

Narrative counterspeech

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open access

Lepoutre, M. ORCID: <https://orcid.org/0000-0001-7573-8585>
(2022) Narrative counterspeech. *Political Studies*. ISSN 0032-3217 doi: <https://doi.org/10.1177/00323217221129253>
Available at <https://centaur.reading.ac.uk/107288/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1177/00323217221129253>

Publisher: Sage

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Narrative Counterspeech

Maxime Lepoutre 

Political Studies

1–20

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/00323217221129253

journals.sagepub.com/home/psx

Abstract

The proliferation of conspiracy theories poses a significant threat to democratic decision-making. To counter this threat, many political theorists advocate countering conspiracy theories with ‘more speech’ (or ‘counterspeech’). Yet conspiracy theories are notoriously resistant to counterspeech. This article aims to conceptualise and defend a novel form of counterspeech – narrative counterspeech – that is singularly well-placed to overcome this resistance. My argument proceeds in three steps. First, I argue that conspiracy theories pose a special problem for counterspeech for three interconnected reasons relating to salience, emotion and internal coherence. Drawing on recent work in social epistemology, philosophy of emotion and cognitive science, I then demonstrate that narrative forms of counterspeech constitute an apt response to this diagnosis. Finally, I forestall two objections: the first questions the likely effectiveness of narrative counterspeech; the second insists that, even if it were effective, it would remain unacceptably manipulative. Neither objection, I contend, is ultimately compelling.

Keywords

conspiracy theories, disinformation, counterspeech, fact-checking, narrative, manipulation, social epistemology, philosophy of emotion, political theory

Accepted: 12 September 2022

Introduction

In March 2020, as the first wave of the coronavirus pandemic was spreading rapidly across the world, a conspiracy theory followed closely in its wake. COVID-19, the theory alleged, was nothing else than a Big Pharma scam engineered to stimulate demand for vaccines (EU vs Disinformation, 2020). As vaccination campaigns subsequently accelerated, so too did the conspiracy theories. COVID-19 vaccines, some claimed, were designed to alter human DNA. Others worried – and continue to worry – that Bill Gates was using these vaccines to implant microchips in recipients (Carmichael and Goodman, 2020).

COVID-19 vaccines are not an isolated case. Conspiracy theories more generally have flourished across the world in recent years – aided, often, by social media platforms that allow them to be shared widely and rapidly in more or less insulated arenas (Uscinski,

Department of Politics and International Relations, University of Reading, Reading, UK

Corresponding author:

Maxime Lepoutre, Department of Politics and International Relations, University of Reading, Shinfield Road, Reading RG6 6EL, UK.

Email: m.c.lepoutre@reading.ac.uk

2020; Walter and Drochon, 2022). MMR (measles, mumps and rubella) vaccines are said to be associated with autism (Boseley, 2019). Climate change is portrayed as nothing more than a hoax (Worland, 2019). And, perhaps most infamously, the US Democratic Party is believed by many Americans to be run by a cabal of Satanic child sex-traffickers (Nagesh, 2021).

This proliferation is dangerous. Exposure to conspiracy theories can meaningfully influence our beliefs and intentions (e.g. Jolley and Douglas, 2014; Jolley et al., 2020; Romer and Jamieson, 2020; Uscinski, 2020: 6–10). This, in turn, may have devastating consequences. For instance, insofar as conspiracy theories fuel resistance to COVID-19 vaccines, they endanger the unvaccinated and prolong a global health crisis. Likewise, by fuelling climate scepticism, conspiracy theories undercut public support for policies aimed at mitigating the climate emergency.¹

This danger calls for an urgent response. One proposal here might be for the state, acting directly or indirectly, to censor dangerous conspiracy theories. For example, some political philosophers have suggested that the state could legally require social media companies to remove posts promoting such conspiracy theories (Brown, in press; Howard, 2021a). But censorship remains controversial for a number of reasons. Some worries are principled – for instance, the worry that censorship would violate freedom of expression. Others are more practical – for instance, the worry that the state (or social media companies) cannot be trusted to implement censorship in an accurate and fair manner (Lepoutre, 2021: chs. 3–4).

In part because of these difficulties, many political philosophers emphasise a different response to conspiracy theories: namely, the use of ‘more speech’ (or ‘counterspeech’). Russell Muirhead Nancy Rosenblum (2020: 143) clearly illustrate this position. Their influential diagnosis of ‘the new conspiracism’ recommends, first and foremost, ‘speaking truth to conspiracy’. Quassim Cassam (2019: 105) concludes his own examination of conspiracy theories with a similar recommendation. ‘[W]hen faced with theories that distort the facts’, he declares, ‘the motto should be: rebut, rebut, rebut’.²

Counterspeech can be deployed instead of censorship, and this is often what advocates of counterspeech appear to be recommending. But it can also be deployed in conjunction with censorship – say, out of recognition of the fact that, in practice, prohibiting dangerous conspiracy theories is highly unlikely to eliminate them altogether. Jeffrey Howard, for example, expresses openness to deploying legal restrictions *and* counterspeech in response to dangerous communications (Howard, 2021a, 2021b). Either way, the fundamental point remains that counterspeech (whether it replaces or complements alternative strategies) has emerged as one of the most popular – if not *the* most popular – response to conspiracy theories.

Yet this popular proposal, too, faces enormous difficulties. Countering conspiracy theories with more speech is notoriously challenging – even more challenging, as I will argue shortly, than countering other forms of disinformation with more speech. Hence, when applied to conspiracy theories, counterspeech runs the risk of being ineffective or even counterproductive.

The present article aims to tackle this problem in two ways: first, by offering a clearer account of why conspiracy theories are unusually resistant to counterspeech; and second – and more positively – by conceptualising and defending a novel form of counterspeech (‘narrative counterspeech’) that is distinctively well-placed to overcome this resistance.

My argument will proceed as follows. First, I argue that conspiracy theories pose a special challenge to counterspeech for three interrelated reasons: their enduring salience, their affective dimension and their internal coherence. Building on this diagnosis, together with recent work in social epistemology, I then recommend the use of narrative – or storytelling – to counteract conspiracy theories. The final two sections forestall two significant worries: that, in practice, narrative counterspeech has proven to be causally ineffective; and that deploying narrative counterspeech is manipulative. Neither, I will suggest, constitutes a decisive objection to narrative counterspeech.

Before proceeding, several clarifications are needed. First, I am not arguing that narrative counterspeech is *only* useful in response to conspiracy theories. Rather, I focus on conspiracy theories because these are especially difficult to counter – and so, the benefits of narrative counterspeech are most salient in this domain.

Second, I am not suggesting that any kind of narrative will do. Narrative structure and style vary significantly. As we will see my argument for thinking that narrative is a promising tool for countering conspiracy theories applies more strongly to some kinds of narratives than to others – and thus, it supports the deployment of *specific types* of narrative.

The third clarification concerns the strength of the claim that I am making – and, in particular, its relationship to empirical evidence. Whether or not (a particular form of) narrative constitutes an effective response to conspiracy theories cannot definitively be settled by political philosophy alone. It is, at bottom, an empirical matter, which depends on social scientific inquiry. But social scientific inquiry itself requires guidance from theoretical hypotheses. In this context, my aim is to articulate such a hypothesis. As I will demonstrate, not only is this hypothesis consistent with existing empirical evidence on counterspeech, but it also helps identify where existence evidence is lacking, and thus provides guidance for future empirical research.

The final point has to do with the novelty of the recommendation I am putting forward. Theories of counterspeech have typically not attended to the potential of narrative. Yet this observation should not be overstated: the use of narrative to counter disinformation is not wholly unprecedented. Indeed, narrative has recently attracted support in the domain of science communication (e.g. Dahlstrom, 2014; ElShafie, 2018). While I am sympathetic to these existing proposals, they remain limited in two interconnected ways. From a diagnostic perspective, they generally do not adequately explain *why* conspiracy theories are so resistant to non-narrative counterspeech – indeed, the vast majority of these proposals are not specifically concerned with conspiracy theories.³ This, in turn, has implications for guidance: it translates into a limited understanding of why narrative counterspeech is so promising, of where it is most useful and of what specific forms it should take.

