Fake News!

James Owen Weatherall, Cailin O'Connor

Department of Logic and Philosophy of Science
University of California, Irvine

Abstract

We review several topics of philosophical interest connected to misleading online content. First we consider proposed definitions of different types of misleading content. Then we consider the epistemology of misinformation, focusing on approaches from virtue epistemology and social epistemology. Finally we discuss how misinformation is related to belief polarization, and argue that models of rational polarization present special challenges for conceptualizing fake news and misinformation.

1. Introduction

From 19th century yellow journalism to state-sponsored propaganda in the 20th century, there is a long history of misleading mass media impacting on major political decision-making. But around the time of the 2016 Brexit vote and U.S. Presidential Election, a new epistemic problem emerged—one continuous with other historical examples, but also distinct in several respects. For the first time, misleading content specifically designed to spread effectively on online social media platforms had become sufficiently pervasive and effective to plausibly influence election outcomes. A particular phenomenon encapsulated this new threat to political epistemology: websites that were deliberately designed to persuade readers that they were small, independent media organizations distributing stories that adhered to standard journalistic norms (Rini, 2017; O'Connor and Weatherall, 2019). These sites had names and URLs reminiscent of news outlets; their page designs mimicked other news sites; and they published content with eye-catching, politically relevant headlines and striking

Email addresses: weatherj@uci.edu (James Owen Weatherall), cailino@uci.edu (Cailin O'Connor)

visuals tailor-made for viral distribution on social media. Articles hosted on these sites were soon given a name: they were *fake news*.

Fake news was just the beginning. The rise of this specific form of political propaganda, apparently with multiple simultaneous purposes (to mislead on specific issues; to generate ad revenue; and to undermine public trust in institutions), both augered and emblemized the new kinds of epistemic challenges of a highly connected, fast-paced world fueled by social media, artificial intelligence, and increasingly sophisticated economic and political groups who recognize the value of manipulating public belief (O'Connor and Weatherall, 2019; Lewandowsky et al., 2020; Novaes and de Ridder, 2021). Forms of misleading content shifted and multiplied, contributing to a fantastically complex online epistemic environment. In the years following, researchers across the sciences and humanities have turned their attention to the epistemic issues arising in light of misleading online content, the harms to democratic decision-making they produce, and possible remedies for them.

This article will introduce and review some of the philosophically rich issues raised by this rapidly changing media and epistemic landscape. The literature is now too large to cover in its entirety, and so we will focus on a handful of related issues. We will begin, in section 2, with a discussion of the now widespread distinction between misinformation and disinformation, along with other alternatives, such as "malinformation". We will engage in particular with recent attempts to define disinformation and explore challenges in maintaining sharp distinctions in this conceptual space. Then, in section 3, we discuss the epistemology of misinformation. We start with work in virtue epistemology on susceptibility to misinformation; and then review work arguing that the persistence and propagation of misinformation must be understood from a social perspective. Finally, in section 4, we will briefly review recent work on belief polarization and discuss how this bears on our assessments of misinformation and its impacts. Section 5 concludes.

2. What is Misleading Content?

Misleading content includes text, images, audio, video, or some combination of these that tends to decrease the accuracy of a consumer's beliefs, at least for some audiences. Such content is sometimes broadly referred to as "misinformation". But this blanket term can itself be misleading. There is enormous heterogeneity in the types of misleading content out there, the intentions of its originators, the mechanisms by which it works, and the social consequences of its dissemination. Moreover, whether and how content may mislead generally depends on a viewer's background beliefs, which are dynamic and influenced by the ever changing ecosystem of information to which they are exposed. And, in addition, there is an ongoing arms race between (1) online platforms, (2) creators of misleading content, and (3) consumers of misleading content. This arms race leads to rapid change in what is produced, how it is shared, and what effects it has. This means that the forms of misinformation, broadly understood, are ever evolving, but also that the sorts of content that successfully mislead are constantly in flux (O'Connor and Weatherall, 2019; Lewandowsky et al., 2020).

A distinction is often drawn between "misinformation" and "disinformation" (Fallis, 2016). Misinformation, in this context, is content that tends to be misleading, but which is generated or shared without intention to mislead. It is authentic in the sense that the person or group who has created and distributed it believes it to be true and shares it either with the intent to educate or as some other, non-epistemic expression—say, of emotion of political affiliation (Funkhauser, 2017). Disinformation, by contrast, is produced and shared with the intent to mislead. Disinformation may be generated to shape public belief in a way that is advantageous for the producer; or it may be intended to introduce confusion, uncertainty, or to undermine public trust in some institution. (Sometimes misinformation is used as an umbrella term that includes disinformation as a special case (e.g. Fallis, 2015; O'Connor and Weatherall, 2019); others, such as Jaster and Lanius (2021) and Pritchard (2021), define "fake news" as essentially synonymous with disinformation as we define it

here: misleading content intended to mislead.)

