stapleford, and Matthias Steup

The Epistemic Injustice of Genocide Denialism
Melanie Altanian

Misinformation, Content Moderation, and Epistemology
Protecting Knowledge
Keith Raymond Harris

For more information about this series, please visit: https://www.routledge.com/Routledge-Studies-in-Epistemology/book-series/RSIE

Misinformation, Content Moderation, and Epistemology

Protecting Knowledge

Keith Raymond Harris

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2024
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Funded by the Ministry of Culture and Science of the Federal State of North Rhine Westphalia, Germany.

ISBN: 978-1-032-63687-0 (hbk)
ISBN: 978-1-032-63691-7 (pbk)
ISBN: 978-1-032-63690-0 (ebk)

DOI: 10.4324/9781032636900

Typeset in Sabon
by SPi Technologies India Pvt Ltd (Straive)

To my parents and to Anna

Contents

Acknowledgments

Many people and institutions helped to make this book possible. I would first like to thank Tobias Schlicht, who gave me the opportunity to write this book and whose seminar on social media sparked many of the ideas developed here. I would also like to thank participants in several conferences—including the 2022 Political Epistemology Network, the Visual Trust conference on Image-making, Changing Minds Online, and the Rijeka conference on Hate Speech, Fake News, and Freedom of Speech. Their comments on earlier material helped to give shape to this book.

I would like to thank Andrew Weckenmann for his early interest in and support for the project and Rosaleah Stammler for her support throughout the publication process. I would also like to thank four anonymous referees for their feedback on early material for this book, which helped to shape the direction of the text. Thanks also to a further anonymous referee, whose extensive comments on an earlier draft of the manuscript proved a substantial help.

I owe a special thanks to the Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen for funding the INTERACT! project (PROFILNRW-2020-135) at Ruhr-Universität Bochum, through which this project, including its open access content, was funded.

Last, but not least, thanks to my dog, Buffy, who was a constant presence and an occasional welcome distraction throughout the writing process.

Introduction
Knowledge under threat

Some time ago, I returned to the United States to attend a friend's wedding. Early in my pre-wedding haircut, I made a mistake with which many philosophers will be familiar: I mentioned that I was a philosopher. Worse still, I mentioned that I was a philosopher whose work dealt, in part, with conspiracy theories. This stopped my hairdresser mid-snip. "Have you heard of NESARA?" she asked eagerly. NESARA, short for National Economic Security and Recovery Act, refers to a supposed plan to radically overhaul the US financial system and is highly anticipated in certain fringe corners of the internet. Conversational journeys that start with NESARA don't typically circle back to the safer shores of typical small talk and, over the course of my haircut, I was apprised of theories about government conspiracies and hidden extra-terrestrials, coupled with the insistence that "It's not about red versus blue, it's about light versus dark." Because it is not advisable to argue from the hairdresser's chair, I kept my protestations to the gentle suggestion that "It's good to keep an open mind."

Many readers have probably experienced similar interactions. Indeed, for many of us, run-ins with extraordinary beliefs in both our personal lives and in political contexts appear to be increasingly common. It is widely thought that such beliefs reflect the influence of misinformation, and especially the prevalence of outlandish misinformation on social media. Especially because some such beliefs appear to be dangerous, it is often thought that there is cause for measures—including the use of social media content moderation—to suppress the spread of misinformation. But the sensational nature of such beliefs, which often attract an almost voyeuristic fascination, as well as dramatic events that—like assault on the US Capitol in January 2021—appear to be inspired by such beliefs, can distract from some of the more subtle effects of misinformation. In this book, I argue that, without a more complete understanding of the effects of misinformation, attempts to deal with the challenge may exacerbate its ill-effects. In particular, I argue that heavy-handed attempts to suppress misinformation through content moderation run the risk of amplifying some of its worst effects.

DOI: 10.4324/9781032636900-1

This book is a study in epistemology, that is, the theory of knowledge. I thus focus on the effects of misinformation and content moderation on the attainment of knowledge. The aim of this book is to explore how misinformation threatens knowledge and how content moderation can either mitigate or worsen this threat. Ultimately, I argue that careful attention to knowledge and the conditions for its attainment can help to inform approaches to content moderation that extend, rather than supplant, the agency of social media users.

I.1 The promise of social media and the threats of misinformation

Social media is commonly defined in terms of its user-driven nature (Fuchs, 2014; Kaplan & Haenlein, 2010, p. 61). In contrast to legacy media, which is characterized by one-directional messaging from the few to the many, social media allows ordinary people to generate and broadcast their own content to large audiences. For many people, social media is creatively empowering. Twitter users can write and share jokes that go viral and come to the attention of millions of other users. Amateur photographers and other artists can use Instagram to share their work. TikTok users can record videos that help them to share their tastes and creativity.

Social media is not just creatively empowering, but *epistemically* empowering. Aided by other technologies, especially smartphones and their built-in cameras, social media allows for the rapid distribution of information. Thus, for example, social media has been instrumental in the distribution of evidence of police misconduct and has thereby shaped subsequent protest movements and calls for greater accountability. Generalizing beyond such concrete cases, the epistemic power of social media consists, in part, in the enhanced ability of ordinary individuals to easily broadcast information to recipients across the planet. Social media platforms thereby go beyond merely connecting people. By allowing for the rapid spread of information, such platforms allow individuals to open windows into their own worlds, through which countless others can look.

But the social dimensions of social media go further still. Social media provides not only for the sharing of experience, but for the shared experience of sharing experiences. When we encounter jokes on Twitter, we thereby connect not only to the joke-teller, but also to the many other people who encounter the same joke. Often, as one watches a livestream on social media, one simultaneously sees the live commentary and emoji reactions of other users. Social media platforms, then, are both conduits for information and shared theaters in which to consume it.

As I will explore throughout Part I, this dual nature of social media platforms gives rise to many epistemic challenges that exist alongside the benefits highlighted thus far. We will see, for example, that the visibility of others'

reactions has, sometimes for better and sometimes for worse, the potential to guide our interpretations of content accessed on social media. More immediately, social media allows not only for the rapid spread of creative products and valuable information, but also for the spread of misinformation.

That the spread of misinformation on social media represents a significant threat to individual and collective knowledge, to decision-making, and even to democracy is a point largely taken for granted in some quarters. Thus, the challenge of misinformation and how best to confront it has been widely discussed among journalists, academics, and politicians, especially since 2016—a year which saw both Brexit and the election of Donald Trump to US President. However, it is worth acknowledging from the outset that the epistemological and political significance of misinformation has recently faced significant scrutiny. Many academics have argued that concerns about the spread and influence of misinformation have the character of a "moral panic" (Altay et al., 2023; Carlson, 2020; Jungherr & Schroeder, 2021) in which the actual impacts of misinformation are blown out of proportion. Some key data in support of this revisionary response include the observations that misinformation is principally consumed (Grinberg et al., 2019) and shared (Osmundsen et al., 2021, p. 1005) by a small minority of social media users.

Although the revisionary response is a useful corrective to sometimes exaggerated claims as to the severity of the misinformation problem, there are several reasons to think this response is somewhat too optimistic. First, studies of the prevalence of misinformation tend to focus on extreme cases. For example, studies of the spread of fake news and other forms of misinformation tend to use relatively narrow definitions of the term such that only content from consistently unreliable publishers counts as fake news. Second, many studies of misinformation focus narrowly on fake news and may underappreciate the impacts of other forms of misinformation. Third, even if few people are deceived by misinformation, small numbers may be enough to have major impacts in some cases, including in the contexts of competitive elections (van der Linden, 2023). Fourth, studies of the impacts of misinformation tend to focus on the western context, but these impacts may be especially potent in geopolitical contexts in which media ecosystems are comparatively underdeveloped. Fifth, even if the role of misinformation in driving large-scale political events is sometimes overstated, misinformation seems to have important effects on individuals' private lives. A subreddit titled r/QAnonCasualties, for example, details stories of estrangement and the destruction of personal relationships due to loved ones' devotion to the QAnon conspiracy theory. Such private costs are difficult to study, but of great significance to the people that experience them. Finally, it is consistent with the concrete impacts of misinformation being limited that the specter of misinformation reduces individuals' trust in online information, generating subtle or extreme doubts that compromise

their abilities to enjoy the full epistemic benefits of social media. This last point, it should be noted, cuts in multiple directions. Both misinformation and exaggerated reports as to the prevalence and impact of misinformation can drive a loss of trust.

In fact, I argue in what follows that reduced trust is one of three prongs by which misinformation threatens knowledge. In the next three sections, I briefly introduce the three-pronged threat of misinformation by way of example and analogy.

I.2 The gunman

A criminal affidavit reveals that, on December 1, 2016, Edgar Maddison Welch received the following text message:

> Tell me we r going to save the Indians from the pipeline.

Welch responded with the following pair of messages:

> Way more important, much higher stakes
> Pizzagate.

As the surrounding string of messages indicates, Welch had been convinced by online misinformation of the need to take drastic action. Whereas Welch's text exchanges suggest some pre-existing interest in protesting against the Dakota Access Pipeline, this interest was quickly superseded by a plan that would land Welch in police custody and would bring international attention to the false and outlandish Pizzagate conspiracy theory.

The substance of the Pizzagate conspiracy theory is that high-profile Democratic politicians and operatives used a range of locations, including the Comet Ping Pong pizzeria in Washington, D.C., for sex trafficking children. Welch arrived, armed, at Comet Ping Pong just three days after participating in the text exchange above. As he drove to the pizzeria, he recorded messages for his family. These recordings included the following message to his children:

> Like I always told you we have a duty to protect people who can't protect themselves…I hope you understand that one day.

Welch's conduct reflects a sincere belief in the Pizzagate conspiracy theory—a belief developed through consumption of misinformation on YouTube and other online platforms.

While Welch's incursion into Comet Ping Pong fortunately ended without injury to himself or any patrons or employees, the case illustrates how

quickly misinformation can deceive its consumers, thereby motivating reckless, violent, or otherwise counter-normative behavior. Welch was not alone in being misled by the Pizzagate allegations. Believers in Pizzagate, and in the QAnon conspiracy theory that grew out of it, have harassed restaurant employees (Rosenberg, 2016), committed murders (Vigdor, 2021; Watkins, 2019), and stormed the US Capitol (Rubin et al., 2021) in the name of these theories. Unsurprisingly, it is the tendency of misinformation to cause false beliefs and thus reckless action that has received the lion's share of attention. However, the threat of misinformation cannot be understood solely in terms of causing false beliefs.

I.3 The holdout

In 1987, the *Sydney Morning Herald* reported the death of Norio Suzuki as follows:

> Yeti Hunter Dies.

Suzuki had indeed died in an avalanche while searching for a Yeti in the Himalayan mountains. This search for the mythical creature might appear less quixotic when one considers that Suzuki had previously managed to find a target nearly as elusive as the Yeti: Second Lieutenant Hiroo Onoda.

Onoda was an intelligence officer for the Japanese Imperial army during the Second World War. As part of his mission, Onoda was sent to Lubang Island, located in the Philippines, in 1944. Suzuki found him on this same island, roughly thirty years later. To that point, Onoda remained committed to his original mission, unaware that the war had long since ended. Only after the Japanese government sent Onoda's former commanding officer to officially relieve him of duty did Onoda accept that the war was over.

Prior to Suzuki's discovery, many unsuccessful attempts had been made to convince Onoda of the war's end. These are detailed in Onoda's (1999) autobiography *No Surrender: My Thirty Years War* in a chapter entitled "Faked Messages." Signed photographs of Onoda's family did not convince him, nor did newspapers illustrating the post-war peace that had settled over Japan. In each case, Onoda suspected that the supposed proof was somehow faked. Onoda attributed his suspicions to his intelligence training in the Futamata class at the Nakano military intelligence school:

> I had been taught at Futamata always to be on the lookout for faked messages, and it did not seem to me that my attitude was overly cautious…I still remembered learning at Futamata about a fake message that had made it easier for Germany to overrun France in 1940.