This is precisely what my argument aims to remedy. By offering a sustained examination of conspiracy theories, it aims to provide a clearer appreciation of why narrative counterspeech is needed; and hence, it aims to provide more fine-grained guidance regarding where and how it should be deployed.

Why Conspiracy Theories Resist Counterspeech

Broadly understood, a conspiracy theory is ‘an effort to explain some event or practice by reference to the machinations of powerful people, who attempt to conceal their role (at least until their aims are accomplished)’ (Sunstein and Vermeule, 2009: 205).

As it stands, this definition is extremely broad. It includes theories that are true; theories that are supported by the best available evidence; and theories that either have no bearing on matters of public concern, or that have a positive impact on such matters. For example, it includes the (accurate) theory that the Nixon administration attempted to break into, and cover up its break-in of, the Democratic National Congress headquarters. This is arguably *too* broad for my purposes. Theories that are true, epistemically justified and/or benign do not urgently call for counterspeech.

As a result, following Cass Sunstein and Adrian Vermeule (2009: 205–211), I will be focusing on a subset of conspiracy theories: those (such as the theory that COVID-19 vaccines are a Big Pharma scam; or that climate change is a hoax) that are false, unsupported by the best existing evidence and at risk of generating significant public harms. These theories *do* urgently call for counterspeech. But they also tend to resist counterspeech.

One obvious reason for this, which is not specific to conspiracy theories, has to do with access. In a fragmented society, where members of different groups select different sources of news, and have comparatively little contact with one another, simply *exposing* potential conspiracy sympathisers to counterspeech may be challenging (Mutz, 2006; Sunstein, 2017; Talisse, 2019: chs. 2–3). Accordingly, any attempt at counterspeech (whether it takes a narrative or non-narrative form) must make strenuous efforts to reach members of different groups.⁴

In what follows, however, I will focus on a deeper problem: even when counterspeech does reach actual or potential conspiracy sympathisers, conspiracy theories have distinctive properties in virtue of which they tend to resist its corrective influence. There are at least three reasons for this resistance.

Salience

Conspiracy theories tell us about powerful agents who scheme, in secret, to achieve (typically sinister) goals. These tales, like the Hollywood thrillers they are modelled on, are generally captivating. The theory that COVID-19 is a scam orchestrated by pharmaceutical companies may be absurd – but there is no denying that it is exciting. It is difficult to hear this theory, without wanting to find out more.

Why does it matter that conspiracy theories are entertaining in this way? It matters because it means that conspiracy theories are inherently disposed to capture and retain our attention: they easily become salient to us, and once this has happened, it is difficult to make them less salient (Cassam, 2019: 58; Muirhead and Rosenblum, 2020: 38; van Prooijen et al., 2021: 2). This salience, in turn, is problematic in several respects.

One reason concerns fluency. The more conspiracy theories command our attention, the more familiar – or ‘fluent’ – they seem to us. This matters, because fluency has an impact on belief. Cognitive scientists have shown that, the more familiar a claim feels to us, the more likely we are to believe it (see, e.g. Lewandowsky et al., 2012, for an overview).

There is a second reason why the salience of conspiracy theories might induce us to believe them. Given that attention is a limited resource, we generally strive to pay attention only to things that are *worth* paying attention to. Accordingly, the fact that conspiracy theories are salient in public debate – that people pay attention to them – might seem to suggest that these theories are worth paying attention to (Cassam, 2019: ch. 4). And, in

particular, it might seem to imply that these theories have greater evidential support than they actually do.

But the enduring salience of conspiracy theories is a problem even when it does not lead us to believe them. Time spent inspecting an alleged COVID-19 plot is time not spent discussing, and attending to, the real human and financial cost of the pandemic. The fact that conspiracy theories absorb our attention is therefore a problem *even when* we ultimately reject them, because it distracts us from real and pressing political issues.

The point so far is that conspiracy theories are designed to become and remain salient – and this salience facilitates their harmful outcomes. This phenomenon, in turn, is doubly problematic for counterspeech. For one thing, it means that, when competing with conspiracy theories, counterspeech may struggle to gain our attention. Consider the standard way of countering conspiracy theories with more speech: fact-checking. Fact-checking typically involves identifying a piece of disinformation, saying that it is false, and, often, methodically explaining why it is false. This intervention is usually far less exciting than the conspiracy theory it is responding to. It might therefore struggle to draw and maintain the attention of its intended targets.

But the deeper problem is that, even when counterspeech does capture our attention, it risks reinforcing the salience of the conspiracy theory it is responding to (see, e.g. Saul, 2021: 147–148; Sunstein and Vermeule, 2009: 222).⁵ Declaring that ‘the COVID-19 pandemic was not orchestrated by pharmaceutical companies’, as Reuters (2021) recently did, is likely to amplify, by repeating it, the theory that it *was* so orchestrated. This amplification is problematic for the reasons canvassed above. By reinforcing the salience of a conspiracy theory, counterspeech risks (1) making the theory in question more fluent to listeners, (2) implicating that it is worthy of consideration and (3) distracting us from real and pressing issues.⁶

Emotion

The second reason conspiracy theories tend to resist counterspeech relates to emotion: support for conspiracy theories tends to be emotionally charged.

To begin with, emotions help explain why many people come to believe conspiracy theories in the first place. In other words, conspiracy theories frequently seem appealing because they satisfy, or resonate with, people’s pre-existing emotions (Cassam, 2019: 59–60; Muirhead and Rosenblum, 2020: 38–40; Sunstein and Vermeule, 2009: 213; van Prooijen et al., 2021: 8).

For example, there is evidence that disempowered groups’ anger or resentment at elites partly explains why some members of those groups were willing to believe that powerful pharmaceutical companies would orchestrate a pandemic for their own profit (Tonkovic et al., 2021). A similar observation applies to partisan conspiracy theories: in contexts marked by affective polarisation, Republicans are more willing to believe that Democrats are involved in child sex-trafficking, and Democrats are more willing to believe that Donald Trump’s campaign colluded with Russia, in part because this resonates with pre-existing feelings of mutual dislike or even hatred (see Uscinski, 2020: 84–86).⁷

In turn, believing conspiracy theories often sustains or amplifies these pre-existing emotions (Albertson and Guiler, 2020). Believing that pharmaceutical companies are running a global scam undoubtedly fuels the anger or resentment that may have made this theory seem appealing to begin with. Likewise, believing that one’s partisan opponents

are conspiring to commit abhorrent crimes seems likely to increase one's existing feelings of dislike or hatred towards them.⁸

This emotional dimension seems *prima facie* problematic for counterspeech – or, at least, for counterspeech as it is commonly practised. Fact-checking provides information that contradicts the targeted conspiracy theories. But one might worry that this does little to counter their emotional appeal. To reiterate, many people support conspiracy theories, not just because they have been exposed to misleading information, but because of how conspiracy theories make them feel. These theories often resonate with, and intensify, powerful emotions.

Note that the present obstacle is closely related to the salience problem. Philosophers of emotion have widely argued that emotions have a cognitive dimension. That is, emotions modify the way we represent the world. More specifically, it is widely held, among philosophers of emotion, that emotions are *sources of salience*. Put differently, emotions exert a strong influence on our patterns of attention, and thereby play a significant role in determining what appears salient to us, and what does not (Brady, 2013: 61–62; Deonna and Teroni, 2012: 121–122; Elgin, 2008: 44–45). Fear, for example, tends to draw our attention to potential sources of danger. Anger makes us fixate on perceived wrongs or injustices. Grief makes loss more salient to us. And so on.