Precisely how to think about disinformation has been a subject of some philosophical dispute, connected to the literatures on lying, bullshitting, and deception more generally (Chisholm and Feehan, 1977; Bok, 1978; Frankfurt, 2005; Carson, 2006; Fallis, 2009; Saul, 2012). In a prescient early paper discussing the likelihood of disinformation playing a significant role on the internet, Floridi (1996) defines disinformation very broadly, as arising in (all and only) cases in which "the process of information is defective" (509). He gives three examples of how this can happen: via propaganda, i.e., content intended to persuade without regard for the truth; incompleteness of information, as when some relevant facts are neglected; and censorship, where certain sources of relevant information are silenced. These examples suggest intentionality, but his definition does not require that, and so it is not clear that disinformation is distinguished from misinformation on this account. In later work, Floridi amends his definition with the additional proviso that the source of the information must *know* that it is false and that the source must purposely convey the content in order to mislead those who receive it (Floridi, 2011). Fetzer (2004) offers a similar analysis, analogizing disinformation to lying: it is something with semantic content that the source believes to be false, which is disseminated in order to mislead.

Fallis (2015) argues that all such accounts are too narrow, because they require that disinformation be *false*. But as he argues, in some cases disinformation is true but nonetheless both misleading and intended to mislead. In this connection, disinformation is sometimes distinguished from "malinformation" (Wardle and Derakhshan, 2017). For those who draw the distinction, disinformation and malinformation are both produced with the intention to mislead, but whereas disinformation involves intentional factual inaccuracies, malinformation is true or accurate information. For instance, it would be disinformation to assert, with the intention to mislead, that "science shows that vaccines cause autism"; but malinformation to assert that "the well-respected medical journal *Lancet* published an article

linking autism and vaccines". The latter statement is entirely accurate, but neglects important context, such as that this article was retracted by the journal in 2010, that hundreds of follow-up studies failed to replicate its findings, and that the overwhelming consensus among medical researchers is that there is no such link. (If either statement was shared by someone who sincerely believed it, with no intent to mislead, then it would be misinformation.)

The mis/dis/mal-information distinction is helpful for some purposes. For instance, drawing attention to the fact that true statements can decrease belief accuracy may be helpful for media literacy training, since it highlights the importance of contextualizing assertions even if they are true. Meanwhile, since many legal frameworks turn on whether an action is taken with corrupt intent, the question of someone's purpose in sharing some information may be important for navigating thorny issues around free speech and misleading content. But there are also many ways in which these distinctions are problematic. For instance, for many authors the distinction between mis- and dis/mal- information is intent. Yet there is little reason to think that the intentions of the person generating and sharing information have any bearing on its ecological impact after it has been generated. A meme making false claims about race and intelligence is equally harmful whether it is generated by a white supremacist who believes the content, a foreign government entity seeking to increase political polarization, or someone seeking to monetize engagement on a website without concern for the other impacts of their content.

Another problem is that many types of misleading content are designed to be widely shareable, but in general different people sharing the content may have different intentions. For instance, one person might assert that sunspots are primarily responsible for driving global climate change with full knowledge that this is not true, and another person, upon hearing it, may be convinced by the statement and then pass it along. Thus the same content can be disinformation when shared by some people but misinformation for others. This means the labels are not easily associated with the content itself, but rather the content

and the specific context of distribution.

More recently, several authors have challenged standard accounts of disinformation along other lines. For instance, Harris (2023) argues that disinformation need not be intended to *mislead*, in the sense of producing false beliefs. Instead, it could be intended to make it more difficult to form true beliefs or to affect an audience's behavior in other ways. Simion (2023), too, argues that it is not essential to disinformation that it be false or misleading, nor that it be intended to mislead. On her account, what makes disinformation distinctive is that it generates *ignorance*, which she argues is distinct from promoting false belief.

There is another way of thinking about what makes disinformation – and malinformation – distinctive that draws on the biological literature on deceptive signaling (see e.g. Skyrms, 2010). The idea is that, by creating inaccurate beliefs, some signals have the effect of benefiting the sender at the expense of the receiver. Such signals may originate even from low-rationality signallers, such as bacteria, and so it would seem that concepts such as "intent to mislead" would be inappropriate; nonetheless, they are deceptive, because of the structural role they play in coordinating behaviors that lead to payoffs different from what the receiver expects or prefers. This conception of deception is indifferent to the "truth" of the content that is transmitted. Inspired by these ideas, Fallis (2015) offers a different account of disinformation, as misleading content whose *function* is to mislead. Function, here, is meant in the sense of biological functions, which is to say that something has a function if it plays a certain role in some broader system and that role explains its existence, in the sense that it has been selected or created because it plays that role (Millikan, 1984; Neander, 1991; Garson, 2019). Something can have a function because it was designed to have that function, but no designer is necessary. Something can evolve (biologically, culturally, or otherwise) to have a function.