Onoda's training taught him not to take anything at face value, and instead to consider how various communications might be moves in the game of intelligence and counterintelligence. Even seeing his own brother on the island with a search party was not enough to overcome Onoda's suspicions:

> A man was standing on the top of Six Hundred speaking earnestly into a microphone. I approached a point about a hundred and fifty yards away from him. I did not dare go nearer, because I would have made too good a target.
>
> I could not see the man's face, but he was built like my brother, and his voice was identical.
>
> "That's really something," I thought. "They've found a Nisei or a prisoner who looks at a distance like my brother, and he's learned to imitate my brother's voice perfectly."
>
> The man started to sing, "East wind blowing in the sky over the capital…" This was a well-known students' song at the Tokyo First High School, which my brother had attended, and I knew he liked it. It started out as a fine performance, and I listened with interest. But gradually the voice grew strained and higher, and at the end it was completely off tune.
>
> I laughed to myself. The impersonator had not been able to keep it up, and his own voice had come through in the end.

Only later did Onoda learn that his brother's voice had cracked due to overwhelming emotion.

Onoda's imperviousness to evidence of the war's end was costly. He describes this realization as follows:

> For thirty years on Lubang I had polished my rifle every day. For what? For thirty years I had thought I was doing something for my country, but now it looked as though I had just caused a lot of people a lot of trouble.

Onoda poignantly describes the source of his confusion and that of another soldier that was his companion for many years:

> Kozuka and I had developed so many fixed ideas that we were unable to understand anything that did not conform with them. If there was anything that did not fit in with them we interpreted it to mean what we wanted it to mean.

Onoda was primed for suspicion by a history of education in intelligence, a domain in which fakery is to be expected. A formal education in intelligence is atypical of the general population. However, widespread awareness of the existence of fake news and manipulated media can likewise prime ordinary people for suspicion, especially of information that conflicts with their existing beliefs. Similarly, while most of us need not worry about fake brothers, we might worry about fake online persons in the form of bots and trolls. Such awareness might arise through direct experience. Noticing fake news in one's social media feed, for example, one might come to fear that there is more fake news present that one has not yet recognized as such. Consequently, one might come to doubt the reliability of even legitimate news.

Especially since the political events of 2016—most notably Brexit and the election to President of Donald Trump—the popular press has been heavy with stories about fake news, bots, Russian trolls, deepfakes, and other forms and disseminators of misinformation. Such stories no doubt inspire some to be more vigilant in their information consumption and thereby help to avoid deception. However, such vigilance can easily slip into excessive skepticism. Focusing on the possibility of deception might lead audiences to make similar mistakes to Onoda's, that is, to reject even accurate information. In this way, even accurate, well-intentioned investigations of fake news and related phenomena can contribute to skepticism toward legitimate sources of information[1]. It is thus essential not to regard the threat of misinformation solely in terms of its deceptive potential. It is likewise essential not to exaggerate the deceptive potential of misinformation, thereby fueling excessive skepticism.

I.4 The counterfeiters

In September of 1939, the Nazi government began to devise a plot to overcome one of its most formidable foes: The British economy. The twentieth century had already seen the industrialization of warfare manifest in the mass production of munitions. Arthur Nebe—head of Germany's criminal police—proposed an innovative application of German industrial capacity. The plan was to counterfeit and distribute British banknotes on a massive scale, thereby thrusting devastating inflation onto the British economy. Although the plan was never fully implemented, the attempt to mass produce convincing British banknotes was ultimately successful. The plotters later turned their attention to the forging of American banknotes.

Several of those involved expressed concerns about the plot, either privately or to the other plotters. Two of the most serious objections concerned the potential for it to backfire. In his diary, Joseph Goebbels wondered what

would happen if the British used the same tactic in retaliation (Malkin, 2006, Chapter 1). Others worried that discovery of the plan would leave the Germans with reputations as counterfeiters, thus doing long-term damage to the value of German currency.

At the heart of the Nazi plot, as well as the potential objections to it, is the recognition that the proliferation of fake currency degrades the value of real currency. Something similar is true of evidence. Ordinarily, news reports provide good evidence in favor of the claims reported. Ordinarily, photos and videos provide good evidence that the events they depict really occurred. But, where fake news, fake photos, and fake videos abound, even their authentic counterparts lose much of their evidential value. Misinformation, in effect, is counterfeit information (Fallis & Mathiesen, 2019) and tends to have an effect on legitimate evidence that is analogous to the effect of counterfeit currency on authentic currency.

Suppose that one forms a true belief that the President of the United States committed a verbal gaffe based on video footage that seems to show this. Suppose, further, that the video footage is authentic and has not been edited in any misleading way. Even in this case, one arguably does not *know* that the President committed a verbal gaffe. After all, it is plausible enough that the President's political foes might have concocted fake video footage or edited authentic footage to give a misleading impression. Even this possibility—increasingly realistic in light of novel techniques for cheaply and quickly editing video footage—is arguably enough to prevent one from having knowledge. This is because a given observation serves as good evidence for a given hypothesis only to the extent that the observation would be relatively unlikely if that hypothesis were false. For this reason, the realistic possibility that there is inauthentic evidence reduces the value of even authentic evidence in a way that undermines knowledge acquisition.

That misinformation reduces the significance of evidence is not a novel point. It is a familiar lesson that crying wolf in the absence of wolves reduces the evidential value of one's cries. It is not merely that falsely crying wolf encourages skepticism of even true reports in the future, such skepticism is *rational* in light of the reduced evidential value of one's cries. But counterfeits—whether informational or monetary—do not simply reduce the value of the authentic counterparts issued by the same authors. Fake British banknotes would reduce the value of even legitimately issued banknotes. Likewise, if they cannot be distinguished, fake news issued by unscrupulous parties can degrade the evidential value of news reports issued by scrupulous parties.

Suppose that the Nazi plot had been carried to fruition and that massive quantities of realistic counterfeit British currency had been dropped over British cities. It would then be reasonable for British citizens to apply extra

scrutiny to payments before accepting them and perhaps to avoid accepting them altogether. Likewise, it is often reasonable for those aware of the possibility of misinformation to apply extra scrutiny to information retrieved online and perhaps to avoid forming beliefs based on this information. Misinformation gives rise to skepticism not because it causes people to be irrational, but because it makes skepticism rational.

I.5 Content moderation and its discontents

Governments take counterfeiting very seriously. Those that engage in counterfeiting and indeed those that merely pass on counterfeit banknotes are typically subject to serious penalty, and counterfeit banknotes themselves are typically destroyed. Such heavy-handed measures are not the only means by which the destructive impacts of counterfeit currencies might be mitigated. In principle, ordinary citizens might be trained in the detection of counterfeits, or equipped with devices for distinguishing between real and fake currency. However, such alternatives would ask much of ordinary people and would slow down many financial transactions. In contrast, the strategy of deterrence and destruction functions to protect the integrity of currency quite generally, without demanding excessive vigilance on the part of ordinary persons.

Given the analogy between counterfeit currency and misinformation, it is at least initially plausible that a comparably heavy-handed approach would be effective and appropriate for mitigating the threats posed by misinformation. If it is appropriate to destroy counterfeit banknotes to protect the value of currency, then it is likewise appropriate to protect knowledge by destroying misinformation—or at least by removing misinformation from the social media platforms on which it proliferates. In this way, content moderation—in the form of interference with who and what can appear on social media and in what context—appears to be an appropriate response to the challenges posed by misinformation. This is not to say that more gentle responses, involving for example better education in information literacy, are unnecessary or undesirable. It is instead to say that content moderation is at least one valuable weapon among the arsenal of those that can be deployed to combat misinformation's ill-effects.

Yet the use of content moderation against misinformation has proven highly controversial. Some objections to content moderation appeal to the inalienable right to free expression. Other objections are more specifically epistemological, alleging for instance that content moderation is, or could be, used to suppress the truth. It is such epistemological issues concerning the threats of misinformation and the efficacy of content moderation in mitigating these threats with which this book is concerned.

I.6 Book overview

Here is the plan for the book. In Part I, I develop an account of how misinformation threatens knowledge, and how the unique social feedback encountered on social media exacerbates this threat. This begins, in Chapter 1, with a distinction between three threats posed by misinformation: the *deceptive threat*, the *skeptical threat*, and the *epistemic threat*. I then argue that misinformation's threat to knowledge does not, strictly speaking, depend on the real existence of misinformation. Even the mere propensity of misinformation to exist and indeed mere concerns as to the possible existence of misinformation are enough to threaten knowledge. Chapter 1 concludes with a discussion of some limitations on misinformation's threats to knowledge.

In Chapter 2, I discuss the skeptical threat of misinformation at greater length, with a special focus on the role of conspiracy theories in driving skepticism of official explanations and evidence presented by epistemic authorities more generally. I argue that the ill-effects of conspiracy theories do not depend on the irrationality of individual believers, but rather that conspiracy theories can make rational people believe absurd things and, often as importantly, fail to believe well-evidenced claims.

Chapter 3 concerns the role of social evidence in the discrimination between accurate and inaccurate information. Such evidence takes many forms on social media, including testimony, likes, and shares. I argue that, while social evidence can in principle help to distinguish between accurate and inaccurate information, the significance of this evidence is compromised by several factors. These include the ambiguity of certain forms of social evidence, the varied motivations that individuals have for engaging in social media communication, and the distorting influence of trolls and bots. Finally, I argue that epistemic virtue on the part of individuals is not enough to secure the value of social evidence.

In Part II, I turn to the effectiveness of content moderation as a response to the challenges of misinformation and to positive proposals as to how content moderation can best be put to this end. This begins in Chapter 4, where I develop several epistemic arguments against the labeling and removal of misinformation from social media. These arguments highlight the likelihood that content moderation will fail to be comprehensive and the lack of trust that many individuals have toward content moderators. I argue that these factors not only limit the effectiveness of content moderation as a way of mitigating the threats of misinformation, but that, in light of such factors, content moderation can exacerbate the threat to knowledge. Finally, I argue that content moderation threatens to reduce the evidential value of testimony and apparent consensus insofar as content moderation amounts to tampering with the social evidence.

Then, in Chapter 5, I argue that there is good reason to engage in content moderation, despite the concerns raised in the previous chapter. In part, this is because social media platforms *must* exert control over the spread of content, and thus—even absent explicit attempts to mitigate the spread of misinformation—some of the epistemic ill-effects of content moderation are unavoidable. Furthermore, content moderation can, at least in many cases, mitigate the three-pronged threat of misinformation. Additionally, individuals are subject to a range of limitations that promote the spread and consumption of misinformation and that can be reduced through content moderation. Finally, the concern that content moderation amounts to social evidence tampering is of limited importance, as individuals already struggle to assess the weight of such evidence.

Chapter 6 offers several proposals as to how content moderation can be conducted to combat the three-pronged threat of misinformation while minimizing the epistemic ill-effects of content moderation itself. Generally, I argue that content moderation ought to aim at enhancing the epistemic agency of ordinary individuals as both consumers and moderators of content. This can be accomplished, in part, by enhancing the quality of social evidence and by assisting individuals in assessing the track records of various sources.

In the epilogue, I emphasize that misinformation is not simply an epistemological problem. Very often, the consumption and sharing of misinformation reflects underlying individual, social, and political problems. Such problems must be confronted but, as I argue, protecting knowledge is an essential part of doing so.