What this indicates is that conspiracy theories' tendency to absorb our attention is not independent of their entanglement with emotion. On the contrary, the enduring salience of conspiracy theories partly *results* from their emotional potency. In light of the fact that emotions are sources of salience, it is plausible to think that the emotions triggered or reinforced by conspiracy theories play a part in sustaining the salience of those theories. Thus, reducing the salience of conspiracy theories also requires countering their emotional appeal.

Coherence

The third problem concerns coherence. Conspiracy theories are typically not isolated claims. Rather, as *theories*, they normally involve a system of claims that cohere or hang together more or less closely (Cassam, 2019: 44; Douglas et al., 2019: 7; Enders et al., 2021: 267).⁹

This coherence poses two challenges for counterspeech. The first is that, as cognitive scientists have shown, we generally have a strong preference for cognitive coherence: we want our cognitive commitments to hang together in a mutually supportive way. As a result, a simple factual rebuttal of a conspiracist claim ('Vaccines *are* safe') is unlikely to be accepted by someone who already adheres to the relevant conspiracy theory (for an overview of relevant empirical evidence, see Lewandowsky et al., 2012: 112–113, 117). This is because accepting the rebuttal would leave them with a system of beliefs that sit awkwardly together (e.g. a belief in the safety of vaccines that jars with conspiratorial beliefs about why, how and by whom vaccines were created).

But the problem goes deeper than this. The problem is not simply that we tend to prefer a coherent system of beliefs over an incoherent one, even if that coherence depends upon falsehoods. It is, in addition, that the coherence of conspiracy theories is *robust*: it is difficult to disrupt the coherence of conspiracy theories to begin with. This is due to conspiracy theories' 'self-sealing' quality. Conspiracy theories hold that powerful actors are acting covertly, and trying to conceal their actions. They therefore *predict* the existence of seemingly contradictory evidence (on this self-sealing quality of conspiracy theories, see,

e.g. Cassam, 2019: 97; Muirhead and Rosenblum, 2020: 141, 143, 157; Napolitano, 2021: 2; Sunstein and Vermeule, 2009: 207). For the ardent conspiracist, contradictory evidence (such as that provided by fact-checkers) is yet another sign that powerful actors are covering up their footsteps.¹⁰

Thus, the coherence of conspiracy theories makes them resistant to counterspeech. Counterevidence supplied by counterspeech can often be reconciled, or made to cohere with, the conspiracy theory. And even if it cannot, we generally have a preference for maintaining a coherent belief system – and so, we may reject the correction anyway.

Let us take stock. I have argued that, for reasons relating to salience, emotion and coherence, conspiracy theories are unusually difficult to counteract with ‘more speech’. This diagnosis clarifies what counterspeech would need to do, in order to constitute an effective response to conspiracy theories. It must be able to capture and retain attention. It must engage the emotions of listeners. And it must have the capacity to unsettle a robustly coherent set of cognitive commitments. In what follows, I will show, drawing on recent work in the epistemology of narrative, that ‘narrative counterspeech’ is especially well-placed to meet these demands.

Narrative Counterspeech

The Idea of Narrative

What is a narrative? ‘Narratives’, Rachel Fraser (2021: 4027) explains, are ‘just those texts and utterances which have the form of a *story*’.

Now, what exactly constitutes a story is itself disputed. But it is nevertheless widely agreed that stories typically contain the following ‘core’ characteristics. Stories characteristically involve a sequence of events that are causally structured (by contrast with annals, which simply enumerate temporally ordered events). This causally structured sequence of events generally revolves around one or more protagonist(s), or principal characters, which may or may not be human. And it usually has a familiar arc: after setting the stage, the typical story is catalysed by a problem or obstacle, which puts something at risk, and which the protagonist(s) must confront – leading, usually, to a resolution (for discussion of these ‘core’ features of narratives, see, e.g. ElShafie, 2018: 1217–1219; Fraser, 2021: 4027; Mandler and Johnson, 1977: 114–115).

Besides these ‘core’ characteristics, stories often – though by no means always – involve a number of distinctive stylistic features. For example, many stories make use of first-personal pronouns (‘I’). Stories also often involve high levels of complexity and detail relating to the internal life of characters (e.g. their emotions and perceptions). Moreover, stories commonly employ figurative language (e.g. metaphors, similes) as well as loaded words (words, like ‘cop’, that have evaluations built into them) (Fraser, 2021: 4035n22, 4039–4040).

These stylistic features are of course not meant to be exhaustive. Nor, to reiterate, are they *essential* features of narratives. Yet, they are commonly associated with narratives. And, as we will see presently, they help explain why narratives – and some narratives more than others – are promising tools for tackling conspiracy theories.

Countering Conspiracy With Narrative

Narratives are well-placed to counter conspiracy theories in a way that maintains the coherence of its audience’s cognitive systems, engages their emotions and captures their

attention. In what follows, I will show how narratives' coherence, structure and style contribute to making this the case.

Narrative Coherence. One of the key problems of conspiracy theories, recall, is that they involve a more or less coherent system of claims. Indeed, a conspiracy theory might involve a whole account of how different events and actors are causally interlinked. This is a problem, in part, because we have a strong preference for cognitive coherence. Accordingly, if we must choose between accepting counterspeech and preserving the coherence of our cognitive system, many are likely to opt for the latter option.

Counterspeech that takes a narrative form seems well positioned to avoid this concern. As we have just seen, narrative characteristically involves, not just a single claim or proposition, but rather an integrated account of how different actors and events are causally interlinked (see, e.g. Fraser, 2021: 4026). Thus, narrative counterspeech offers its audience, not simply a bare rejection of a conspiracist claim ('Vaccines are safe'), but rather a coherent cognitive system that aims to replace the conspiracist system. And so, its audience can accept it without giving up cognitive coherence.

Consider a brief illustration. In 2021, Forbes published an extended profile of the scientist Ian MacLachlan, titled: 'Covid's Forgotten Hero: The Untold Story of the Scientist Whose Breakthrough Made the Vaccines Possible' (Vardi, 2021). The story does not simply claim, against COVID-19 conspiracy theories, that vaccines are safe. Rather, it tells an elaborate story that purports to reveal how vaccines were developed, by whom, what their motivations were and how this process led to crucial innovations that ensured their safety. Better yet, the story explicitly accommodates important components of COVID-19 conspiracies – for example, distrust of elites, and of pharmaceutical companies in particular – by alleging that pharmaceutical companies appropriated MacLachlan's work without acknowledging it. Thus, the story connects with, and strives to do justice to, some of the core beliefs and concerns underpinning support for COVID-19 conspiracy theories.

The upshot is that readers can accept the story's claims without being left with an incoherent system of beliefs and attitudes. For one thing, the claim that vaccines are safe is explained by an account of how, why and by whom they were made. For another, the account itself coheres with an attitude of elite distrust to which many conspiracists are already sympathetic.

The fact that narrative counterspeech offers listeners a way of retaining cognitive coherence is important. It shows that narrative counterspeech is capable of overcoming at least one of the key hurdles posed by conspiracy theories. And thus it is a necessary step in establishing that narrative could be effective in countering conspiracy theories.

But even if preserving coherence is a *necessary* ingredient of successful counterspeech – that is, of counterspeech that is likely to be accepted – it is not sufficient. The self-sealing quality of conspiracy theories means that they predict (and so, are able to accommodate) the existence of such countervailing narratives. So considerations of coherence alone cannot explain why listeners should *prefer* the story supplied by narrative counterspeech to the conspiracy theory it aims to counter.

Taken alone, considerations of coherence are also insufficient to explain why we should prefer narrative counterspeech to non-narrative counterspeech. A very rudimentary fact-check might simply involve negating a false claim ('Vaccines are safe'). But a

more sophisticated fact-check could easily elaborate on this correction by explaining how the vaccine came into being, who created it and why this origin makes it safe (for discussion of sophisticated fact-checks, see Cassam, 2019: 102–104). So, non-narrative counterspeech can (and often does) provide a cohesive system of claims.