Fallis's account avoids concerns about intent, since content can have the function of misleading irrespective of whether downstream sharers intend to mislead anyone. On the

other hand, it erodes the distinction between disinformation and misinformation in other ways, since inadvertently misleading content or content that is shared by someone who believes it to be true – plausibly, misinformation – can nonetheless play the right sorts of structural roles in an information ecosystem to have the function of misleading. Consider, for instance, information shared between concerned parents expressing skepticism about the safety and efficacy of vaccines. Often this information is false, but authentically believed by the person distributing it; there is no intention to mislead, but rather to educate or assist; and yet, because of the source and means of transmission of the content, it can be especially well-suited to misleading people.[1] This account also does not capture cases like the ones Harris (2023) or Simion (2023) have in mind, where the purpose of disinformation is a different kind of epistemic failure than producing false beliefs.

In our view, it is important to have the conceptual tools to recognize that not all misleading content is false, that the motivations of those consuming and sharing misleading content can come apart from those that create it, that some content can be particularly well-suited to mislead, and that false beliefs are not always the goal of disinformation campaigns. Less clear is that the best way to chart this conceptual space is by introducing sharp distinctions between disinformation, misinformation, and other terms in the neighborhood—or to argue about what definition properly captures disinformation, as opposed to some other closely related concept. For this reason, we will use "misinformation" as an umbrella term in what follows, and we will elaborate on the species of misinformation at issue as necessary.

In some ways, the attitude just described resonates with the arguments of Habgood-Coote (2019), who argues against using the expression "fake news", in part because its meaning is contested, and in part because it is not clear how to generalize from the various examples one often finds in popular discussions to a clear concept. For him, it is best

[1]That said, it is not clear if such content is deceptive in the Skyrmsian sense, since parents sharing vaccine misinformation do not obviously benefit from transmitting that information.

to simply describe specific epistemic failures at issue in any particular case – say, failures of transmission of reliable information on social media, or biases in news media – rather than use a generic term with unclear meaning. We tend to disagree with him about "fake news" in particular, both because it is rarely presented as precise philosophical terminology (pace Gelfert, 2018; Fallis and Mathiesen, 2019; Pritchard, 2021; Jaster and Lanius, 2021) and because it has a clear core meaning in contemporary discourse, flowing from a specific phenomenon from 2016. But we do find his arguments persuasive when applied to attempts to distinguish misinformation from disinformation, malinformation, and other similar terms. There is more heterogeneity in the practices that get labeled with these terms than they can reflect, and they are often presented as if they clearly map a conceptual space that is in fact far more textured.[2]

3. The Epistemology of Misinformation

So much for what misinformation is. What does it do? A preliminary answer would be that it causes its consumers to form or strengthen false beliefs. (We have already pointed out that some misinformation does not exactly function this way, but set this aside for now.) There is an active literature in cognitive science that seeks to adjudicate whether mere exposure (or repeated exposure) to some (false) content is sufficient to induce belief (Hasher et al., 1977; Gilbert et al., 1990, 1993; Brasher and Marsh, 2020), or if instead human reasoners generally exercise "epistemic vigilance" by assessing information for plausibility, coherence with other beliefs, and markers of epistemic quality before forming beliefs on its basis (Sperber et al., 2010; Mercier and Sperber, 2019). Suppose we assume that at least sometimes, humans do exercise judgment about what they accept. (This is apparently accepted even by advocates of the view that humans are generally biased towards accepting information irrespective of its

[2]That said, we also disagree that the correct remedy is to simply drop the problematic terminology, rather than to use it with a clear picture of its limits.

epistemic merits.) What needs to be true about an epistemic agent and their environment for them to be persuaded by misinformation? What makes some misinformation more effective than other misinformation? In this section we will discuss two (compatible) approaches to these questions: virtue epistemology and social epistemology.

3.1. Virtue Epistemology

One influential way of thinking about these questions draws on virtue epistemology. Virtue epistemology is an umbrella term that encompasses may different approaches; reviewing all of them would be beyond the scope of this article (but see Greco, 1993, 2003; Axtell, 1997; Kvanig, 2010; Battaly, 2008; Turri et al., 2021).[3] For present purposes, it suffices to note that virtue epistemology emphasizes the role of good epistemic practices and personal characteristics ("epistemic virtues") – and poor ones ("epistemic vices") – in knowledge production and transmission. Work on misinformation in this context tends to emphasize the importance of virtues and vices in explaining susceptibility to misinformation and the tendency to share it (Montmarquet, 1993; Zagzebski, 1996; Roberts and Wood, 2007; Battaly, 2008; Cassam, 2018a,b).

For instance, Cassam (2018a) has discussed the epistemic vice of *insouciance* in this context (c.f. Frankfurt, 2005). Epistemic insouciance is indifference to the truth: "a casual lack of concern about the facts or an indifference to whether [one's] ... statements have any basis in reality" (2). Cassam argues that epistemic insouciance is characteristic of "post-truth" politics, wherein politicians and their followers simply dismiss expert knowledge and evidence in favor of freely sharing content that elicits an epistemic (or affective) reaction

[3]Virtue epistemology also has prominent critics, often echoing well-known criticisms of virtue ethics. For instance, the "situationist" response to virtue epistemology is that human cognition is too sensitive to irrelevant contextual factors for human agents to reliably embody epistemic virtues, and thus the presence of those virtues cannot underwrite knowledge claims (Olin and Doris, 2014). But this line of thought does not undermine the idea that epistemic vices contribute to susceptibility to misinformation, or that cultivating virtues, to the extent it is possible, can combat it.

friendly to their political goals. Thus when Michael Gove famously declared, in the lead-up to the Brexit vote in 2016, "The people of this country have had enough of experts", he was not rebutting arguments that Brexit would be economically harmful, so much as rejecting the idea that arguments or evidence were relevant to policy discussions. Those who are epistemically insouciant may be more likely to accept – and share – claims that are false or poorly evidenced, but which are broadly compatible with their worldview, problematic for their political opponents, or seen as a signal of political affiliation. Epistemic insouciance may help explain why easily refuted claims seem to persist in some online communities, or why clear examples of disinformation can spread widely, without apparent regard for their reliability.