I.7 Epistemology, social media, and content moderation

Social media misinformation and content moderation may strike some readers as strange topics to address through a philosophical lens. Often, philosophy is associated with perennial issues like the meaning of life and the nature of fundamental concepts like knowledge and justice, rather than issues introduced by new technologies. Yet, if philosophy is to be relevant, it must engage with changing conditions in the world, including novel technologies. Moreover, both misinformation and content moderation are natural topics for epistemology. Much of epistemology is devoted to exploring ways in which appearances might systematically diverge from reality and the consequences of such possibilities for human knowledge. Thus, for example, the ancient skeptics considered how the peculiarities of our sense organs and the conditions of our bodies might distort our perceptions of the world. The ancient Chinese philosopher Zhuangzi and later the early modern philosopher René Descartes considered the possibility that our experiences of the world were mere dreams or perhaps due to the

machinations of a powerful and deceptive demon. Twentieth-century philosophers modernized such anxieties, considering the possibility that our experiences might be manufactured by deceptive neuroscientists. More recently, epistemologists have devised elaborate thought experiments to consider how fakes in one's environment might compromise the acquisition of knowledge. Epistemological consideration of social media misinformation is thus not a radical departure from the discipline's history. Instead, it amounts to the application of existing conceptual tools to situations encountered in the real world.

Content moderation is likewise a suitable object of epistemological theorizing, albeit one that has thus far received limited attention from epistemologists[2]. The epistemology of content moderation is not wholly uncharted territory, as it is partly an extension of long-running discussions of the epistemology of free speech and censorship. Philosophers have argued that free speech carries important epistemic benefits and that the restriction of speech does great damage to the individual and collective pursuit of knowledge (Wright, 2021). Social media offers ordinary individuals radically expanded abilities to broadcast their ideas, while content moderation potentially restricts such abilities. In short, while social media creates new opportunities for freedom of expression, it also affords new means of control. Exploring these issues through an epistemological lens serves to bring theoretical tools to bear on issues of practical significance while also promising to enrich epistemology itself.

All this said, social media and content moderation present unique challenges for epistemological study, insofar as certain relevant facts—concerning the existence, popularity, and policies of various platforms—are subject to rapid change[3]. Thus, although I refer to specific platforms and policies where appropriate, this book does not attempt to provide an in-depth study of the specificities of content moderation policies on various platforms. Rather, I aim to assess, through an epistemological lens, various types of content moderation policies that have been used and that might be used in the future.

Notes

1 Relatedly, there is evidence that warnings about misinformation (Van Der Meer et al., 2023) as well as attempts to train individuals to better identify misinformation (Modirrousta-Galian & Higham, 2023) reduce credulity toward misinformation at the cost of also reducing credulity toward legitimate information.
2 But see Karen Frost-Arnold (2023) for a rare exception.
3 For example, the name of Twitter was recently changed to "X." Because relevant theoretical and empirical studies referenced throughout this book were carried out when the platform was called Twitter, I continue to use that name to refer to the platform.

References

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

Carlson, M. (2020). Fake News as an Informational Moral Panic: The Symbolic Deviancy of Social Media during the 2016 US Presidential Election. *Information, Communication & Society*, *23*(3), 374–388. https://doi.org/10.1080/1369118X.2018.1505934

Fallis, D., & Mathiesen, K. (2019). Fake News Is Counterfeit News. *Inquiry*, 1–20. https://doi.org/10.1080/0020174X.2019.1688179

Frost-Arnold, K. (2023). *Who Should We Be Online? A Social Epistemology for the Internet*. Oxford University Press.

Fuchs, C. (2014). *Social Media: A Critical Introduction*. SAGE.

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

Jungherr, A., & Schroeder, R. (2021). Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy. *Social Media + Society*, *7*(1). https://doi.org/10.1177/2056305121988928

Kaplan, A. M., & Haenlein, M. (2010). Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Malkin, L. (2006). *Krueger's Men: The Secret Nazi Counterfeit Plot and the Prisoners of Block 19* (1st ed.). Little, Brown and Co.

Modirrousta-Galian, A., & Higham, P. A. (2023). Gamified Inoculation Interventions Do Not Improve Discrimination between True and Fake News: Reanalyzing Existing Research with Receiver Operating Characteristic Analysis. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001395

Onoda, H. (1999). *No Surrender: My Thirty-Year War*. Naval Institute Press.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. *American Political Science Review*, *115*(3), 999–1015. https://doi.org/10.1017/S0003055421000290

Rosenberg, E. (2016, December 7). Roberta's, Popular Brooklyn Restaurant, Is Pulled Into 'Pizzagate' Hoax. *The New York Times*. https://www.nytimes.com/2016/12/07/nyregion/robertas-restaurant-brooklyn-threatened-fake-news-pizzagate-conspiracy.html

Rubin, O., Bruggeman, L., & Steakin, W. (2021). *QAnon Emerges as Recurring Theme of Criminal Cases Tied to US Capitol Siege*. ABC News. https://abcnews.go.com/US/qanon-emerges-recurring-theme-criminal-cases-tied-us/story?id=75347445

van der Linden, S. (2023). *Foolproof: Why Misinformation Infects Our Minds and How to Build Immunity*. W.W. Norton & Company.

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

Vigdor, N. (2021, August 12). Surf Instructor Killed His Children and Claimed QAnon Made Him Do It, F.B.I. Says. *The New York Times*. https://www.nytimes.com/2021/08/12/us/matthew-coleman-children-mexico.html

Watkins, A. (2019, December 6). Accused of Killing a Gambino Mob Boss, He's Presenting a Novel Defense. *The New York Times*. https://www.nytimes.com/2019/12/06/nyregion/gambino-shooting-anthony-comello-qanon.html

Wright, A. T. (2021). Mill's Social Epistemic Rationale for the Freedom to Dispute Scientific Knowledge: Why We Must Put Up with Flat-Earthers. *Philosopher's Imprint*, *21*(14). http://hdl.handle.net/2027/spo.3521354.0021.014

Introduction

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

Carlson, M. (2020). Fake News as an Informational Moral Panic: The Symbolic Deviancy of Social Media during the 2016 US Presidential Election. Information, Communication & Society, 23(3), 374–388. https://doi.org/10.1080/1369118X.2018.1505934

Fallis, D. , & Mathiesen, K. (2019). Fake News Is Counterfeit News. Inquiry, 1–20. https://doi.org/10.1080/0020174X.2019.1688179

Frost-Arnold, K. (2023). Who Should We Be Online? A Social Epistemology for the Internet. Oxford University Press.

Fuchs, C. (2014). Social Media: A Critical Introduction. SAGE.

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

Jungherr, A. , & Schroeder, R. (2021). Disinformation and the Structural Transformations of the Public Arena: Addressing the Actual Challenges to Democracy. Social Media + Society, 7(1). https://doi.org/10.1177/2056305121988928

Kaplan, A. M. , & Haenlein, M. (2010). Users of the World, Unite! The Challenges and Opportunities of Social Media. Business Horizons, 53(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Malkin, L. (2006). Krueger's Men: The Secret Nazi Counterfeit Plot and the Prisoners of Block 19 (1st ed.). Little, Brown and Co.

Modirrousta-Galian, A. , & Higham, P. A. (2023). Gamified Inoculation Interventions Do Not Improve Discrimination between True and Fake News: Reanalyzing Existing Research with Receiver Operating Characteristic Analysis. Journal of Experimental Psychology: General. https://doi.org/10.1037/xge0001395

Onoda, H. (1999). No Surrender: My Thirty-Year War. Naval Institute Press.

Osmundsen, M. , Bor, A. , Vahlstrup, P. B. , Bechmann, A. , & Petersen, M. B. (2021). Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. American Political Science Review, 115(3), 999–1015. https://doi.org/10.1017/S0003055421000290

Rosenberg, E. (2016, December 7). Roberta's, Popular Brooklyn Restaurant, Is Pulled Into 'Pizzagate' Hoax. The New York Times. https://www.nytimes.com/2016/12/07/nyregion/robertas-restaurant-brooklyn-threatened-fake-news-pizzagate-conspiracy.html

Rubin, O. , Bruggeman, L. , & Steakin, W. (2021). QAnon Emerges as Recurring Theme of Criminal Cases Tied to US Capitol Siege. ABC News. https://abcnews.go.com/US/qanon-emerges-recurring-theme-criminal-cases-tied-us/story?id=75347445

van der Linden, S. (2023). Foolproof: Why Misinformation Infects Our Minds and How to Build Immunity. W.W. Norton & Company.

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

Vigdor, N. (2021, August 12). Surf Instructor Killed His Children and Claimed QAnon Made Him Do It, F.B.I. Says. The New York Times. https://www.nytimes.com/2021/08/12/us/matthew-coleman-children-mexico.html

Watkins, A. (2019, December 6). Accused of Killing a Gambino Mob Boss, He's Presenting a Novel Defense. The New York Times. https://www.nytimes.com/2019/12/06/nyregion/gambino-shooting-anthony-comello-qanon.html

Wright, A. T. (2021). Mill's Social Epistemic Rationale for the Freedom to Dispute Scientific Knowledge: Why We Must Put Up with Flat-Earthers. Philosopher's Imprint, 21(14). http://hdl.handle.net/2027/spo.3521354.0021.014

How misinformation prevents knowing

Alston, W. P. (1988). The Deontological Conception of Epistemic Justification. Philosophical Perspectives, 2, 257–299. https://doi.org/10.2307/2214077

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

Anvari, F. , & Lakens, D. (2018). The Replicability Crisis and Public Trust in Psychological Science. Comprehensive Results in Social Psychology, 3(3), 266–286. https://doi.org/10.1080/23743603.2019.1684822

Bernecker, S. , Flowerree, A. K. , & Grundmann, T. (Eds.). (2021). The Epistemology of Fake News (1st ed.). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198863977.001.0001

Blake-Turner, C. (2020). Fake News, Relevant Alternatives, and the Degradation of Our Epistemic Environment. Inquiry, 1–21. https://doi.org/10.1080/0020174X.2020.1725623

Brenan, M. (2022, October 18). Americans' Trust In Media Remains Near Record Low. Gallup.Com. https://news.gallup.com/poll/403166/americans-trust-media-remains-near-record-low.aspx

Carlson, M. (2021). Skepticism and the Digital Information Environment. SATS, 22(2), 149–167. https://doi.org/10.1515/sats-2021-0008

Cavedon-Taylor, D. (2013). Photographically Based Knowledge. Episteme, 10(3), 283–297. https://doi.org/10.1017/epi.2013.21

Cohen, L. J. (1989). Belief and Acceptance. Mind, 98(391), 367–389. https://doi.org/10.1093/mind/xcviii.391.367

Colledani, D. , Anselmi, P. , & Robusto, E. (2021). COVID-19 Emergency: The Influence of Implicit Attitudes, Information Sources, and Individual Characteristics on Psychological Distress, Intentions to Get Vaccinated, and Compliance with Restrictive Rules. Health Psychology Report, 10(1), 1–12. https://doi.org/10.5114/hpr.2021.111292

Cox, J. (2019, October 7). Most Deepfakes Are Used for Creating Non-Consensual Porn, Not Fake News. Vice. https://www.vice.com/en/article/7x57v9/most-deepfakes-are-porn-harassment-not-fake-news

de Ridder, J. (2022). Online Illusions of Understanding. Social Epistemology, 1–16. https://doi.org/10.1080/02691728.2022.2151331

Fallis, D. (2021). The Epistemic Threat of Deepfakes. Philosophy & Technology, 34(4), 623–643. https://doi.org/10.1007/s13347-020-00419-2

Fallis, D. , & Mathiesen, K. (2019). Fake News Is Counterfeit News. Inquiry, 1–20. https://doi.org/10.1080/0020174X.2019.1688179

Foer, F. (2018, April 8). The Era of Fake Video Begins. The Atlantic. https://www.theatlantic.com/magazine/archive/2018/05/realitys-end/556877/

Gelfert, A. (2018). Fake News: A Definition. Informal Logic, 38(1), 84–117. https://doi.org/10.22329/il.v38i1.5068

Gendler, T. S. (2008). Alief and Belief. Journal of Philosophy, 105(10), 634–663. https://doi.org/10.5840/jphil20081051025

Genot, E. J. , & Olsson, E. J. (2021). The Dissemination of Scientific Fake News. In S. Bernecker , A. K. Flowerree , & T. Grundmann (Eds.), The Epistemology of Fake News (pp. 228–242). Oxford University Press.