We are therefore left with the following questions: Why would listeners prefer the coherent system offered by narrative counterspeech to that offered by the conspiracy theory being countered? And, relatedly, what is it about the coherent system offered by narrative counterspeech that makes it more enticing than that supplied by a sophisticated fact-check?

To answer these questions, we need to return to the other two problems raised by conspiracy theories. Conspiracy theories are attractive, not merely due to their internal coherence, but also in virtue of their capacity to engage emotion and capture attention.

The problem with sophisticated fact-checks is that they fare poorly on both counts. Take, for instance, Gerald Posner's *Case Closed*, a 500-page takedown of the conspiracy theories that swirled around John F. Kennedy's assassination. Cassam (2019) holds this up as an exemplar of counterspeech. But however methodical it is – in fact, arguably *because* of how methodical it is – this takedown will struggle to retain the attention, and to engage the emotions, of all but an infinitesimally small minority of political aficionados.¹¹

Herein lies the comparative advantage of narrative counterspeech. The way it draws its targets away from conspiracy theories has more to do with conversion than with the force of better argument. The reason someone abandons or avoids a conspiracy theory, and instead embraces the account delivered by narrative counterspeech, is not necessarily that the conspiracy in question has been exposed as incoherent or incapable of being reconciled with existing evidence. (As we have seen, conspiracy theories' self-sealing quality makes it unclear to what extent this is even possible.) The reason, instead, is that the listener has been told a story that is seductive – it is captivating and emotionally engaging. Indeed, it is widely held that narrative constitutes a powerful tool for capturing attention and enlisting the emotions (see, e.g. ElShafie, 2018: 1214–1215; Fraser, 2021: 4035–4037; Prescott-Couch, manuscript).

This, however, cannot simply be assumed. To do so would come close to assuming that narrative counterspeech is an effective antidote for conspiracy theories. And this, of course, is what I aim to demonstrate. In the remainder of this section, I will therefore investigate *why* narrative is distinctively well-placed to capture attention and enlist the emotions. As mentioned earlier, these two issues are interconnected in virtue of the fact that emotions govern our attentional patterns. So, instead of considering them separately, I will proceed by examining the characteristics of narratives that give rise to these interconnected effects.

Narrative Structure. As explained earlier, narratives have a characteristic structure, which is sometimes referred to as the 'story schema'. Narratives characteristically involve a series of causally interconnected events, featuring at least one protagonist, who confronts a meaningful obstacle or problem, leading to some form of resolution.

This schema matters for three reasons. First, and most obviously, its content involves important *stakes*. The story schema involves an obstacle, where that obstacle places something meaningful at risk. This risk, and the events revolving around it, help explain why stories tend to be emotionally engaging and attention-grabbing. The threat to

something meaningful may feel frightening, actions taken to overcome this threat may be awe-inspiring, a happy resolution may be uplifting and so on (Dahlstrom, 2014: 13614–13615). And this has implications for attention: as we have seen, to experience emotions such as fear, awe or joy *just is*, in part, to have one's attention absorbed and directed in a certain way.

Second, the standard narrative structure facilitates *first-personal simulation*. Information can be encoded in different representational formats. For example, the layout of a town might be encoded propositionally (e.g. as a list of claims specifying the coordinates of each landmark), spatially (e.g. as a map) or through simulation (e.g. by imagining, from a first-personal perspective, how one would get from one landmark to another) (Fraser, 2021: 4033–4034).

How does this relate to narrative structure? In her recent analysis of the epistemology of narrative, Fraser argues that narratives have the power to cue representational formats. That is, they influence which representational format we adopt when we encode information. In particular, the standard story structure has a tendency to cue a simulationist format (Fraser, 2021: 4036; see also Dahlstrom, 2014: 13614–13615; ElShafie, 2018: 1217).¹² This is, in part, because stories tend to revolve around a relatable protagonist. The protagonist is relatable in several ways. They are usually human or anthropomorphised. They, like everyone, have personal goals that are threatened by obstacles. Moreover, ElShafie (2018: 1217) notes that protagonists often (though not always) have likeable yet flawed personalities.

All of this is visible, for instance, in the 'Untold Story' of Ian MacLachlan. The story tells the reader about the development of COVID-19 vaccines through the lens of this one person's perseverance through setbacks – some of them self-inflicted – that left him 'exhausted and demoralised' (Vardi, 2021). So, the story follows a more or less relatable protagonist – a human being whose interests, struggles and victories we can identify with – and thereby invites us to imagine the story's events through his eyes. Nor is this an unusual example. A similar observation notably applies in the context of climate communication. Personal stories about ordinary people whose lives have been disrupted by climate change – for instance, about individuals at risk of being displaced by sea-level rises, or about firefighters desperately trying to bring climate-driven wildfires under control – invite us to imaginatively put ourselves into their protagonists' shoes.

This matters because first-personal simulation tends to facilitate stronger emotional experience – and by implication, it helps capture our attention more effectively. Projecting ourselves into someone else's perspective helps us resonate, or feel along with, them (Camp, 2017: 77; Dahlstrom, 2014: 13616; ElShafie, 2018: 1215; Fraser, 2021: 4035). Thus, it contributes to making us experience events in a more visceral and captivating way.

Consider, for example, Shelby and Ernst's (2013: 1798) narrative of a vaccine-sceptical father who witnesses his unvaccinated daughter suffering from tetanus. This narrative invites us to see the situation through the father's eyes. It fixes our attention on the harrowing scene, and invites us, almost irresistibly, to share his feelings of terror, powerlessness and guilt. By comparison, a sophisticated fact-check could familiarise us with statistics regarding the risks tetanus poses for children. But it would struggle to prompt such an engrossing experience.

Up to this point, I have argued that the story structure exerts a meaningful influence on our attention and emotions because it involves meaningful *stakes*, and because, via its impact on representational format, it tends to invite *first-personal simulation*. Yet

narrative structure also matters because it is *familiar*. Cognitive scientists have long argued that the story schema is built into our cognition (Mandler and Johnson, 1977; see also ElShafie, 2018: 1215; Fraser, 2021: 4042 for discussion). In other words, we are psychologically disposed to process and organise information in a way that involves a story structure.

This has significant implications for our patterns of attention (and by extension, for the emotional responses associated with these patterns). As Fraser (2021: 4045) explains, ‘story schemata funnel attention, most obviously towards the events positioned as central within the schema’. Put differently, the familiarity of the story structure means that our attention fixes with relative ease on ‘core’ features of the story: the protagonist, the problem, the resolution and so on. And there is evidence that this effect endures over time. The familiarity of the story structure helps us *recall* these core elements (Fraser, 2021: 4042). Thus, narrative forms of counterspeech may be able to influence our patterns of attention – and with them, our emotions – in a lasting way.

Narrative Style. But narrative structure is not the only reason why narrative counterspeech seems well-placed to counteract conspiracy theories’ effect on emotion and attention. As we saw previously, narratives commonly involve stylistic features which, though they are not present in all narratives, are nonetheless more prevalent in narratives than in non-narrative speech. These stylistic features are significant because they, too, can have a meaningful impact on attention and emotion.

Some stylistic features achieve this impact by facilitating first-personal simulation. Consider the narrative deployed by the Center for Disease Control and Prevention (2017) to raise awareness of cervical cancer. Two stylistic features stand out. First, like many stories, it is told in the first-person, using first-person pronouns (‘Hi, my name is Jackie. I am a mom to a couple of wonderful kinds and I’m a cervical cancer survivor. This is my story’.) Second, it involves a rich level of detail about its protagonist’s internal life (‘Every time the phone rings, I’m *terrified* it’s my oncologist calling [. . .] the thought of cancer returning is always in the back of my mind’).