Similarly, Wright (2021) has argued that cultivating epistemic trustworthiness – the virtue of only sharing information that you have vetted – can help mitigate the epistemic risks associated with re-posting information online. These arguments are related to work by Lynch (2018, 2019), who has argued that epistemic arrogance – "an unwillingness to learn from others arising from a distorted relationship with the truth" (Lynch, 2018, p. 284) – can undermine public discourse, because it leads agents to reject relevant evidence and information shared by others. While Lynch does not specifically address the role of epistemic arrogance in the acceptance and distribution of misinformation, there is little doubt they are linked, and subsequent work has suggested that intellectual humility, the virtue opposed to epistemic arrogance, is an important counterweight to misinformation (Porter et al., 2022) (though see Levy (2021) for an interesting counterpoint). Roughly, the idea is that arrogance can lead agents to share content without scrutinizing it, reducing trustworthiness, whereas humility can make agents more cautious about sharing, and thus more trustworthy.

Priest (2021), meanwhile, offers a different perspective on the role of epistemic vice in the spread of misinformation. She argues that it is the vices of epistemic elites – journalists, scientists, left-leaning celebrities – that drive political divides. Specifically, she argues, epis-

temic *insensitivity* – failing to recognize the ways in which non-epistemic factors influence one's beliefs – and epistemic *obstruction* – overreliance on technical jargon and complex theory – are particularly damaging when embodied by experts, because they erode trust.

There is some empirical work that supports the claim that epistemic vices contribute to susceptibility to misinformation. For instance, Meyer et al. (2021b) conduct an observational study in which they survey participants to determine whether they subscribed to various false beliefs concerning COVID-19. They then studied the relationship between these false beliefs and demographic, political, and other factors. They found that the strongest correlation with holding false beliefs about COVID-19 was with participants' score on an "epistemic vice scale" (Meyer et al., 2021a). This scale measured epistemic vice across several dimensions, including indifference to the truth and rigidity. Similarly, Koetke et al. (2022) show that the virtue of intellectual humility is correlated with a more skeptical attitude towards misinformation about COVID-19, leading to more careful assessments of COVID information and improving resilience. More generally, Vasilyeva et al. (2021) found a large sample of participants across six countries to be effective at distinguishing true and false statements about COVID under a range of experimental conditions, which they argued was evidence for epistemic vigilance across different populations. (See Porter et al., 2022, for a partial review of this empirical literature.)

3.2. Social Epistemology

We now turn to a different approach to the epistemology of misinformation, which emphasizes the social aspects of misinformation receptivity (O'Connor et al., 2024). (As discussed below, virtue and social epistemology are compatible and often combined.) The basic observation is that humans rely on one another for knowledge about the world. Most of what we believe we learn from (or with) others, whether through testimony, reporting, or collaboration. We make judgments about what to believe and how to act on the basis of what

others in our communities believe and do; and we make judgments about whom we should trust and who has expertise on various issues. In some cases, expressed belief can even function as a social signal indicating group membership or status (Funkhauser, 2017). For some philosophers, it is the way misinformation interacts with these social factors that ultimately explains why misinformation is an epistemic problem, particularly in the context of social media.

One way of understanding how misinformation can persist and propagate concerns the dynamics of differential trust. Agents are presented with a broad spectrum of information with differing levels of quality and reliability. Even virtuous agents must make judgments about what is ultimately a reliable source of information—including, in the extreme case where someone embarks on an independent investigation of some issue, whether there may be good reasons why one's own findings are less reliable than reported findings of others with more experience, better equipment, larger sample sizes, and so forth. How are these judgments to be made? The challenge is ultimately a version of the two experts, one novice problem: how can someone with little expertise adjudicate expert disagreements (Goldman, 2001)?

Consider, for instance, the particular species of misinformation discussed in the introduction: fake news. Fallis and Mathiesen (2019) argue that what makes fake news effective is that it is *counterfeit* news. In other words, it attempts to gain trust and legitimacy by mimicking a source of information generally taken to be reliable. It is because journalists play a certain social role as accurate reporters, and they are governed, at least in principle, by a professional ethics code intended to preserve that social role, that mimicking journalistic content can generate false beliefs. Fake news, as it appeared in 2016, had the trappings of reliable content. And changes in media literacy, social media sharing policy, and broader cultural awareness were necessary before many people could distinguish reliable news from fake news.