Gettier, E. L. (1963). Is Justified True Belief Knowledge? Analysis, 23(6), 121–123. https://doi.org/10.1093/analys/23.6.121

Goldberg, J. (2020, November 16). Why Obama Fears for Our Democracy. The Atlantic. https://www.theatlantic.com/ideas/archive/2020/11/why-obama-fears-for-our-democracy/617087/

Goldman, A. I. (1976). Discrimination and Perceptual Knowledge. The Journal of Philosophy, 73(20), 771. https://doi.org/10.2307/2025679

Grundmann, T. (2020). Fake News: The Case for a Purely Consumer-Oriented Explication. Inquiry, 1–15. https://doi.org/10.1080/0020174X.2020.1813195

Habgood-Coote, J. (2019). Stop Talking About Fake News! Inquiry, 62(9–10), 1033–1065. https://doi.org/10.1080/0020174X.2018.1508363

Habgood-Coote, J. (2023). Deepfakes and the Epistemic Apocalypse. Synthese, 201(3), 103. https://doi.org/10.1007/s11229-023-04097-3
Harris, K. R. (2021). Video On Demand: What Deepfakes Do and How They Harm. Synthese, 199(5–6), 13373–13391. https://doi.org/10.1007/s11229-021-03379-y
Harris, K. R. (2022). Real Fakes: The Epistemology of Online Misinformation. Philosophy & Technology, 35(3), 83. https://doi.org/10.1007/s13347-022-00581-9
Hughes, H. C. , & Waismel-Manor, I. (2021). The Macedonian Fake News Industry and the 2016 US Election. PS: Political Science & Politics, 54(1), 19–23. https://doi.org/10.1017/S1049096520000992
Jaster, R. , & Lanius, D. (2018). What Is Fake News? Versus, 2(127), 207–227.
Kerner, C. , & Risse, M. (2021). Beyond Porn and Discreditation: Epistemic Promises and Perils of Deepfake Technology in Digital Lifeworlds. Moral Philosophy and Politics, 8(1), 81–108. https://doi.org/10.1515/mopp-2020-0024
Kyaw, N. N. (2020). Social Media, Hate Speech and Fake News during Myanmar's Political Transition. In A. Sinpeng & R. Tapsell (Eds.), From Grassroots Activism to Disinformation (pp. 86–104). ISEAS–Yusof Ishak Institute Singapore. https://doi.org/10.1355/9789814951036-006
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, 359(6380), 1094–1096. https://doi.org/10.1126/science.aao2998
Levy, N. (2017, July 24). The Bad News About Fake News, Neil Levy. Social Epistemology Review and Reply Collective. https://social-epistemology.com/2017/07/24/the-bad-news-about-fake-news-neil-levy/
Levy, N. (2022). Bad beliefs: Why they happen to good people (1sted.). Oxford University Press.
Mandelbaum, E. (2013). Against Alief. Philosophical Studies, 165(1), 197–211. https://doi.org/10.1007/s11098-012-9930-7
Matthews, T. (2023). Deepfakes, Fake Barns, and Knowledge from Videos. Synthese, 201(2), 41. https://doi.org/10.1007/s11229-022-04033-x
Mercier, H. (2020). Not Born Yesterday: The Science of Who We Trust and What We Believe. Princeton University Press. https://doi.org/10.1515/9780691198842
Michaelson, E. , Sterken, R. , & Pepp, J. (2019). What's New About Fake News? Journal of Ethics and Social Philosophy, 16(2). https://doi.org/10.26556/jesp.v16i2.629
Novaes, C. D. , & De Ridder, J. (2021). Is Fake News Old News? In S. Bernecker , A. K. Flowerree , & T. Grundmann (Eds.), The Epistemology of Fake News (1st ed., pp. 156–179). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198863977.003.0008
Rini, R. (2017). Fake News and Partisan Epistemology. Kennedy Institute of Ethics Journal, 27(S2), 43–64.
Rini, R. (2020). Deepfakes and the Epistemic Backstop. Philosophers' Imprint, 20(24), 1–16.
Russell, B. (1948). Human Knowledge: Its Scope and Limits. London and New York: Routledge.
Silverman, C. (2016, November 16). This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook. BuzzFeed News. https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook
Skyrms, B. (2010). Signals: Evolution, Learning, and Information (1st ed.). Oxford University PressOxford. https://doi.org/10.1093/acprof:oso/9780199580828.001.0001
Stroud, B. (1984). The Significance of Philosophical Scepticism. Oxford University Press.
Tucher, A. (2017). "I Believe in Faking": The Dilemma of Photographic Realism at the Dawn of Photojournalism. Photography and Culture, 10(3), 195–214. https://doi.org/10.1080/17514517.2017.1322397
Wallace, C. (2016, September 2). Obama Did Not Ban the Pledge. FactCheck.Org. https://www.factcheck.org/2016/09/obama-did-not-ban-the-pledge/
Williams, B. (1973). Problems of the Self: Philosophical Papers 1956–1972 (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511621253

Wingen, T. , Berkessel, J. B. , & Englich, B. (2020). No Replication, No Trust? How Low Replicability Influences Trust in Psychology. Social Psychological and Personality Science, 11(4), 454–463. https://doi.org/10.1177/1948550619877412

Conspiracy theories and runaway skepticism

Bakshy, E. , Messing, S. , & Adamic, L. A. (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. Science, 348(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Barkun, M. (2013). A Culture of Conspiracy: Apocalyptic Visions in Contemporary America (2nd ed.). University of California Press.

Basham, L. (2003). Malevolent Global Conspiracy. Journal of Social Philosophy, 34(1), 91–103. https://doi.org/10.1111/1467-9833.00167

Beam, M. A. , Hutchens, M. J. , & Hmielowski, J. D. (2018). Facebook News and (De)polarization: Reinforcing Spirals in the 2016 US Election. Information, Communication & Society, 21(7), 940–958. https://doi.org/10.1080/1369118X.2018.1444783

Boudry, M. (2022). Why We Should Be Suspicious of Conspiracy Theories: A Novel Demarcation Problem. Episteme, 1–21. https://doi.org/10.1017/epi.2022.34

Cassam, Q. (2019). Conspiracy Theories. Polity Press.

Cassam, Q. (2023). Conspiracy Theories. Society, 60(2), 190–199. https://doi.org/10.1007/s12115-023-00816-1

Coady, D. (2003). Conspiracy Theories and Official Stories: International Journal of Applied Philosophy, 17(2), 197–209. https://doi.org/10.5840/ijap200317210

Dellsén, F. (2021). Consensus versus Unanimity: Which Carries More Weight? The British Journal for the Philosophy of Science, 718273. https://doi.org/10.1086/718273

Dentith, M. R. X. (2014). The Philosophy of Conspiracy Theories. Palgrave Macmillan UK. https://doi.org/10.1057/9781137363169

Dentith, M. R. X. (2018). Expertise and Conspiracy Theories. Social Epistemology, 32(3), 196–208. https://doi.org/10.1080/02691728.2018.1440021

Dentith, M. R. X. (2019). Conspiracy Theories on the Basis of the Evidence. Synthese, 196(6), 2243–2261. https://doi.org/10.1007/s11229-017-1532-7

Dentith, M. R. X. (2022). Suspicious Conspiracy Theories. Synthese, 200(3), 243. https://doi.org/10.1007/s11229-022-03602-4

Dentith, M. R. X. (2023). Some Conspiracy Theories. Social Epistemology, 37(4), 522–534. https://doi.org/10.1080/02691728.2023.2173539

Douglas, K. M. , Sutton, R. M. , & Cichocka, A. (2017). The Psychology of Conspiracy Theories. Current Directions in Psychological Science, 26(6), 538–542. https://doi.org/10.1177/0963721417718261

Feldman, S. (2011). Counterfact Conspiracy Theories. International Journal of Applied Philosophy, 25(1), 15–24. https://doi.org/10.5840/ijap20112512

Gardiner, G. (2021). Relevance and Risk: How the Relevant Alternatives Framework Models the Epistemology of Risk. Synthese, 199(1–2), 481–511. https://doi.org/10.1007/s11229-020-02668-2

Goertzel, T. (1994). Belief in Conspiracy Theories. Political Psychology, 15(4), 731. https://doi.org/10.2307/3791630

Goldman, A. (2016, December 7). The Comet Ping Pong Gunman Answers Our Reporter's Questions. The New York Times. https://www.nytimes.com/2016/12/07/us/edgar-welch-comet-pizza-fake-news.html

Hagen, K. (2022). Is Conspiracy Theorizing *Really* Epistemically Problematic? Episteme, 19(2), 197–219. https://doi.org/10.1017/epi.2020.19

Hardwig, J. (1985). Epistemic Dependence. The Journal of Philosophy, 82(7), 335. https://doi.org/10.2307/2026523

Harris, K. R. (2018). What's Epistemically Wrong with Conspiracy Theorising? Royal Institute of Philosophy Supplement, 84, 235–257.

Harris, K. R. (2022). Some Problems with Particularism. Synthese, 200(6), 447. https://doi.org/10.1007/s11229-022-03948-9
Harris, K. R. (2023). Conspiracy Theories, Populism, and Epistemic Autonomy. Journal of the American Philosophical Association, 9(1), 21–36.
Ichino, A. , & Räikkä, J. (2020). Non-doxastic conspiracy theories. Argumenta. https://doi.org/10.14275/2465-2334/20200.ich
Keeley, B. L. (1999). Of Conspiracy Theories. The Journal of Philosophy, 96(3), 109. https://doi.org/10.2307/2564659
Lamoureux, M. (2016). This Dude Accidentally Convinced the Internet that Finland Doesn't Exist. Vice. https://www.vice.com/en/article/xyd48w/this-dude-accidentally-convinced-the-internet-that-finland-doesnt-exist
Levy, N. (2007). Radically Socialized Knowledge and Conspiracy Theories. Episteme, 4(2), 181–192. https://doi.org/10.3366/epi.2007.4.2.181
Lu, Y. , & Lee, J. K. (2019). Stumbling Upon the Other Side: Incidental Learning of Counter-Attitudinal Political Information on Facebook. New Media & Society, 21(1), 248–265. https://doi.org/10.1177/1461444818793421
McMurry, E. , Kapetaneas, J. , Park, C. , McNiff, E. , & Chang, J. (2020). QAnon, once a Fringe Conspiracy Theory, Edges into the Mainstream: "Things Could Get Much, Much Worse." ABC News. https://abcnews.go.com/Politics/qanon-fringe-conspiracy-theory-edges-mainstream-things-worse/story?id=72751829
Miller, M. E. (2021). The Pizzagate Gunman Is Out of Prison. Conspiracy Theories Are Out of Control. Washington Post. https://www.washingtonpost.com/dc-md-va/2021/02/16/pizzagate-qanon-capitol-attack/
Min, S. J. , & Wohn, D. Y. (2020). Underneath the Filter Bubble: The Role of Weak Ties and Network Cultural Diversity in Cross-Cutting Exposure to Disagreements on Social Media. The Journal of Social Media in Society, 9(1), Article 1.
Napolitano, M. G. (2021). Conspiracy Theories and Evidential Self-Insulation. In S. Bernecker , A. K. Flowerree , & T. Grundmann (Eds.), The Epistemology of Fake News (1st ed., pp. 82–106). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198863977.003.0005
Napolitano, M. G. , & Reuter, K. (2023). What is a Conspiracy Theory? Erkenntnis, 88(5), 2035–2062. https://doi.org/10.1007/s10670-021-00441-6
Negroponte, N. (1996). Being Digital (1. Vintage Books ed). Vintage Books.
Nguyen, C. T. (2020). Echo Chambers and Epistemic Bubbles. Episteme, 17(2), 141–161. https://doi.org/10.1017/epi.2018.32
Nguyen, C. T. (2022). Playfulness Versus Epistemic Traps. In M. Alfano , C. Klein , & J. D. Ridder , Social Virtue Epistemology (1st ed., pp. 269–290). Routledge. https://doi.org/10.4324/9780367808952-36
Oreskes, N. , & Conway, E. M. (2010). Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming (1st U.S. ed). Bloomsbury Press.
Pariser, E. (2011). The Filter Bubble: What the Internet Is Hiding From You. Penguin Press.
Pigden, C. (1995). Popper Revisited, or What Is Wrong With Conspiracy Theories? Philosophy of the Social Sciences, 25(1), 3–34. https://doi.org/10.1177/004839319502500101
Pigden, C. (2006). Complots of Mischief. In D. Coady (Ed.), Conspiracy Theories: The Philosophical Debate (pp. 139–166). Ashgate. https://philarchive.org/rec/PIGCOM
Pigden, C. (2007). Conspiracy Theories and the Conventional Wisdom. Episteme, 4(2), 219–232. https://doi.org/10.3366/epi.2007.4.2.219
Pomerantsev, P. (2014, September 9). How Vladimir Putin Is Revolutionizing Information Warfare. The Atlantic. https://www.theatlantic.com/international/archive/2014/09/russia-putin-revolutionizing-information-warfare/379880/
Poth, N. , & Dolega, K. (2023). Bayesian Belief Protection: A Study of Belief in Conspiracy Theories. Philosophical Psychology, 36(6), 1182–1207. https://doi.org/10.1080/09515089.2023.2168881
Räikkä, J. (2023, May 30). *Why a Pejorative Definition of "Conspiracy Theory" Need Not Be Unfair*. Social Epistemology Review and Reply Collective. https://social-