These two stylistic features tend to cue a specific representational format: they invite us, as readers or listeners, to imagine the story from the protagonist’s perspective. And this matters because, as argued earlier, first-personal simulation has a distinctive capacity to provoke an emotionally charged and engrossing experience.¹³

But this is not the only way that the stylistic features of narratives can help capture our attention or elicit our emotions. Fraser argues that, besides being able to cue a specific representational format, narratives can also cue certain *characterisations*. Roughly, the idea is that stylistic devices that are commonplace in narratives (such as figurative language or loaded words) have the capacity to make us see something through specific interpretive lenses. Those lenses, in turn, tend to focus our attention and evoke our emotions in powerful ways (Fraser, 2021: 4038–4042).

This is especially visible with metaphors. Metaphors tend to make us see something *as something else*. This affects what properties of that thing appear salient or prominent to us – and, often as a result, how we feel about it. And, as Elisabeth Camp’s (2017: 47–54) philosophical analysis of metaphor shows, this effect is difficult to resist (see also Fraser, 2021: 4039; Moran, 1989: 90–91). Whether or not we imagine the subject of the metaphor as something else – and thus, whether or not our attention and emotions are engaged in this way – is often outside of our volitional control. For instance, describing COVID-19 vaccines as ‘shields’ invites us, almost irresistibly, to see ourselves as under attack

from an external threat. This metaphor therefore tends to elicit fear, and to focus our attention on this perceived threat.

The present section suggests that narrative style can help cue a specific representational format, or a specific characterisation, in ways that exert a significant influence on our attention and emotions. But this argument requires qualification in two respects. First, I am not suggesting that these stylistic devices are unique to narrative. Fact-checks, for instance, do occasionally employ metaphors. Nevertheless, these stylistic devices are considerably less common in non-narrative speech. Extensive uses of the first-person, loaded words and metaphors would seem out of place in a fact-check – and so too, typically, would a detailed account of someone’s internal life. So it remains the case that these stylistic devices, and their benefits for countering conspiracy theories, are more strongly associated with narrative than non-narrative counterspeech.

Second, it is important to emphasise, once more, that not all narratives include these stylistic devices. Some narratives are not told in the first-person. Some tell us relatively little about what characters think or feel. And some eschew figurative language. I do not mean to deny this. Rather, the point is that, for the reasons outlined above, narratives that do include these characteristic stylistic devices may be more effective at capturing attention and eliciting emotional responses.

Let us take a step back. I have argued that, due to their coherence, structure and style, narrative forms of counterspeech are comparatively well-placed to counteract the distinctive challenges presented by conspiracy theories. But, as we have just seen, the foregoing argument goes further than this initial conclusion. In addition, it also supplies guidance as to which *kinds* of narratives are especially likely to be effective. Notably, we have seen that the use of the first-person, of metaphors and figurative language, of detailed accounts of characters’ internal lives and the deployment of a relatable protagonist are *prima facie* likely to make for more effective narrative counterspeech.

Yet, in what follows, I wish to consider two possible concerns with this proposal. The first concern claims that we lack empirical grounds for thinking that narrative counterspeech is effective. The second insists that, even if narrative counterspeech *were* effective, it would still be manipulative, and therefore unacceptable.

The Effectiveness Objection

The most immediate concern with narrative counterspeech is empirical. According to this objection, the claim that narrative counterspeech constitutes an effective response to conspiracy theories is at bottom an empirical claim. Yet that claim, the objection continues, is unsupported by empirical evidence.

There are different versions of this objection. The first version holds that we simply do not have evidence relating to narrative counterspeech. And so, we simply do not know whether it is effective.

Even if this first version of the objection were true, it would not be a problem for my argument. When social scientists investigate the effects of various forms of counterspeech, they do not do so randomly. Instead, as mentioned in the Introduction, they rely on informed theoretical hypotheses regarding which interventions may be effective. Political theorists and philosophers have an important role to play in generating and articulating such hypotheses (for a defence of this methodological point, see Lepoutre, 2021: 58–59). Accordingly, my aim has been to put forward such a hypothesis – a hypothesis,

informed by insights from social epistemology, philosophy of emotion, and cognitive science, regarding which forms of counterspeech might reasonably be expected to be effective at countering conspiracy theories.

What this means is that the lack of existing evidence in support of narrative counterspeech need not be an objection to my proposal. Rather, it is part of what motivates it. My argument aims to motivate, and offer guidance for, further empirical investigation. It suggests that, going forward, we should test the effectiveness of narrative counterspeech at countering conspiracy theories.

Yet there is a second and more problematic version of the ‘effectiveness’ objection. According to the second version, we *do* have empirical evidence relating to narrative counterspeech. The problem is that this evidence demonstrates that narrative counterspeech is ineffective.

Some recent studies might seem to support this second objection. Ullrich Ecker et al. (2020: 2) report, based on experimental evidence, that ‘narrative corrections are no more effective than non-narrative corrections’. Likewise, in a review of existing empirical evidence, Aleksandra Lazic and Iris Zezelj (2021: 655) find little difference in effectiveness between narrative and non-narrative interventions aimed at promoting vaccines. Narratives sometimes outperformed non-narrative interventions, but the reverse also happened. And even when narratives outperformed non-narrative interventions, the difference tended to be small.

On closer inspection, however, this existing evidence does not warrant scepticism regarding the hypothesis I have put forward. For one thing, even if the evidence succeeded in showing that narrative counterspeech is no more effective than non-narrative counterspeech, this would not necessarily mean that narrative counterspeech is ineffective. In fact, the evidence suggests the opposite. Narrative counterspeech appears to have a positive effect on beliefs and intentions when compared with no counterspeech (Lazic and Zezelj, 2021: 655). Moreover, Lazic and Zezelj (2021) also find that the most successful interventions are those that *combine* narrative and non-narrative forms of counterspeech.

These positive results are tentative. But insofar as they are valid, they already have implications for practice. First, they suggest that it is better to engage in narrative counterspeech than in no counterspeech whatsoever. This implication is far from obvious. For the reasons canvassed in the first half of this paper, some social psychologists and philosophers of language have expressed the worry that counterspeech might be ineffective – or worse, counterproductive.¹⁴ Second, these positive results also suggest that we should not limit our interventions to non-narrative counterspeech. Even where fact-checks are already in place, the evidence outlined above suggests that they would be more successful if paired with narrative counterspeech.

Still, this first response is not entirely satisfactory. Although existing evidence does not suggest that narrative counterspeech is ineffective, it may still seem to contradict the specific theoretical hypothesis I have offered. My theoretical hypothesis is not merely that narrative counterspeech may be an effective tool for responding to conspiracy theories. It is that, in virtue of its distinctive characteristics, narrative counterspeech may be *more* effective at doing so than non-narrative counterspeech. On the face of it, the two studies introduced above appear to contradict this prediction.

Yet this appearance is misleading. The existing evidence is in fact too coarse to contradict or corroborate my theoretical hypothesis. There are two reasons for this. The first is that existing evidence relating to narrative interventions – including the two studies

outlined above – typically does not focus on conspiracy theories.¹⁵ This is an important limitation in our context. The theoretical argument I have offered suggests that narrative will be comparatively effective (relative to non-narrative counterspeech, such as fact-checking) *in response to conspiracy theories*. Indeed, I have argued that conspiracy theories present a distinctive problem (due to their salience, emotional dimension and coherence) that storytelling is comparatively well-placed to handle. To assess *this* claim, more targeted empirical investigation are needed.

There is a second problem with existing evidence. Some stories are better than others. They are more cohesive, more engrossing and more emotionally engaging. The problem is that the studies that tend to be most pessimistic about the prospects of narratives *also* tend to deploy poorly designed narratives – narratives that, in light of the analysis I have offered, seem unlikely to engage readers' attention or elicit strong emotions.