Similar mechanisms are at work with certain forms of scientific misinformation. For instance, as discussed by Oreskes and Conway (2010) and O'Connor and Weatherall (2019), industrial and political propagandists often use legitimate scientific results – that is, scientific articles published in peer-reviewed journals by scientists acting in epistemic good faith – to persuade people, including lay people, politicians, and even medical professionals, of things that are beneficial to the propagandist, irrespective of whether they are true or well-supported by scientific evidence. How can this happen? Scientific evidence is generally statistical in character, which means that one should expect some fraction of published studies to show results that are spurious, in the sense that they suggest conclusions that are false (and, generally, not supported by the total body of evidence available). By selectively sharing results of this sort (Weatherall et al., 2018), or via "Industrial Selection", the process of selectively funding research that uses methods more likely to generate spurious results (Holman and Bruner, 2017), propagandists can give a misleading picture of what the (total) evidence really shows. This approach can be especially powerful because the results shared are not counterfeit at all.

These phenomena can also occur without any propagandist, or any intention to mislead. Mohseni et al. (2024) consider how scientific "curators", including science journalists, social media algorithms, textbook writers, and even review articles, can introduce evidential biases. They use a model to show how curation methods that are generally considered ethically acceptable, such as only reporting on extreme cases or reporting on both sides of an issue, can be problematic, especially when media consumers exhibit confirmation bias. Meanwhile, practices within science itself can give rise to similar sorts of endogenous scientific misinformation. For instance, it is now well-documented that retracted scientific articles continue to be cited and shared as if they are reliable (see e.g. Neale et al., 2010; Schneider et al., 2020; LaCroix et al., 2021; Genot and Olsson, 2021)—a phenomenon that O'Connor and Weatherall (2020) have dubbed "information zombies". These authentic, but now question-

able or refuted, articles have all the hallmarks of authoritative information, making them an especially dangerous form of persistent misinformation. West and Bergstrom (2020) argue that other practices internal to science, such as hyperbole and citation misdirection, where papers providing at best thin evidence for a claim are repeatedly cited as definitive, can also lead to distinctive forms of scientific misinformation.

Virtue epistemology and social epistemology are not incompatible, and many authors emphasize the importance of epistemic virtues like those already discussed in social settings. Perhaps most striking, though, is that some authors argue that individual social vices may become virtues at a social level. Smart (2018) describes this sort of phenomenon as "Mandevillian intelligence", i.e., group level intelligence that emerges as a result of individual cognitive vices. To take a concrete example, Levy and Alfano (2020) argue that "cumulative culture" (or "cumulative knowledge"), that is, knowledge that is primarily learned from others, perhaps elaborated or developed, and then passed on further, depends on accepting the testimony of epistemic authorities, often with minimal skepticism or independent verification. The development of experimental sciences illustrates the point: no individual researcher can conduct every experiment themselves, and so the totality of scientific knowledge depends on researchers accepting the work of other scientists without repeating those experiments. Levy and Alfano argue that the dispositions needed for the effective development and transmission of cumulative knowledge tend to look more like epistemic vices, such as blind trust of authority and dogmatism about established results, than epistemic virtues. And thus it seems that a powerful form of knowledge depends on epistemic vice.

This argument supports a version of what Mayo-Wilson et al. (2011) call the "Independence Thesis", which is the claim that individual-level rationality and group-level rationality come apart. This suggests an interesting relationship between the social epistemological factors that drive the persistence and spread of false belief and virtue epistemological explanations, since it suggests that at least some traits that make individuals susceptible to

misinformation, and which contribute to its spread at a social level, may nonetheless have large-scale epistemic benefits that outweigh the harms.

4. Polarization

In this final section, we turn to belief polarization. This is a topic of recent philosophical interest that does not directly concern misinformation, but which bears on how we assess misinformation and its impacts. Roughly, belief polarization occurs when agents in an epistemic community fail to converge to shared beliefs about matters of fact, even in the face of debate and discussion. Of particular interest are cases where all agents have access to the same evidence (or evidence of the same kind, e.g., showing the same statistical regularities). One might imagine – in light of Bayesian merging theorems, for instance – that polarized agents in such communities must exhibit some form of irrationality; and various media commentators have suggested that misinformation may contribute to polarization. But if polarization can arise endogenously, without introducing misleading content, such proposals are less compelling.

Several authors have shown how polarization can emerge in just this way (for further refernces see Bramson et al., 2017). Using different modeling frameworks, Hegselmann and Krause (2006) and Olsson (2013) both show that when agents discount testimony from other agents whose opinions are too distant from their own, polarization can occur. O'Connor and Weatherall (2018) show that similar effects can arise when agents share evidence instead of opinions, but where agents treat evidence shared by sources whose beliefs differ from theirs as uncertain (see also Weatherall and O'Connor, 2021b). Singer et al. (2019) show that boundedly rational agents can likewise polarize over matters of fact, while Weatherall and O'Connor (2021a) show that polarization can arise due to network effects when agents conform their actions to those in their immediate environment.

Each of these models show how various plausible mechanisms can generate polarization.