epistemology.com/2023/05/30/why-a-pejorative-definition-of-conspiracy-theory-need-not-be-unfair-juha-raikka/
Rini, R. (2021). Weaponized Skepticism: An Analysis of Social Media Deception as Applied Political Epistemology. In E. Edenburg & M. Hannon (Eds.), Political Epistemology (pp. 31–48). Oxford: Oxford University Press.
Sunstein, C. R. (2001). Republic.com. Princeton University Press.
Sunstein, C. R. , & Vermeule, A. (2009). Conspiracy Theories: Causes and Cures*. Journal of Political Philosophy, 17(2), 202–227. https://doi.org/10.1111/j.1467-9760.2008.00325.x
Van Alstyne, M. , & Brynjolfsson, E. (2005). Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities. Management Science, 51(6), 851–868.
van der Wal, R. C. , Sutton, R. M. , Lange, J. , & Braga, J. P. N. (2018). Suspicious Binds: Conspiracy Thinking and Tenuous Perceptions of Causal Connections between Co-Occurring and Spuriously Correlated Events. European Journal of Social Psychology, 48(7), 970–989. https://doi.org/10.1002/ejsp.2507
van Prooijen, J.-W. , Douglas, K. M. , & De Inocencio, C. (2018). Connecting the Dots: Illusory Pattern Perception Predicts Belief in Conspiracies and the Supernatural. European Journal of Social Psychology, 48(3), 320–335. https://doi.org/10.1002/ejsp.2331
Wood, M. J. , & Douglas, K. M. (2013). “What About Building 7?” A Social Psychological Study of Online Discussion of 9/11 Conspiracy Theories. Frontiers in Psychology, 4. https://doi.org/10.3389/fpsyg.2013.00409

Ambiguity, fakery, and social evidence

Adler, J. E. (1994). Testimony, Trust, Knowing. Journal of Philosophy, 91(5), 264–275.
Alexander, S. (2013, April 12). Lizardman’s Constant Is 4%. Slate Star Codex. https://slatestarcodex.com/2013/04/12/noisy-poll-results-and-reptilian-muslim-climatologists-from-mars/
Alsmadi, I. , & O'Brien, M. J. (2020). How Many Bots in Russian Troll Tweets? Information Processing & Management, 57(6), 102303. https://doi.org/10.1016/j.ipm.2020.102303
Arechar, A. A. , Allen, J. , Berinsky, A. J. , Cole, R. , Epstein, Z. , Garimella, K. , Gully, A. , Lu, J. G. , Ross, R. M. , Stagnaro, M. N. , Zhang, Y. , Pennycook, G. , & Rand, D. G. (2023). Understanding and Combatting Misinformation across 16 Countries on Six Continents. Nature Human Behaviour. https://doi.org/10.1038/s41562-023-01641-6
Avram, M. , Micallef, N. , Patil, S. , & Menczer, F. (2020). Exposure to Social Engagement Metrics Increases Vulnerability to Misinformation. Harvard Kennedy School Misinformation Review. https://doi.org/10.37016/mr-2020-033
Baehr, J. (2010). Epistemic Malevolence. Metaphilosophy, 41(1–2), 189–213. https://doi.org/10.1111/j.1467-9973.2009.01623.x
Baehr, J. (2011). The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology. Oxford University Press.
Bastos, M. , & Farkas, J. (2019). “Donald Trump Is My President!”: The Internet Research Agency Propaganda Machine. Social Media + Society, 5(3). https://doi.org/10.1177/2056305119865466
Bastos, M. , & Mercea, D. (2018). The Public Accountability of Social Platforms: Lessons from a Study on Bots and Trolls in the Brexit Campaign. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2128), 20180003. https://doi.org/10.1098/rsta.2018.0003
Bastos, M. T. , & Mercea, D. (2019). The Brexit Botnet and User-Generated Hyperpartisan News. Social Science Computer Review, 37(1), 38–54. https://doi.org/10.1177/0894439317734157
Battaly, H. (2008). Virtue Epistemology. Philosophy Compass, 3(4), 639–663. https://doi.org/10.1111/j.1747-9991.2008.00146.x
Broniatowski, D. A. , Jamison, A. M. , Qi, S. , AlKulaib, L. , Chen, T. , Benton, A. , Quinn, S. C. , & Dredze, M. (2018). Weaponized Health Communication: Twitter Bots and Russian

Trolls Amplify the Vaccine Debate. American Journal of Public Health, 108(10), 1378–1384. https://doi.org/10.2105/AJPH.2018.304567
Bullock, J. G. , Gerber, A. S. , Hill, S. J. , & Huber, G. A. (2015). Partisan Bias in Factual Beliefs about Politics. Quarterly Journal of Political Science, 10(4), 519–578. https://doi.org/10.1561/100.00014074
Cassam, Q. (2016). Vice Epistemology. The Monist, 99(2), 159–180.
Chen, A. (2015, June 2). The Agency. The New York Times. https://www.nytimes.com/2015/06/07/magazine/the-agency.html
Coady, C. A. J. (1992). Testimony: A Philosophical Study. Oxford University Press.
Fallis, D. (2021). The Epistemic Threat of Deepfakes. Philosophy & Technology, 34(4), 623–643. https://doi.org/10.1007/s13347-020-00419-2
Feldman, R. (1985). Reliability and Justification. The Monist, 68(2), 159–174. https://doi.org/10.5840/monist198568226
Ferrara, E. , Varol, O. , Davis, C. , Menczer, F. , & Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96–104. https://doi.org/10.1145/2818717
Fichman, P. , & Sanfilippo, M. R. (2016). Online Trolling and Its Perpetrators: Under the Cyberbridge. Rowman & Littlefield.
Fricker, E. (1994). Against Gullibility. In A. Chakrabarti & B. K. Matilal (Eds.), Knowing from Words (pp. 125–161). Kluwer Academic Publishers.
Ganapini, M. B. (2023). The Signaling Function of Sharing Fake Stories. Mind & Language, 38(1), 64–80. https://doi.org/10.1111/mila.12373
Goldman, A. I. (1979). What is Justified Belief? In G. S. Pappas (Ed.), Justification and Knowledge (pp. 1–23). Springer Netherlands. https://doi.org/10.1007/978-94-009-9493-5_1
Goldman, A. I. (1999). Knowledge in a Social World. Oxford University Press.
Greco, J. (2010). Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511844645
Hampton, R. (2019, April 23). The Black Feminists Who Saw the Alt-Right Threat Coming. Slate. https://slate.com/technology/2019/04/black-feminists-alt-right-twitter-gamergate.html
Hannon, M. , & de Ridder, J. (2021). The Point of Political Belief. In M. Hannon & J. de Ridder (Eds.), Routledge Handbook of Political Epistemology (pp. 156–166). Routledge.
Hardaker, C. (2010). Trolling in Asynchronous Computer-Mediated Communication: From User Discussions to Academic Definitions. Journal of Politeness Research. Language, Behaviour, Culture, 6(2). https://doi.org/10.1515/jplr.2010.011
Harris, K. R. (2018). What's Epistemically Wrong with Conspiracy Theorising? Royal Institute of Philosophy Supplement, 84, 235–257.
Harris, K. R. (2022). Outward-Facing Epistemic Vice. Synthese, 200(6), 516. https://doi.org/10.1007/s11229-022-03995-2
Harris, K. R. (2023). Liars and Trolls and Bots Online: The Problem of Fake Persons. Philosophy & Technology, 36(2), 35. https://doi.org/10.1007/s13347-023-00640-9
Hartman, R. (2021, April 20). Did 4% of Americans Really Drink Bleach Last Year? Harvard Business Review. https://hbr.org/2021/04/did-4-of-americans-really-drink-bleach-last-year
Hayes, R. A. , Carr, C. T. , & Wohn, D. Y. (2016). One Click, Many Meanings: Interpreting Paralinguistic Digital Affordances in Social Media. Journal of Broadcasting & Electronic Media, 60(1), 171–187. https://doi.org/10.1080/08838151.2015.1127248
Hristova, D. , Jovicic, S. , Goebl, B. , & Slunecko, T. (2020). The Social Media Game? In R. Dillon (Ed.), The Digital Gaming Handbook (1st ed., pp. 63–94). CRC Press. https://doi.org/10.1201/9780429274596-8
Kawall, J. (2002). Other-Regarding Epistemic Virtues. Ratio, 15(3), 257–275. https://doi.org/10.1111/1467-9329.00190
Kleemans, M. , Daalmans, S. , Carbaat, I. , & Anschütz, D. (2018). Picture Perfect: The Direct Effect of Manipulated Instagram Photos on Body Image in Adolescent Girls. Media Psychology, 21(1), 93–110. https://doi.org/10.1080/15213269.2016.1257392
Koch, T. K. , Frischlich, L. , & Lermer, E. (2023). Effects of Fact-Checking Warning Labels and Social Endorsement Cues on Climate Change Fake News Credibility and Engagement on Social Media. Journal of Applied Social Psychology, 53(6), 495–507. https://doi.org/10.1111/jasp.12959