This is most visible in Ecker et al.'s experimental study – the study, to date, that offers the most sustained comparison of narrative and non-narrative counterspeech, and that most strongly rejects the claim that narrative counterspeech is comparatively effective. In terms of narrative structure, the study's narrative vignettes generally have a weak plot. Indeed, their brevity means that the story's arc of causally connected events – in particular, the triggering and subsequent resolution of the problem – tends to be extremely bare and compressed. And, for the same reason, very little is known about the protagonist, making it difficult to relate to them.

A similar observation goes for narrative style. The vignettes are usually written in the third person. They seldom employ emotional language, let alone detailed information about the internal lives of its characters. And they make little, if any, use of figurative language. In short, these stories tend to lack the characteristic structural and stylistic features that, according to the argument I have offered, would allow narratives to capture people's attention and arouse their emotions.¹⁶ It is therefore unsurprising, from the perspective of this argument, that they do not prove comparatively effective.

This is significant. Just as there can be high-quality and low-quality fact-checks, so too there can be high-quality and low-quality narratives. The fact that low-quality narratives are not distinctively effective does not tell us much. It is compatible with thinking that, when designed in accordance with the recommendations outlined earlier, narrative may be a distinctively useful tool for countering conspiracy theories.

The broader upshot is that existing empirical evidence remains too limited to test, or indeed to refute, the hypothesis I have advanced. This upshot in turn has implications for future research. To assess my proposal, future empirical investigations should specifically examine the comparative merits of narrative and non-narrative counterspeech *that targets conspiracy theories*. And they should conform, in structure and style, to the characteristics I have outlined.”

The Manipulation Objection

But there is another possible concern with narrative counterspeech. Even if we suppose that it is effective at getting its audience to reject conspiratorial claims, one might worry that it does so in an ethically problematic manner. Specifically, one might worry that narrative counterspeech is unacceptably manipulative. Why might this be? The general idea is that narratives induce listeners to accept true propositions, or reject false propositions, in a way that bypasses their reason.

Critics of narrative usually cite one of two justifications for this thought. The first is simply that, as we have seen, narratives appeal to people's emotions. Accordingly, one

might worry that people will reject conspiracy theories, not because they have epistemically good reasons to do so – for example, good evidence that the conspiracy is not taking place – but rather because of how they feel (Lazic and Zezelj, 2021: 656).

The other reason is that narratives are often partly fictionalised. Stories usually omit some features of the events they report, while distorting facts that they do include within the story (ElShafie, 2018: 1220). For example, a pro-vaccine story about the development of COVID-19 vaccines might disproportionately emphasise positive facts about the scientists who developed it (e.g. their commitment and selflessness) while omitting, say, negative details about office politics. Moreover, the story might oversimplify the process of development, for instance, by focusing exclusively on the contributions of one or two ‘main’ individuals.

The tendency to fictionalise is no accident. These omissions and distortions might be needed to make events fit a standard story structure, involving an identifiable protagonist who faces up to a meaningful problem (ElShafie, 2018: 1220). And they might be needed, moreover, to make the story an *effective* story: a story that successfully grabs people’s attention, and causes them to feel strongly with, or strongly about, the protagonist and their journey.

Put together, these observations might seem to imply that narrative counterspeech is manipulative. It may succeed in getting people to reject conspiratorial claims. But it does so, one might think, in a disrespectful way. It bypasses their reason, either by playing on their emotions or by feeding them falsehoods.

There are several things to say in response to this objection. The first is simply that, even if narrative counterspeech did bypass its targets’ reason, there may nonetheless be cases where doing so is all-things-considered justifiable. In this vein, Jason Stanley (2015: 78, 112) argues that speech that bypasses the rational faculties (which he refers to as ‘positive propaganda’) can play a legitimate role in liberal democracies. The thought underpinning this response is that moral reasons not to manipulate are not absolute. Even if they are weighty, they are in principle capable of being overridden. Thus, we can accept that adopting a manipulative messaging strategy is an evil – but if the alternative is the immensely dangerous proliferation of conspiracy theories about COVID-19 vaccines, it may nevertheless be the lesser evil.

Yet there is a further, and arguably more important, point: the fact that narratives engage the emotions and incorporate falsehoods does not in and of itself show that narrative counterspeech bypasses reason.

This is easiest to see with emotion. The objection at hand assumes that emotions are divorced from rational thought. But we have already encountered grounds for doubting this claim. Recall that emotions have a cognitive dimension: they are sources of salience, that highlight features of the object they are directed at (see Section ‘Emotion’). This salience role is epistemically valuable. It helps us manage informationally rich environments, by drawing our attention to features of that environment we may otherwise have overlooked. For example, feeling anger can highlight possible sources of injustice. Likewise, feeling fear can cast a spotlight on dangers we would otherwise have overlooked. What this suggests is that there is no necessary connection between experiencing emotions, and failing to reason. On the contrary, emotions can and often do enrich the information on the basis of which we reason (see, e.g. Brady, 2013; Elgin, 2008).

The point extends to falsehoods as well. Catherine Elgin (2017) has argued that some falsehoods are *felicitous*, in the sense that they help improve our understanding. And they too, do so by helping to render salient important properties of a target object. For example, the standard map of the London Underground distorts the geographical relations

between stations. But this misrepresentation serves an epistemic purpose. It makes it more visible how many stops separate two stations, as well as where one can change from one line to another. And, by doing so, it makes it easier to reason about the best way to travel to one's destination.¹⁷

Elgin (2017) argues that this phenomenon applies across a vast range of domains, including scientific practice, philosophical methodology and literary fiction. Narrative counterspeech is no exception. A narrative's distortions and omissions can, in principle, help to highlight important features of the narrative's object. For example, a narrative might abstract away from the office politics of scientists working on COVID-19 vaccines in order better to highlight something real about those scientists: namely, their public-facing motivations and commitment to scientific rigour. Similarly, emphasising the contributions of one or two scientists underplays the collective dimension of vaccine development. At the same time, however, it may help audiences notice and retain core features of this process. Giving the events a clear protagonist and with it, a standard story structure, can make it easier for audiences to simulate, imagine and recall, core stages of vaccine development.

To be clear: I am not claiming that all falsehoods, and all emotions, facilitate reasoning. This is clearly false. After all, the falsehoods and emotions promulgated by conspiracy theories tend to do the opposite.

My point is therefore more modest. It is simply that the relationship between emotions, falsehoods and reasoning is more ambivalent than the objection assumes. Falsehoods, like emotions, can facilitate as well as impair reasoning. And so, from the fact that narrative counterspeech arouses emotions, and incorporates simplifying or distorting falsehoods, it does not necessarily follow that it bypasses reason – and thus, it does not necessarily follow that it manipulates its targets.

Conclusion

Conspiracy theories pose a distinctive challenge for counterspeech. Due to their self-sealing internal coherence, they are nearly impervious to logical or empirical refutation. What is more, their emotional intensity and enduring salience make them dangerously seductive – more seductive, typically, than rigorous fact-checking.

I have argued that, to address this problem, we should deploy *narrative* forms of counterspeech. In other words, we should oppose conspiracy theories not simply by stating and explaining facts, but by telling a compelling story. Because of their characteristic cohesion, structure and stylistic features, narratives are well-placed to overcome the resistance of conspiracy theories.

But not all forms of narrative counterspeech will do. The diagnosis I have offered entails that some forms of narrative counterspeech are more likely than others to succeed in eroding support for dangerous conspiracy theories. Among other things, narrative forms of counterspeech are more likely to provoke an emotionally intense and engrossing experience when they cue a simulationist representational format. And the extent to which they do so, as we have seen, itself depends on meaningful structural and stylistic choices.