But none of them establish that fully rational agents can polarize, and so one might argue, say from the perspective of the virtue epistemologists discussed above, that polarization can be avoided if agents are more rational or avoid problematic heuristics. Another class of models makes the case that even this is not enough. For instance, Jern et al. (2014), Cook and Lewandowsky (2016), and Freeborn (2023) all describe situations in which fully Bayesian agents can diverge in their beliefs when presented with the same evidence. In all of these cases, agents are represented as having Bayesian networks of belief (Pearl, 1986), but differing in their priors on some nodes in a way that affects how they respond to evidence directly bearing on other nodes. More striking still is recent work by Dorst (2023), who argues that higher order uncertainty regarding how to rationally respond to information can lead to polarization of a particularly strong kind, where agents can *predict* that they will mutually polarize. Dorst's models involve an extension of standard Bayesian representations of belief, but he offers an argument that the extension is suitable for describing rational belief revision.

What do these models and arguments have to do with misinformation? There are undoubtedly cases where available evidence effectively settles some matter of fact; and yet agents nonetheless accept or share content that denies those facts. Such content is surely misinformation. But it is not at all clear that this is the generic situation; and in any case, when trying to assess what kind of content should count as misinformation, the assessors are generally embedded in epistemic communities themselves, with no external position from which to evaluate the information shared by others. One may be inclined to label as misinformation any content that tends to mislead some portion of that agent's community, in the sense of leading those members to update their beliefs in what the agent perceives to be the wrong direction. The foregoing models suggest that there are a broad range of contexts in which this can occur symmetrically, even for rational agents. In such cases, judgments concerning what count as misinformation amount to little more than expressions about one's

own background beliefs and/or higher order uncertainties.

5. Conclusion

The relatives of fake news—misinformation, disinformation, malinformation and the like—present serious challenges for reliable belief formation, collective action, and ultimately for democratic governance, insofar as democracy depends on a well-informed populace. They also raises important philosophical questions such as how to characterize (and recognize) misleading content and under what circumstances such content is most likely to influence public belief. In this brief review, we have discussed several aspects of the philosophical literature on misinformation. We have focused on questions related to how to define misinformation, on the role of epistemic vices in susceptibility to misinformation, on social epistemological aspects of the transmission and uptake of misinformation, and on rational polarization.

Acknowledgments

This work is supported by the National Science Foundation under grant number 1922424 "Consensus, Democracy, and the Public Understanding of Science". We are grateful to Edouard Machery and an anonymous reviewer for comments that improved the manuscript.

References

Axtell, G., 1997. Recent work on virtue epistemology. American Philosophical Quarterly 34, 1–26.

Battaly, H., 2008. Virtue epistemology. Philosophy Compass 3, 639–663.

Bok, S., 1978. Lying: Moral Choice in Public and Private Life. Random House, New York.

Bramson, A., Grim, P., Singer, D.J., Berger, W.J., Sack, G., Fisher, S., Flocken, C., Holman, B., 2017. Understanding polarization: Meanings, measures, and model evaluation. Philosophy of Science 84.

Brasher, N.M., Marsh, E.J., 2020. Judging truth. Annual Review of Psychology 71.

Carson, T.L., 2006. The definition of lying. Noûs 40, 284–306.

Cassam, Q., 2018a. Epistemic insouciance. Journal of Philosophical Research 43, 1–20.

Cassam, Q., 2018b. Vices of the Mind. Oxford University Press, Oxford.

Chisholm, R.M., Feehan, T.D., 1977. The intent to deceive. Journal of Philosophy 74, 143–159.

Cook, J., Lewandowsky, S., 2016. Rational irrationality: Modeling climate change belief polarization using bayesian networks. Topics in Cognitive Science 8, 160–179.

Dorst, K., 2023. Rational polarization. The Philosophical Review Forthcoming.

Fallis, D., 2009. What is lying. Journal of Philosophy 106, 29–56.

Fallis, D., 2015. What is disinformation? Library Trends 63, 401–426.

Fallis, D., 2016. Mis- and dis- information, in: Floridi, L. (Ed.), The Routledge Handbook of Philosophy of Information. Routledge, New York, pp. 332–346.

Fallis, D., Mathiesen, K., 2019. Fake news is counterfeit news. Inquiry .

Fetzer, J.H., 2004. Disinformation: The use of false information. Minds and Machines 2, 231–240.

Floridi, L., 1996. Brave.net.world: The internet as a disinformation superhighway? Electronic Library 14, 509—514.

Floridi, L., 2011. The Philosophy of Information. Oxford University Press, Oxford, UK.

Frankfurt, H., 2005. On Bullshit. Princeton University Press, Princeton, NJ.

Freeborn, D., 2023. Polarization and Factionalization for Agents with Multiple, Related Beliefs. Ph.D. thesis. University of California, Irvine.

Funkhauser, E., 2017. Beliefs as signals: A new function for belief. Philosophical Psychology 30, 809–831.

Garson, J., 2019. What Biological Functions Are and Why They Matter. Cambridge University Press, Cambridge, UK.