Lackey, J. (2008). Learning from Words: Testimony as a Source of Knowledge. Oxford University Press.
Lackey, J. (2021). Echo Chambers, Fake News, and Social Epistemology. In S. Bernecker , A. K. Flowerree , & T. Grundmann (Eds.), The Epistemology of Fake News (pp. 208–227). Oxford University Press.
Lee, S. “Sage,” 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
Levy, N. (2022). Bad Beliefs: Why They Happen to Good People (1st ed.). Oxford University Press.
Linvill, D. L. , & Warren, P. L. (2020). Troll Factories: Manufacturing Specialized Disinformation on Twitter. Political Communication, 37(4), 447–467. https://doi.org/10.1080/10584609.2020.1718257
Luo, M. , Hancock, J. T. , & Markowitz, D. M. (2022). Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues. Communication Research, 49(2), 171–195. https://doi.org/10.1177/0093650220921321
Marsili, N. (2021). Retweeting: Its Linguistic and Epistemic Value. https://philarchive.org/rec/MARRIL-2
McDonald, L. (2021). Please Like This Paper. Philosophy, 96(3), 335–358. https://doi.org/10.1017/S0031819121000152
Mercier, H. (2020). Not Born Yesterday: The Science of Who We Trust and What We Believe. Princeton University Press. https://doi.org/10.1515/9780691198842
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
Meyer, M. , & Alfano, M. (2022). Fake News, Conspiracy Theorizing, and Intellectual Vice. In M. Alfano , C. Klein , & J. de Ridder (Eds.), Social Virtue Epistemology (pp. 236–251). Routledge.
Murray, M. (2022, September 27). *Poll: 61% of Republicans Still Believe Biden Didn’t Win Fair and Square in 2020*. NBC News. https://www.nbcnews.com/meet-the-press/meetthepressblog/poll-61-republicans-still-believe-biden-didnt-win-fair-square-2020-rcna49630
Nguyen, C. T. (2020). Games: Agency as Art. Oxford University Press.
Nguyen, C. T. (2021). How Twitter Gamifies Communication. In J. Lackey (Ed.), Applied Epistemology (pp. 410–436). Oxford University Press.
O'Sullivan, D. (2017). A Notorious Russian Twitter Troll Came Back, and for a Week Twitter Did Nothing. CNN. https://money.cnn.com/2017/11/17/media/new-jenna-abrams-account-twitter-russia/index.html
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. (2021). The Psychology of Fake News. Trends in Cognitive Sciences, 25(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007
Pennycook, G. , & Rand, D. G. (2022). Nudging Social Media toward Accuracy. The ANNALS of the American Academy of Political and Social Science, 700(1), 152–164. https://doi.org/10.1177/00027162221092342
Prier, J. (2017). Commanding the Trend: Social Media as Information Warfare. Strategic Studies Quarterly, 11(4), 50–85.
PRRI . (2022). The Persistence of QAnon in the Post-Trump Era: An Analysis of Who Believes the Conspiracies. PRRI. https://www.prri.org/research/the-persistence-of-qanon-in-the-post-trump-era-an-analysis-of-who-believes-the-conspiracies/
Rini, R. (2017). Fake News and Partisan Epistemology. Kennedy Institute of Ethics Journal, 27(S2), 43–64.

Rollins, J. (2015). Beliefs and Testimony as Social Evidence: Epistemic Egoism, Epistemic Universalism, and Common Consent Arguments: Common Consent Arguments. Philosophy Compass, 10(1), 78–90. https://doi.org/10.1111/phc3.12184
Ross, R. M. , & Levy, N. (2023). Expressive Responding in Support of Donald Trump: An Extended Replication of Schaffner and Luks (2018). Collabra: Psychology, 9(1), 68054. https://doi.org/10.1525/collabra.68054
Schaffner, B. F. , & Luks, S. (2018). Misinformation or Expressive Responding? Public Opinion Quarterly, 82(1), 135–147. https://doi.org/10.1093/poq/nfx042
Sosa, E. (2007). A Virtue Epistemology. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199297023.001.0001
Tu, F. , Pan, Z. , & Jia, X. (2023). Facts Are Hard to Come By: Discerning and Sharing Factual Information on Social Media. Journal of Computer-Mediated Communication, 28(4), zmad021. https://doi.org/10.1093/jcmc/zmad021
Weiner, M. (2003). Accepting Testimony. Philosophical Quarterly, 53(211), 256–264. https://doi.org/10.1111/1467-9213.00310
Zagzebski, L. T. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (Issue 4). Cambridge University Press.

Damned if you do

Alvarado, R. (2022). Should We Replace Radiologists with Deep Learning? Pigeons, Error and Trust in Medical AI. Bioethics, 36(2), 121–133. https://doi.org/10.1111/bioe.12959
Baghramian, M. , & Croce, M. (2021). Experts, Public Policy, and the Question of Trust. In M. Hannon & J. de Ridder (Eds.), The Routledge Handbook of Political Epistemology (1st ed., pp. 446–457). Routledge. https://doi.org/10.4324/9780429326769-53
Baier, A. (1986). Trust and Antitrust. Ethics, 96(2), 231–260. https://doi.org/10.1086/292745
Bakshy, E. , Messing, S. , & Adamic, L. A. (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. Science, 348(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160
Caplan, R. , & Gillespie, T. (2020). Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy. Social Media + Society, 6(2). https://doi.org/10.1177/2056305120936636
Christenson, D. P. , Kreps, S. E. , & Kriner, D. L. (2021). Contemporary Presidency: Going Public in an Era of Social Media: Tweets, Corrections, and Public Opinion. Presidential Studies Quarterly, 51(1), 151–165. https://doi.org/10.1111/psq.12687
Clayton, J. (2023, July 22). Intel's Deepfake Detector Tested on Real and Fake Videos. BBC News. https://www.bbc.com/news/technology-66267961
Conger, K. (2020, November 12). Twitter Says It Labeled 0.2% of All Election-Related Tweets as Disputed. The New York Times. https://www.nytimes.com/2020/11/12/technology/twitter-says-it-labeled-0-2-of-all-election-related-tweets-as-disputed.html
Crawford, K. , & Gillespie, T. (2016). What Is a Flag For? Social Media Reporting Tools and the Vocabulary of Complaint. New Media & Society, 18(3), 410–428. https://doi.org/10.1177/1461444814543163
de Keulenaar, E. , Burton, A. G. , & Kisjes, I. (2021). Deplatforming, Demotion and Folk Theories of Big Tech Persecution. Fronteiras - Estudos Midiáticos, 23(2), 118–139. https://doi.org/10.4013/fem.2021.232.09
DiResta, R. (2020, September 20). The Supply of Disinformation Will Soon Be Infinite. The Atlantic. https://www.theatlantic.com/ideas/archive/2020/09/future-propaganda-will-be-computer-generated/616400/
Durán, J. M. , & Jongsma, K. R. (2021). Who Is Afraid of Black Box Algorithms? On the Epistemological and Ethical Basis of Trust in Medical AI. Journal of Medical Ethics, medethics-2020-106820. https://doi.org/10.1136/medethics-2020-106820

Flaxman, S. , Goel, S. , & Rao, J. M. (2016). Filter Bubbles, Echo Chambers, and Online News Consumption. Public Opinion Quarterly, 80(S1), 298–320. https://doi.org/10.1093/poq/nfw006

Gillespie, T. (2018). Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media. Yale University Press.

Gillespie, T. (2022). Do Not Recommend? Reduction as a Form of Content Moderation. Social Media + Society, 8(3). https://doi.org/10.1177/20563051221117552

Goldberg, S. C. (2020). Trust and Reliance. In J. Simon (Ed.), The Routledge Handbook of Trust and Philosophy (1st ed., pp. 97–108). Routledge. https://doi.org/10.4324/9781315542294-8

Gorwa, R. , Binns, R. , & Katzenbach, C. (2020). Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance. Big Data & Society, 7(1). https://doi.org/10.1177/2053951719897945

Heatherly, K. A. , Lu, Y. , & Lee, J. K. (2017). Filtering Out the Other Side? Cross-Cutting and Like-Minded Discussions on Social Networking Sites. New Media & Society, 19(8), 1271–1289. https://doi.org/10.1177/1461444816634677

Horta Ribeiro, M. , Jhaver, S. , Zannettou, S. , Blackburn, J. , Stringhini, G. , De Cristofaro, E. , & West, R. (2021). Do Platform Migrations Compromise Content Moderation? Evidence from r/The_Donald and r/Incels. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1–24. https://doi.org/10.1145/3476057

Kleinman, Z. (2022, September 16). Anti-vax groups use carrot emojis to hide Facebook posts. *BBC News*. https://www.bbc.com/news/technology-62877597

Kozyreva, A. , Herzog, S. M. , Lewandowsky, S. , Hertwig, R. , Lorenz-Spreen, P. , Leiser, M. , & Reifler, J. (2023). Resolving Content Moderation Dilemmas between Free Speech and Harmful Misinformation. Proceedings of the National Academy of Sciences, 120(7), e2210666120. https://doi.org/10.1073/pnas.2210666120

Krishnan, N. , Gu, J. , Tromble, R. , & Abroms, L. C. (2021). Research Note: Examining How Various Social Media Platforms Have Responded to COVID-19 Misinformation. Harvard Kennedy School Misinformation Review. https://doi.org/10.37016/mr-2020-85

Levy, N. (2022). Bad Beliefs: Why They Happen to Good People (1st ed.). Oxford University Press.

Maloy, A. F. , & Vynck, G. D. (2021, November 22). How Wellness Influencers Are Fueling the Anti-Vaccine Movement. Washington Post. https://www.washingtonpost.com/technology/2021/09/12/wellness-influencers-vaccine-misinformation/

Masood, M. , Nawaz, M. , Malik, K. M. , Javed, A. , Irtaza, A. , & Malik, H. (2023). Deepfakes Generation and Detection: State-of-the-Art, Open Challenges, Countermeasures, and Way Forward. Applied Intelligence, 53(4), 3974–4026. https://doi.org/10.1007/s10489-022-03766-z

Mirsky, Y. , & Lee, W. (2022). The Creation and Detection of Deepfakes: A Survey. ACM Computing Surveys, 54(1), 1–41. https://doi.org/10.1145/3425780

Newton, C. (2019, February 25). The Secret Lives of Facebook Moderators in America. The Verge. https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona

Nguyen, C. T. (2020). Echo Chambers and Epistemic Bubbles. Episteme, 17(2), 141–161. https://doi.org/10.1017/epi.2018.32

O'Connor, C. , & Weatherall, J. O. (2019). The Misinformation Age: How False Beliefs Spread. Yale University Press.

Oremus, W. (2022, March 11). Analysis | In Putin's Russia, 'Fake News' Now Means Real News. Washington Post. https://www.washingtonpost.com/technology/2022/03/11/russia-fake-news-law-misinformation/

Paul, K. (2020, October 15). Facebook and Twitter Restrict Controversial New York Post Story on Joe Biden. The Guardian. https://www.theguardian.com/technology/2020/oct/14/facebook-twitter-new-york-post-hunter-biden

Pennycook, G. , Bear, A. , Collins, E. T. , & Rand, D. G. (2020). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of

Headlines Without Warnings. Management Science, 66(11), 4944–4957. https://doi.org/10.1287/mnsc.2019.3478
Post Editorial Board . (2020, October 14). *Facebook Censors the Post to Help Joe Biden's 2020 Campaign*. https://nypost.com/2020/10/14/facebook-censors-the-post-to-help-joe-bidens-2020-campaign/
Roberts, S. T. (2021). Behind the Screen: Content Moderation in the Shadows of Social Media; with a New Preface. Yale University Press.
Ryan, M. (2020). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. Science and Engineering Ethics, 26(5), 2749–2767. https://doi.org/10.1007/s11948-020-00228-y
Schulze, H. , Hohner, J. , Greipl, S. , Girgnhuber, M. , Desta, I. , & Rieger, D. (2022). Far-Right Conspiracy Groups on Fringe Platforms: A Longitudinal Analysis of Radicalization Dynamics on Telegram. Convergence: The International Journal of Research into New Media Technologies, 28(4), 1103–1126. https://doi.org/10.1177/13548565221104977
Slovic, P. (1993). Perceived Risk, Trust, and Democracy. Risk Analysis, 13(6), 675–682. https://doi.org/10.1111/j.1539-6924.1993.tb01329.x
Southern, M. G. (2020, September 26). Twitter Says More Users Clicking Links Before Retweeting. Search Engine Journal. https://www.searchenginejournal.com/twitter-says-more-users-clicking-links-before-retweeting/382128/
Swart, J. , Peters, C. , & Broersma, M. (2018). Shedding Light on the Dark Social: The Connective Role of News and Journalism in Social Media Communities. New Media & Society, 20(11), 4329–4345. https://doi.org/10.1177/1461444818772063
Talamanca, G. F. , & Arfini, S. (2022). Through the Newsfeed Glass: Rethinking Filter Bubbles and Echo Chambers. Philosophy & Technology, 35(1), 20. https://doi.org/10.1007/s13347-021-00494-z
Walker, M. , & Gottfried, J. (2019). 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/
Wiseman, J. (2020, October 3). Rush to Pass 'Fake News' Laws during Covid-19 Intensifying Global Media Freedom Challenges. International Press Institute. https://ipi.media/rush-to-pass-fake-news-laws-during-covid-19-intensifying-global-media-freedom-challenges/
Wright, A. T. (2021). Mill's Social Epistemic Rationale for the Freedom to Dispute Scientific Knowledge: Why We Must Put Up with Flat-Earthers. Philosopher's Imprint, 21(14). http://hdl.handle.net/2027/spo.3521354.0021.014
Yaqub, W. , Kakhidze, O. , Brockman, M. L. , Memon, N. , & Patil, S. (2020). Effects of Credibility Indicators on Social Media News Sharing Intent. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–14. https://doi.org/10.1145/3313831.3376213
Zeng, J. , & Schäfer, M. S. (2021). Conceptualizing "Dark Platforms". Covid-19-Related Conspiracy Theories on 8kun and Gab. Digital Journalism, 9(9), 1321–1343. https://doi.org/10.1080/21670811.2021.1938165