This diversity in narrative forms has implications for both research and practice. It explains, in the first place, why existing empirical evidence relating to narrative's effects remains mixed. To assess the full potential of narrative counterspeech, future empirical investigations must discriminate, in the ways I have suggested, between different forms of narrative. Relatedly, the upshot for practice is that it is not enough simply to engage in

storytelling. Successfully responding to conspiracy theories requires us to engage in *skilled* storytelling: storytelling that, while grounded in science and expertise, harnesses the full creative potential of literary fiction.

Acknowledgements

For helpful discussions and feedback on this paper, the author is deeply grateful to Laura Caponetto, Bianca Cepollaro, Joanna Demaree-Cotton, Álvaro Domínguez Armas, Rachel Fraser, Corrado Fumagalli, Jeff Howard, James Hutton, Teresa Marques, Mihaela Popa-Wyatt, Alexander Prescott-Couch, Miriam Ronzoni, Jonathan Seglow, Robert Simpson and the participants at the MANCEPT panel on Hate Speech, Disinformation and Counterspeech. This paper also benefitted from the extremely helpful comments of two anonymous reviewers for *Political Studies*.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This research was generously supported by a Leverhulme Research Fellowship (RF-2022-001: ‘Felicitous Fictions: Legitimate Falsehoods in Public Discourse’).

ORCID iD

Maxime Lepoutre  <https://orcid.org/0000-0001-7573-8585>

Notes

1. Note that conspiracy theories can also generate problems that go beyond the specific policy issues on which they bear. According to Muirhead and Rosenblum (2020), conspiracy theories also promote generalised distrust in epistemic institutions, and erode the democratic processes that depend on these institutions.
2. These two remedies – legal restrictions and counterspeech – are not meant to be exhaustive. Drochon (2018: 337) has argued that exclusion (in particular, political, economic and social exclusion) constitutes ‘the strongest explanatory factor for belief in conspiracy theories’. As a result, he recommends countering conspiracy theories at a more structural level – namely by adopting socioeconomic policies aimed at tackling this multifaceted exclusion. I am sympathetic to this proposal. However, tackling multifaceted exclusion is an extremely long-term project. Accordingly, even if Drochon’s diagnosis and proposed remedy are correct, the conditions in which conspiracy theories flourish will remain in place for the foreseeable future. During this time, we need more immediate ways of countering conspiracy theories. I am putting forward narrative counterspeech as an important part of this more immediate solution.
3. Neither Dahlstrom (2014) nor ElShafie (2018) is specifically concerned with responding to misinformation (let alone the specific form of misinformation constituted by conspiracy theories). Lasic and Zezelj (2021: 655–656) come closer to doing this. But as they acknowledge, their recommendations are extremely tentative. They do not offer a sustained account of why conspiracy theories generally resist non-narrative counterspeech. This limits their ability to explain why narrative is so promising, and what specific form it should take.
4. Political theorists and social scientists have devised various strategies for overcoming this problem of access, including reforming social media algorithms to promote exposure to diverse perspectives, the ‘cognitive infiltration’ of echo-chambers, and more robust policies aimed at achieving intergroup spatial integration. For discussion, see, for example, Sunstein and Vermeule (2009), Sunstein (2017), Settle (2018) and Lepoutre (2021: ch. 7).
5. Worries about salience do not apply exclusively to conspiracy theories. But, as explained above, conspiracy theories are distinctly likely to become salient, due to their characteristically entertaining content.
6. The potential impact of counterspeech on a conspiracy theory’s familiarity and credibility does not necessarily mean that counterspeech will *increase* belief in those conspiracy theories. Evidence of a so-called

- 'backfire' effect – whereby attempts at fact-checking misinformation reinforce belief in that piece of misinformation – has proven difficult to replicate (Wood and Porter, 2020). But what the evidence does suggest is that, even when fact-checking does not actually increase support for the theories it targets, it scarcely reduces support for those theories. On this latter point, see Guess and Coppock (2018: 13) and Wood and Porter (2020: 150–151).
7. Other scholars highlight different conspiracy theory-inviting emotions. For example, Guilhot (2021) instead explains the appeal of conspiracy theories by pointing to the way they help us cope with 'apocalyptic anxieties', relating, for example, to the climate emergency, war, pandemics and so on. Drochon (2018), for his part, underscores the role played by feelings of powerlessness and exclusion.
 8. Not all conspiracy theories reinforce the emotions that initially made them appealing. Take, for example, the 'apocalyptic anxieties' discussed by Guilhot (2021). The purpose of some conspiracy theories is partly to alleviate these anxieties. For example, believing climate change to be a hoax may well *reduce* one's climate-related anxieties.
 9. Muirhead and Rosenblum (2020) argue that this is not always true. They suggest that contemporary conspiracism often involves 'conspiracy without the theory': allegations of conspiracy without an accompanying account of how the conspiracy works. I am not fully convinced by their diagnosis. What sometimes appears to be an undeveloped allegation of conspiracy (e.g. a Tweet simply saying 'The election was rigged!') may actually operate as a cue that activates a broader perspective (e.g. a worldview involving distrust of government; belief in a deep state populated by university graduates and urbanites; theories about voter fraud; etc.). But insofar as Muirhead and Rosenblum's diagnosis is accurate, the point introduced in the present section is less applicable to their 'new conspiracism'.
 10. The self-sealing quality of conspiracy theories plays an important role in explaining away the following piece of seemingly contradictory evidence. Conspiracy theories often ascribe extraordinary levels of competence and effectiveness to elites – and this might seem inconsistent with elites' performance in many other domains. Yet conspiracy theorists might respond, once more, that their theory *predicts* this apparent inconsistency. Elites, they might say, pretend to be incompetent in public in order better to hide their tracks. I am grateful to a reviewer for pressing me on this point.
 11. Cassam (2019: 102–104) recognises this problem, and recommends that readers should then do all they can to disseminate Posner's conclusions. And he also acknowledges, ultimately, that engaging the emotions is an important part of the process (Cassam, 2019: 122–124). But this raises the question of *how* these facts should be disseminated; and, relatedly, of how to do so in a way that engages the emotions. Narrative counterspeech, I am suggesting, may well be the answer.
 12. Fraser (2021: 4036) rightly notes that *not all* narratives cue a first-personal format, and that whether they do so depends partly on stylistic choices – a point to which I will later return. But the point for now is that the standard story structure tends to invite first-personal simulation.
 13. There is preliminary evidence supporting this claim. In their review of empirical evidence regarding narratives, De Graaf et al. (2016: 98–99) find that use of the first-personal standpoint and 'descriptions of emotional experiences' were both associated with greater persuasiveness. However, this evidence remains tentative. First, the review does not distinguish between studies that compared narrative interventions to non-narrative interventions, and studies that compared narrative interventions to no intervention. So, the review cannot decisively tell us that these stylistic features make a difference relative to non-narrative interventions. Second, the studies canvassed by the review typically do not involve responses to conspiracy theories. This is a limitation because, based on the diagnosis I have offered, the benefits of (certain) narratives are especially likely to be salient when responding to conspiracy theories.
 14. For the worry about counterproductivity, see Nyhan and Reifler (2010) and Simpson (2013). Now, as mentioned in Note 6, Wood and Porter (2020) have argued that this worry does not stand up to empirical scrutiny. Yet even if they are right about this, this is compatible with worrying that counterspeech is likely to be ineffective (even if it does not strictly speaking backfire). For this more moderate worry, see, for example, Berinsky (2017: 242), Guess and Coppock (2018: 13) and Wood and Porter (2020: 150–151).
 15. Lazić and Zezelj (2021: 650) explicitly note that this is a limitation of their data. Ecker et al. (2020) explore the use of narrative in response to disinformation. But, as discussed above, conspiracy theories are a very specific form of disinformation, whose distinctive properties explain why narrative counterspeech is comparatively well-placed to counter it.
 16. See, for example, the Wildfire story (Ecker et al., 2020: 15). It involves little sense of a challenge or problem (the vignette describes a routine investigation), little sense of something meaningful being at stake (the fire has already been brought under control when the story starts), no description of emotions, no use of metaphors and nearly no use of a first-personal standpoint.