Gelfert, A., 2018. Fake news: A definition. Informal Logic 38, 84–117.

Genot, E.J., Olsson, E.J., 2021. The dissemination of scientific fake news: On the ranking of retracted articles in google, in: Bernecker, S., Flowerree, A.K., Grundmann, T. (Eds.), The Epistemology of Fake News. Oxford University Press, Oxford, UK, pp. 228–242.

Gilbert, D.T., Krull, D.S., Malone, P.S., 1990. Unbelieving the unbelievable: Some problems in the rejection of false information. Journal of Personality and Social Psychology 59, 601–613.

Gilbert, D.T., Tafarodi, R.W., Malone, P.S., 1993. You can't not believe everything you read. Journal of Personality and Social Psychology 65, 221–233.

Goldman, A.I., 2001. Experts: Why ones should you trust. Philosophy and Phenomenological Research 63, 85–110.

Greco, J., 1993. Virtues and vices of virtue epistemology. Canadian Journal of Philosophy 23, 413–432.

Greco, J., 2003. Virtues in epistemology, in: Moser, P.K. (Ed.), Oxford Handbook of Epistemology. Oxford University Press, Oxford, pp. 287–315.

Habgood-Coote, J., 2019. Stop talking about fake news! Inquiry 62, 1033–1065.

Harris, K.R., 2023. Beyond belief: On disinformation and manipulation. Erkenntnis Forthcoming.

Hasher, L., Goldstein, D., Toppino, T., 1977. Frequency and the conference of referential validity. Journal of Verbal Learning Verbal Behavior 16, 107–112.

Hegselmann, R., Krause, U., 2006. Truth and cognitive division of labour: First steps towards a computer aided social epistemology. Journal of Artificial Societies and Social Simulation 9, 10.

Holman, B., Bruner, J., 2017. Experimentation by industrial selection. Philosophy of Science 84, 1008–1019.

Jaster, R., Lanius, D., 2021. Speaking of fake news: Definitions and dimensions, in: Bernecker, S., Flowerree, A.K., Grundmann, T. (Eds.), The Epistemology of Fake News. Oxford University Press, Oxford, UK, pp. 19–45.

Jern, A., min K. Chang, K., Kemp, C., 2014. Belief polarization is not always irrational. Psychological Review 121, 206–224.

Koetke, J., Schumann, K., Porter, T., 2022. Intellectual humility predicts scrutiny of covid-19 misinformation. Social Psychological and Personality Science 13, 277–284.

Kvanig, J.L., 2010. Virtue epistemology, in: Bernecker, S., Pritchard, D. (Eds.), Routledge Companion to Epistemology. Routledge, New York, NY, pp. 199–207.

LaCroix, T., Geil, A., O'Connor, C., 2021. The dynamics of retraction in epistemic networks. Philosophy of Science 88, 415–438.

Levy, N., 2021. Arrogance and servility online: Humility is not the solution, in: Alfano, M., Lynch, M.P., Tanesini, A. (Eds.), The Routledge Handbook of Philosophy of Humility. Routledge, New York, pp. 470–483.

Levy, N., Alfano, M., 2020. Knowledge from vice: Deeply social epistemology. Mind 129, 887–915.

Lewandowsky, S., Smilie, L., Garcia, D., Hertwig, R., Weatherall, J., Egidy, S., Robertson, R.E., O'Connor, C., Kozyreva, A., Lorenz-Spreen, P., Blaschke, Y., Leiser, M., 2020. Technology and Democracy: Understanding the influence of online technologies on political behaviour and decision-making. Publications of the European Union.

Lynch, M.P., 2018. Arrogance, truth, and public discourse. Episteme 15, 283–296.

Lynch, M.P., 2019. Know-It-All Society: Truth and Arrogance in Political Culture. Liverwright, New York,

NY.

Mayo-Wilson, C., Zollman, K.J., Danks, D., 2011. The independence thesis: When individual and social epistemology diverge. Philosophy of Science 78, 653–677.

Mercier, H., Sperber, D., 2019. The Enigma of Reason. Harvard University Press, Cambridge, MA.

Meyer, M., Alfano, M., de Bruin, B., 2021a. The development and validation of the epistemic vice scale. Review of Philosophy and Psychology .

Meyer, M., Alfano, M., de Bruin, B., 2021b. Epistemic vice predicts acceptance of covid-19 misinformation. Episteme Forthcoming.

Millikan, R.G., 1984. Language, Thought, and other Biological Categories. MIT Press, Cambridge, MA.

Mohseni, A., O'Connor, C., Weatherall, J.O., 2024. The best paper you'll read today: Media biases and the public understanding of science. Philosophical Topics Forthcoming.

Montmarquet, J.A., 1993. Epistemic Virtue and Doxastic Responsibility. Rowman and Littlefield, Lanham, MD.

Neale, A.V., Dailey, R.K., Abrams, J., 2010. Analysis of citations to biomedical articles affected by scientific misconduct. Science and Engineering Ethics 16, 251–261.

Neander, K., 1991. Functions as selected effects. Philosophy of Science 58, 168–184.