Damned if you don't

Aklin, M. , & Urpelainen, J. (2014). Perceptions of Scientific Dissent Undermine Public Support for Environmental Policy. Environmental Science & Policy, 38, 173–177. https://doi.org/10.1016/j.envsci.2013.10.006
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. , De Araujo, E. , & Mercier, H. (2022). "If This Account Is True, It is Most Enormously Wonderful": Interestingness-If-True and the Sharing of True and False News. Digital Journalism, 10(3), 373–394. https://doi.org/10.1080/21670811.2021.1941163
Benkler, Y. , Faris, R. , & Roberts, H. (2018). Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics (1st ed.). Oxford University Press.

https://doi.org/10.1093/oso/9780190923624.001.0001
Bergmann, M. (2006). Justification Without Awareness: A Defense of Epistemic Externalism (1st ed.). Oxford University Press. https://doi.org/10.1093/0199275742.001.0001
Bjerring, J. C. , Hansen, J. U. , & Pedersen, N. J. L. L. (2014). On the Rationality of Pluralistic Ignorance. Synthese, 191(11), 2445–2470. https://doi.org/10.1007/s11229-014-0434-1
Brownstein, M. , & Saul, J. M. (Eds.). (2016). Implicit Bias and Philosophy (1st ed.). Oxford University Press.
Bryanov, K. , Vasina, D. , Pankova, Y. , & Pakholkov, V. (2022). The Other Side of Deplatforming: Right-Wing Telegram in the Wake of Trump's Twitter Ouster. In D. A. Alexandrov , A. V. Boukhanovsky , A. V. Chugunov , Y. Kabanov , O. Koltsova , I. Musabirov , & S. Pashakhin (Eds.), Digital Transformation and Global Society (Vol. 1503, pp. 417–428). Springer International Publishing. https://doi.org/10.1007/978-3-030-93715-7_30
Bunker, C. J. , & Varnum, M. E. W. (2021). How Strong Is the Association between Social Media Use and False Consensus? Computers in Human Behavior, 125, 106947. https://doi.org/10.1016/j.chb.2021.106947
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
Dasgupta, N. (2013). Implicit Attitudes and Beliefs Adapt to Situations. In Advances in Experimental Social Psychology (Vol. 47, pp. 233–279). Elsevier. https://doi.org/10.1016/B978-0-12-407236-7.00005-X
De Ridder, J. (2021). What's So Bad about Misinformation? Inquiry, 1–23. https://doi.org/10.1080/0020174X.2021.2002187
Dechêne, A. , Stahl, C. , Hansen, J. , & Wänke, M. (2010). The Truth About the Truth: A Meta-Analytic Review of the Truth Effect. Personality and Social Psychology Review, 14(2), 238–257. https://doi.org/10.1177/1088868309352251
Dellsén, F. (2018). When Expert Disagreement Supports the Consensus. Australasian Journal of Philosophy, 96(1), 142–156. https://doi.org/10.1080/00048402.2017.1298636
Dellsén, F. (2021). Consensus versus Unanimity: Which Carries More Weight? The British Journal for the Philosophy of Science, 718273. https://doi.org/10.1086/718273
Dellsén, F. (2022). Excessive Testimony: When Less Is More. Philosophy and Phenomenological Research, phpr.12928. https://doi.org/10.1111/phpr.12928
Efstratiou, A. , & Caulfield, T. (2021). Misrepresenting Scientific Consensus on COVID-19: The Amplification of Dissenting Scientists on Twitter. https://doi.org/10.48550/ARXIV.2111.10594
Fallis, D. (2015). What Is Disinformation? Library Trends, 63(3), 401–426.
Fazio, L. K. , Rand, D. G. , & Pennycook, G. (2019). Repetition Increases Perceived Truth Equally for Plausible and Implausible Statements. Psychonomic Bulletin & Review, 26(5), 1705–1710. https://doi.org/10.3758/s13423-019-01651-4
Fetzer, J. H. (2004). Disinformation: The Use of False Information. Minds and Machines, 14(2), 231–240. https://doi.org/10.1023/b:mind.0000021683.28604.5b
Floridi, L. (2012). Steps Forward in the Philosophy of Information. Etica E Politica, 14(1), 304–310.
Gendler, T. S. (2008a). Alief and Belief. Journal of Philosophy, 105(10), 634–663. https://doi.org/10.5840/jphil20081051025
Gendler, T. S. (2008b). Alief in Action (and Reaction). Mind & Language, 23(5), 552–585. https://doi.org/10.1111/j.1468-0017.2008.00352.x
Ghaffary, S. (2023, March 23). Why Advertisers Aren't Coming Back to Twitter. Vox. https://www.vox.com/technology/2023/3/23/23651151/twitter-advertisers-elon-musk-brands-revenue-fleeing
Gillespie, T. (2018). Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. Yale University Press.
Glenski, M. , Volkova, S. , & Kumar, S. (2020). User Engagement with Digital Deception. In K. Shu , S. Wang , D. Lee , & H. Liu (Eds.), Disinformation, Misinformation, and Fake News

in Social Media (pp. 39–61). Springer International Publishing. https://doi.org/10.1007/978-3-030-42699-6_3
Harris, K. R. (2023a). Beyond Belief: On Disinformation and Manipulation. Erkenntnis. https://doi.org/10.1007/s10670-023-00710-6
Harris, K. R. (2023b). Liars and Trolls and Bots Online: The Problem of Fake Persons. Philosophy & Technology, 36(2), 35. https://doi.org/10.1007/s13347-023-00640-9
Hasher, L. , Goldstein, D. , & Toppino, T. (1977). Frequency and the Conference of Referential Validity. Journal of Verbal Learning and Verbal Behavior, 16(1), 107–112. https://doi.org/10.1016/S0022-5371(77)80012-1
Hassan, A. , & Barber, S. J. (2021). The Effects of Repetition Frequency on the Illusory Truth Effect. Cognitive Research: Principles and Implications, 6(1), 38. https://doi.org/10.1186/s41235-021-00301-5
Huebner, B. (2016). Implicit Bias, Reinforcement Learning, and Scaffolded Moral Cognition. In M. Brownstein & J. Saul (Eds.), Implicit Bias and Philosophy (pp. 47–79). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198713241.003.0003
Innes, H. , & Innes, M. (2023). De-Platforming Disinformation: Conspiracy Theories and their Control. Information, Communication & Society, 26(6), 1262–1280. https://doi.org/10.1080/1369118X.2021.1994631
Jaster, R. , & Lanius, D. (2021). Speaking of Fake News: Definitions and Dimensions. In S. Bernecker , A. K. Flowerree , & T. Grundmann (Eds.), The Epistemology of Fake News (pp. 19–45). Oxford University Press.
Jhaver, S. , Boylston, C. , Yang, D. , & Bruckman, A. (2021). Evaluating the Effectiveness of Deplatforming as a Moderation Strategy on Twitter. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1–30. https://doi.org/10.1145/3479525
Lackey, J. (2021). Echo Chambers, Fake News, and Social Epistemology. In S. Bernecker , A. K. Flowerree , & T. Grundmann (Eds.), The Epistemology of Fake News (pp. 208–227). Oxford University Press.
Lanius, C. , Weber, R. , & MacKenzie, W. I. (2021). Use of Bot and Content Flags to Limit the Spread of Misinformation among Social Networks: A Behavior and Attitude Survey. Social Network Analysis and Mining, 11(1), 32. https://doi.org/10.1007/s13278-021-00739-x
Levy, N. (2017, July 24). The Bad News about Fake News, Neil Levy. Social Epistemology Review and Reply Collective. https://social-epistemology.com/2017/07/24/the-bad-news-about-fake-news-neil-levy/
Levy, N. (2022). Bad Beliefs: Why They Happen to Good People (1st ed.). Oxford University Press.
Lossau, T. (2023). Knowledge as a Social Kind. Acta Analytica. https://doi.org/10.1007/s12136-023-00561-4
Luzsa, R. , & Mayr, S. (2021). False Consensus in the Echo Chamber: Exposure to Favorably Biased Social Media News Feeds Leads to Increased Perception of Public Support for Own Opinions. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 15(1). https://doi.org/10.5817/CP2021-1-3
Marks, G. , & Miller, N. (1987). Ten Years of Research on the False-Consensus Effect: An Empirical and Theoretical Review. Psychological Bulletin, 102(1), 72–90. https://doi.org/10.1037/0033-2909.102.1.72
McIntyre, L. C. (2018). Post-truth. MIT Press.
Mercier, H. (2020). Not Born Yesterday: The Science of Who We Trust and What We Believe. Princeton University Press. https://doi.org/10.1515/9780691198842
Merrill, J. B. , & Oremus, W. (2021, October 26). Five Points for Anger, One for a ‘Like’: How Facebook’s Formula Fostered Rage and Misinformation. Washington Post. https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/
Mullen, B. , Atkins, J. L. , Champion, D. S. , Edwards, C. , Hardy, D. , Story, J. E. , & Vanderklok, M. (1985). The False Consensus Effect: A Meta-Analysis of 115 Hypothesis Tests. Journal of Experimental Social Psychology, 21(3), 262–283. https://doi.org/10.1016/0022-1031(85)90020-4
Nicas, J. (2018a, August 6). Alex Jones and Infowars Content Is Removed From Apple, Facebook and YouTube. The New York Times. https://www.nytimes.com/2018a/08/06/technology/infowars-alex-jones-apple-facebook-

spotify.html
Nicas, J. (2018b, September 4). Alex Jones Said Bans Would Strengthen Him. He Was Wrong. The New York Times. https://www.nytimes.com/2018b/09/04/technology/alex-jones-infowars-bans-traffic.html
Nyhan, B. (2021). Why the Backfire Effect Does Not Explain the Durability of Political Misperceptions. Proceedings of the National Academy of Sciences, 118(15), e1912440117. https://doi.org/10.1073/pnas.1912440117
Nyhan, B. , & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. Political Behavior, 32(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2
O'Connor, C. , & Weatherall, J. O. (2019). The Misinformation Age: How False Beliefs Spread. Yale University Press.
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. , & 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
Pollock, J. L. (1984). Reliability and Justified Belief. Canadian Journal of Philosophy, 14(1), 103–114. https://doi.org/10.1080/00455091.1984.10716371
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
Rini, R. (2019). Social Media Disinformation and the Security Threat to Democratic Legitimacy.*Disinformation and Digital Democracies in the 21st Century*. Publication of the NATO Association of Canada 10–14. http://natoassociation.ca/wp-content/uploads/2019/10/NATO-publication-.pdf
Rini, R. (2021). Weaponized Skepticism: An Analysis of Social Media Deception as Applied Political Epistemology. In E. Edenburg & M. Hannon (Eds.), Political Epistemology (pp. 31–48). Oxford University Press.
Rogers, R. (2020). Deplatforming: Following Extreme Internet Celebrities to Telegram and Alternative Social Media. European Journal of Communication, 35(3), 213–229. https://doi.org/10.1177/0267323120922066
Ross, L. , Greene, D. , & House, P. (1977). The “False Consensus Effect”: An Egocentric Bias in Social Perception and Attribution Processes. Journal of Experimental Social Psychology, 13(3), 279–301. https://doi.org/10.1016/0022-1031(77)90049-X
Ross, L. , Lepper, M. R. , & Hubbard, M. (1975). Perseverance in Self-Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm. Journal of Personality and Social Psychology, 32(5), 880–892. https://doi.org/10.1037/0022-3514.32.5.880
Sherman, S. J. , Presson, C. C. , Chassin, L. , Corty, E. , & Olshavsky, R. (1983). The False Consensus Effect in Estimates of Smoking Prevalence: Underlying Mechanisms. Personality and Social Psychology Bulletin, 9(2), 197–207. https://doi.org/10.1177/0146167283092003
Smelter, T. J. , & Calvillo, D. P. (2020). Pictures and Repeated Exposure Increase Perceived Accuracy of News Headlines. Applied Cognitive Psychology, 34(5), 1061–1071. https://doi.org/10.1002/acp.3684
Swire-Thompson, B. , DeGutis, J. , & Lazer, D. (2020). Searching for the Backfire Effect: Measurement and Design Considerations. Journal of Applied Research in Memory and Cognition, 9(3), 286–299. https://doi.org/10.1016/j.jarmac.2020.06.006
Unkelbach, C. , & Koch, A. (2019). Gullible but Functional? In J. P. Forgas & R. F. Baumeister (Eds.), The Social Psychology of Gullibility (1st ed., pp. 42–60). Routledge. https://doi.org/10.4324/9780429203787-3
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
Wojcieszak, M. (2008). False Consensus Goes Online. Public Opinion Quarterly, 72(4), 781–791. https://doi.org/10.1093/poq/nfn056
Wood, T. , & Porter, E. (2019). The Elusive Backfire Effect: Mass Attitudes’ Steadfast Factual Adherence. Political Behavior, 41(1), 135–163. https://doi.org/10.1007/s11109-018-9443-y

Woolley, S. (2022). Manufacturing Consensus: Understanding Propaganda in the Era of Automation and Anonymity. Yale University Press.

Collaborative content moderation

Allen, J. , Arechar, A. A. , Pennycook, G. , & Rand, D. G. (2021). Scaling Up Fact-Checking Using the Wisdom of Crowds. Science Advances, 7(36), eabf4393. https://doi.org/10.1126/sciadv.abf4393

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

Altay, S. , Hacquin, A.-S. , & Mercier, H. (2022). Why Do So Few People Share Fake News? It Hurts Their Reputation. New Media & Society, 24(6), 1303–1324. https://doi.org/10.1177/1461444820969893

Altay, S. , Nera, K. , Ejaz, W. , Schöpfer, C. , & Tomas, F. (2023b). Conspiracy Believers Claim to Be Free Thinkers But (Under)Use Advice Like Everyone Else. British Journal of Social Psychology, bjso.12655. https://doi.org/10.1111/bjso.12655

Arechar, A. A. , Allen, J. , Berinsky, A. J. , Cole, R. , Epstein, Z. , Garimella, K. , Gully, A. , Lu, J. G. , Ross, R. M. , Stagnaro, M. N. , Zhang, Y. , Pennycook, G. , & Rand, D. G. (2023). Understanding and Combatting Misinformation across 16 Countries on Six Continents. Nature Human Behaviour. https://doi.org/10.1038/s41562-023-01641-6

Ballantyne, N. , Celniker, J. B. , & Dunning, D. (2022). Do Your Own Research. Social Epistemology, 1–16. https://doi.org/10.1080/02691728.2022.2146469

Buzzell, A. , & Rini, R. (2023). Doing Your Own Research and Other Impossible Acts of Epistemic Superheroism. Philosophical Psychology, 36(5), 906–930. https://doi.org/10.1080/09515089.2022.2138019

Dechêne, A. , Stahl, C. , Hansen, J. , & Wänke, M. (2010). The Truth About the Truth: A Meta-Analytic Review of the Truth Effect. Personality and Social Psychology Review, 14(2), 238–257. https://doi.org/10.1177/1088868309352251

Duggan, M. (2017, July 11). Online Harassment 2017. Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/

Eady, G. , Paskhalis, T. , Zilinsky, J. , Bonneau, R. , Nagler, J. , & Tucker, J. A. (2023). Exposure to the Russian Internet Research Agency Foreign Influence Campaign on Twitter in the 2016 US Election and Its Relationship to Attitudes and Voting Behavior. Nature Communications, 14(1), 62. https://doi.org/10.1038/s41467-022-35576-9

Frost-Arnold, K. (2023). Who Should We Be Online? A Social Epistemology for the Internet. Oxford University Press.

Goldman, A. I. (2001). Experts: Which Ones Should You Trust? Philosophy and Phenomenological Research, 63(1), 85–110. https://doi.org/10.1111/j.1933-1592.2001.tb00093.x

Golovchenko, Y. , Buntain, C. , Eady, G. , Brown, M. A. , & Tucker, J. A. (2020). Cross-Platform State Propaganda: Russian Trolls on Twitter and YouTube during the 2016 U.S. Presidential Election. The International Journal of Press/Politics, 25(3), 357–389. https://doi.org/10.1177/1940161220912682

Groh, M. , Epstein, Z. , Firestone, C. , & Picard, R. (2022). Deepfake Detection by Human Crowds, Machines, and Machine-Informed Crowds. Proceedings of the National Academy of Sciences, 119(1), e2110013119. https://doi.org/10.1073/pnas.2110013119

Grundmann, T. (2021). Preemptive Authority: The Challenge from Outrageous Expert Judgments. Episteme, 18(3), 407–427. https://doi.org/10.1017/epi.2021.30

Guerrero, A. A. (2017). Living with Ignorance in a World of Experts. In R. Peels (Ed.), Perspective on Ignorance from Moral and Social Philosophy (pp. 156–185). Routledge.

Harris, K. R. (2023). Epistemic Domination. Thought: A Journal of Philosophy. https://doi.org/10.5840/tht202341317

Kosseff, J. (2022). Real Name Policies: How Facebook and Others Decided – Protocol. https://www.protocol.com/policy/anonymity-real-names-jeff-kosseff

Lackey, J. (2018). Experts and Peer Disagreement (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780198798705.003.0012
Levy, N. (2022). Do Your Own Research! Synthese, 200(5), 356. https://doi.org/10.1007/s11229-022-03793-w
Mai, K. T. , Bray, S. D. , Davies, T. , & Griffin, L. D. (2023). *Warning: Humans Cannot Reliably Detect Speech Deepfakes* (arXiv:2301.07829). arXiv. http://arxiv.org/abs/2301.07829
Marwick, A. E. , & Partin, W. C. (2022). Constructing Alternative Facts: Populist Expertise and the QAnon Conspiracy. New Media & Society. https://doi.org/10.1177/14614448221090201
Nichols, T. M. (2017). The Death of Expertise: The Campaign against Established Knowledge and Why It Matters. Oxford University Press.
Oliver, J. E. , & Rahn, W. M. (2016). Rise of the Trumpenvolk: Populism in the 2016 Election. The ANNALS of the American Academy of Political and Social Science, 667(1), 189–206. https://doi.org/10.1177/0002716216662639
Pennycook, G. , Bear, A. , Collins, E. T. , & Rand, D. G. (2020). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. Management Science, 66(11), 4944–4957. https://doi.org/10.1287/mnsc.2019.3478
Pennycook, G. , & Rand, D. G. (2019). Fighting Misinformation on Social Media Using Crowdsourced Judgments of News Source Quality. Proceedings of the National Academy of Sciences, 116(7), 2521–2526. https://doi.org/10.1073/pnas.1806781116
Portes, R. (2017). The People of This Country Have Had Enough of Experts. London Business School. https://www.london.edu/think/who-needs-experts
Rathgeb, C. , Nichols, R. , Ibsen, M. , Drozdowski, P. , & Busch, C. (2022). *Crowd-Powered Face Manipulation Detection: Fusing Human Examiner Decisions* (arXiv:2201.13084). arXiv. http://arxiv.org/abs/2201.13084
Rini, R. (2017). Fake News and Partisan Epistemology. Kennedy Institute of Ethics Journal, 27(S2), 43–64.
Saurette, P. , & Gunster, S. (2011). Ears Wide Shut: Epistemological Populism, Argutainment and Canadian Conservative Talk Radio. Canadian Journal of Political Science, 44(1), 195–218. https://doi.org/10.1017/S0008423910001095
Skurnik, I. , Yoon, C. , Park, D. C. , & Schwarz, N. (2005). How Warnings about False Claims Become Recommendations. Journal of Consumer Research, 31(4), 713–724. https://doi.org/10.1086/426605
Surowiecki, J. (2005). *The Wisdom of Crowds* (Nachdr.). Anchor Books.
Waruwu, B. K. , Tandoc, E. C. , Duffy, A. , Kim, N. , & Ling, R. (2021). Telling Lies Together? Sharing News as a Form of Social Authentication. New Media & Society, 23(9), 2516–2533. https://doi.org/10.1177/1461444820931017
Ylä-Anttila, T. (2018). Populist Knowledge: 'Post-Truth' Repertoires of Contesting Epistemic Authorities. European Journal of Cultural and Political Sociology, 5(4), 356–388. https://doi.org/10.1080/23254823.2017.1414620

Epilogue

Benkler, Y. , Faris, R. , & Roberts, H. (2018). Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics (1st ed.). Oxford University Press. https://doi.org/10.1093/oso/9780190923624.001.0001
Donovan, J. , Dreyfuss, E. , & Friedberg, B. (2022). Meme Wars: The Untold Story of the Online Battles Upending Democracy in America. Bloomsbury Publishing.
Douglas, K. M. , Sutton, R. M. , & Cichocka, A. (2017). The Psychology of Conspiracy Theories. Current Directions in Psychological Science, 26(6), 538–542. https://doi.org/10.1177/0963721417718261

Funkhouser, E. (2017). Beliefs as Signals: A New Function for Belief. Philosophical Psychology, 30(6), 809–831. https://doi.org/10.1080/09515089.2017.1291929
Funkhouser, E. (2022). A Tribal Mind: Beliefs that Signal Group Identity or Commitment. Mind & Language, 37(3), 444–464. https://doi.org/10.1111/mila.12326
Ganapini, M. B. (2023). The signaling function of sharing fake stories. Mind & Language, 38(1), 64–80. https://doi.org/10.1111/mila.12373
Hannah, M. N. (2021). A Conspiracy of Data: QAnon, Social Media, and Information Visualization. Social Media + Society, 7(3). https://doi.org/10.1177/20563051211036064
Ichino, A. , & Räikkä, J. (2020). Non-Doxastic Conspiracy Theories. Argumenta, https://doi.org/10.14275/2465-2334/20200.ich
Mercier, H. (2020). Not Born Yesterday: The Science of Who We Trust and What We Believe. Princeton University Press. https://doi.org/10.1515/9780691198842
Uscinski, J. , Enders, A. M. , Klofstad, C. , & Stoler, J. (2022). Cause and Effect: On the Antecedents and Consequences of Conspiracy Theory Beliefs. Current Opinion in Psychology, 47, 101364. https://doi.org/10.1016/j.copsyc.2022.101364
Whitehurst, L. (2022, December 16). QAnon Follower Who Chased Capitol Officer on Jan. 6 gets 5 years. PBS NewsHour. https://www.pbs.org/newshour/politics/qanon-follower-who-chased-capitol-officer-on-jan-6-gets-5-years
Williams, D. (2022). Signalling, Commitment, and Strategic Absurdities. Mind & Language, 37(5), 1011–1029. https://doi.org/10.1111/mila.12392
Williams, D. (2023). The Marketplace of Rationalizations. Economics and Philosophy, 39(1), 99–123. https://doi.org/10.1017/S0266267121000389