Novaes, C.D., de Ridder, J., 2021. Is fake news old news?, in: Bernecker, S., Flowerree, A.K., Grundmann, T. (Eds.), The Epistemology of Fake News. Oxford University Press, Oxford, UK, pp. 156–179.

O'Connor, C., Weatherall, J.O., 2018. Scientific polarization. European Journal for Philosophy of Science 8, 855–875.

O'Connor, C., Weatherall, J.O., 2019. The Misinformation Age. Yale University Press, New Haven, CT.

O'Connor, C., Weatherall, J.O., 2020. Why false claims about covid-19 refuse to die. Nautilus .

Olin, L., Doris, J.M., 2014. Vicious minds. Philosophical Studies 168, 665–692.

Olsson, E.J., 2013. A bayesian simulation model of group deliberation and polarization, in: Zenker, F. (Ed.), Bayesian Argumentation. Springer, Dordrecht.

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. Bloomsbury, New York, NY.

O'Connor, C., Goldberg, S., Goldman, A., 2024. "social epistemology", in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. summer 2024 edition ed. URL: https://plato.stanford.edu/archives/win2021/entries/epistemology-social/.

Pearl, J., 1986. Fusion, propagation, and structuring in belief networks. Artificial Intelligence 29, 241–288.

Porter, T., Elnakouri, A., Meyers, E.A., Shibayama, T., Jayawickreme, E., Grossmann, I., 2022. Predictors and consequences of intellectual humility. Nature Reviews Psychology 1, 524–536.

Priest, M., 2021. How vice can motivate distrust in elites and trust in fake news, in: Bernecker, S., Flowerree, A.K., Grundmann, T. (Eds.), The Epistemology of Fake News. Oxford University Press, Oxford, UK, pp. 180–205.

Pritchard, D., 2021. Good news, bad news, fake news, in: Bernecker, S., Flowerree, A.K., Grundmann, T. (Eds.), The Epistemology of Fake News. Oxford University Press, Oxford, UK, pp. 46–67.

Rini, R., 2017. Fake news and partisan epistemology. Kennedy Institute of Ethics Journal 27, 43–64.

Roberts, R.C., Wood, W.J., 2007. Intellectual Virtues: An Essay in Regulative Epistemology. Oxford University Press, Oxford.

Saul, J.M., 2012. Lying, Misleading, and What is Said: An Exploration in Philosophy of Language and in Ethics. Oxford University Press, Oxford, UK.

Schneider, J., Ye, D., Hill, A.M., Whitehorn, A.S., 2020. Continued post-retraction citation of a fraudulent clinical trial report, 11 years after it was retracted for falsifying data. Scientometrics 125, 2877–2913.

Simion, M., 2023. Knowledge and disinformation. Episteme Forthcoming.

Singer, D.J., Bramson, A., Grim, P., Holman, B., Jung, J., Kovaka, K., Ranginani, A., Berger, W.J., 2019. Rational social and political polarization. Philosophical Studies 176, 2243–2267.

Skyrms, B., 2010. Signals: Evolution, Learning, and Information. Oxford University Press, Oxford, UK.

Smart, P.R., 2018. Mandevillian intelligence. Synthese 195, 4169–4200.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., Wilson, D., 2010. Epistemic vigilance. Mind & Language 25, 359–393.

Turri, J., Alfano, M., Greco, J., 2021. Virtue epistemology, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. winter 2021 ed. URL: https://plato.stanford.edu/archives/win2021/entries/epistemology-virtue/.

Vasilyeva, N., Smith, K.M., Barr, K., Kiper, J., Stich, S., Machery, E., Barrett, H.C., 2021. Evaluating information and misinformation during the covid-19 pandemic: Evidence for epistemic vigilance. Proceedings of the Annual Meeting of the Cognitive Science Society 43, 583–589.

Wardle, C., Derakhshan, H., 2017. Information Disorder: Toward an interdisciplinary framework for research and policy making. Technical Report. Council of Europe. URL: https://www.coe.int/en/web/freedom-expression/information-disorder.

Weatherall, J.O., O'Connor, C., Bruner, J.P., 2018. How to beat science and influence people: Policymakers

and propaganda in epistemic networks. British Journal for the Philosophy of Science 71, 1157–1186.

Weatherall, J.O., O'Connor, C., 2021a. Conformity in scientific networks. Synthese 198, 7257–7278.

Weatherall, J.O., O'Connor, C., 2021b. Endogenous epistemic factionalization. Synthese 198, 6179—-6200.

West, J.D., Bergstrom, C.T., 2020. Misinformation in and about science. Proceedings of the National Academy of Science 118, e1912444117.

Wright, S., 2021. The virtue of epistemic trustworthiness and re-posting on social media, in: Bernecker, S., Flowerree, A.K., Grundmann, T. (Eds.), The Epistemology of Fake News. Oxford University Press, Oxford, UK, pp. 245–264.

Zagzebski, L.T., 1996. Virtues of the Mind: An Inquiry Into the Nature of Virtue and the Ethical Foundations of Knowledge. Cambridge University Press, Cambridge.