17. Note that this epistemic benefit does not depend on the map user realising that the map distorts geographical properties. A user will find it easier to navigate the Underground using the map *even if* they are unaware of these geographical distortions. On this point, see Lepoutre (2022: 16).

References

- Albertson A and Guiler K (2020) Conspiracy Theories, Election Rigging, and Support for Democratic Norms. *Research and Politics*. Epub ahead of print 28 September. DOI: 10.1177/2053168020959859.
- Berinsky A (2017) Rumors and Health Care Reform: Experiments in Political Misinformation. *British Journal of Political Science* 47: 241–262.
- Boseley S (2019) No Link between Autism and MMR, Affirms Major Study. *The Guardian*. Available at: <https://www.theguardian.com/society/2019/mar/04/no-link-between-autism-and-mmr-affirms-major-study> (accessed 2 June 2022).
- Brady M (2013) *Emotional Insight*. Oxford: Oxford University Press.
- Brown E (forthcoming) Free Speech and the Legal Prohibition of Fake News. *Social Theory and Practice*.
- Camp E (2017) Why Metaphors Make Good Insults. *Philosophical Studies* 17 (4): 47–64.
- Carmichael F and Goodman J (2020) Vaccine Rumours Debunked: Microchips, ‘Altered DNA’ and More. *BBC*. Available at: <https://www.bbc.co.uk/news/54893437> (accessed 2 June 2022).
- Cassam Q (2019) *Conspiracy Theories*. Cambridge: Polity Press.
- Center for Disease Control and Prevention (2022) *Why HPV Vaccine is Important to My Family: The Story of a Cervical Cancer Survivor*. Available at: https://tools.cdc.gov/podcasts/media/pdf/CDC_JackiesStory_POD.pdf (accessed 2 June 2022).
- Dahlstrom M (2014) Using Narratives and Storytelling to Communicate Science with Nonexpert Audiences. *Proceedings of the National Academy of Sciences* 111 (4): 13614–13620.
- De Graaf A, Sanders J and Hoeken H (2016) Characteristics of Narrative Interventions and Health Effects. *Review of Communication Research* 4: 88–131.
- Deonna J and Teroni F (2012) *The Emotions: A Philosophical Introduction*. New York: Routledge.
- Douglas K, Uscinski J, Sutton R, et al. (2019) Understanding Conspiracy Theories. *Advances in Political Psychology* 40 (1): 3–35.
- Drochon H (2018) Who Believes in Conspiracy Theories in Great Britain and Europe? In: Uscinski J (ed.) *Conspiracy Theories and the People Who Believe Them*. Oxford: Oxford University Press, pp.337–46.
- Ecker U, Butler L and Hamby A (2020) You Don’t Have to Tell a Story. *Cognition Research* 5: 64.
- Elgin C (2008) Emotion and Understanding. In: Brun G, Doguoglu U and Kuenzle D (eds) *Epistemology and Emotions*. Farnham: Ashgate, pp.33–50.
- Elgin C (2017) *True Enough*. Oxford: Oxford University Press.
- ElShafie SJ (2018) Making Science Meaningful for Broad Audiences through Stories. *Comparative Biology* 58 (6): 1213–1223.
- Enders A, Uscinski J, Klofstad C, et al. (2021) Do Conspiracy Beliefs Form a Belief System? Examining the Structure and Organization of Conspiracy Beliefs. *Journal of Social and Political Psychology* 9 (1): 255–271.
- EU vs Disinformation (2020) Disinfo: Vaccines Are a Big Pharma Fraud Led by Bill Gates’. *EU vs Disinformation*. Available at: <https://euvsdisinfo.eu/report/coronavirus-vaccines-big-pharma-fraud-bill-gates/> (accessed 2 June 2022).
- Fraser R (2021) Narrative Testimony. *Philosophical Studies* 178: 4025–4052.
- Guess A and Coppock A (2018) Does Information Cause Backlash? *British Journal of Political Science* 50: 1497–1515.
- Guilhot N (2021) Bad Information. *Boston Review*. Available at: <https://bostonreview.net/articles/bad-information/> (accessed 9 September 2022).
- Howard J (2021a) Coronavirus Misinformation, Social Media and Freedom of Speech. In: Niker F and Bhattacharya A (eds) *Political Philosophy in a Pandemic: Routes to a More Just Future*. London: Bloomsbury, pp.167–176.
- Howard J (2021b) Terror, Hate, and the Demands of Counterspeech. *British Journal of Political Science* 51: 924–939.
- Jolley D and Douglas K (2014) The Effects of Anti-Vaccine Conspiracy Theories on Vaccination Intentions. *PLoS ONE* 9 (2): e89177.
- Jolley D, Mari S and Douglas K (2020) Consequences of Conspiracy Theories. In: Butter M and Knight P (eds) *Routledge Handbook of Conspiracy Theories*. London: Routledge, pp.231–41.

- Lazic A and Zezelj I (2021) A Systematic Review of Narrative Interventions: Lessons for Countering Conspiracy Theories and Misinformation. *Public Understanding of Science* 30 (6): 644–670.
- Lepoutre M (2021) *Democratic Speech in Divided Times*. Oxford: Oxford University Press.
- Lepoutre M (2022) Political Understanding. *British Journal of Political Science*. Epub ahead of print 23 March. DOI: 10.1017/S0007123422000023.
- Lewandowsky S, Ecker U, Seifert C, et al. (2012) Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest* 13 (2): 106–131.
- Mandler J and Johnson N (1977) Remembrance of Things Parsed: Story Structure and Recall. *Cognitive Psychology* 9: 111–151.
- Moran (1989) Seeing and believing: Metaphor, image, and force. *Critical Inquiry* 16: 87–112.
- Muirhead R and Rosenblum N (2020) *A Lot of People Are Saying*. Princeton, NJ: Princeton University Press.
- Mutz D (2006) *Hearing the Other Side*. Cambridge: Cambridge University Press.
- Nagesh A (2021) The Moment QAnon Took the Person I Love Most. *BBC*. Available at: <https://www.bbc.co.uk/news/world-us-canada-57369349> (accessed 2 June 2022).
- Napolitano G (2021) Conspiracy Theories and Evidential. In: Bernecker S, Floweree A and Grundmann T (eds) *The Epistemology of Fake News*. Oxford: Oxford University Press, pp.82–108.
- Nyhan R and Reifler J (2010) When Corrections Fail. *Political Behavior* 32: 303–330.
- Prescott-Couch A (manuscript) Narrative Understanding, June 2021.
- Reuters (2021) Fact Check: The Pandemic Was Not Orchestrated by Pharmaceutical Companies, Investment Groups, and Philanthropists. *Reuters*. Available at: <https://www.reuters.com/article/uk-factcheck-pharmaceuticals-philanthrop-idUSKBN29Z0TM> (accessed 2 June 2022).
- Romer D and Jamieson K (2020) Conspiracy Theories as Barriers to Controlling the Spread of in the U.S. *Social Science & Medicine* 263: 113356.
- Saul J (2021) Someone Is Wrong on the Internet: Is There an Obligation to Correct False and Oppressive Speech on Social Media? In: MacKenzie A, Rose J and Bhatt I (eds) *The Epistemology of Deceit in a Postdigital Era: Dupery by Design*. New York: Springer, pp.139–57.
- Settle J (2018) *Frenemies: How Social Media Polarizes America*. Cambridge: Cambridge University Press.
- Shelby A and Ernst K (2013) Story and Science. *Human Vaccines & Immunotherapeutics* 9: 1795–1801.
- Simpson R (2013) Un-ringing the Bell. *Australasian Journal of Philosophy* 91: 555–575.
- Stanley J (2015) *How Propaganda Works*. Princeton, NJ: Princeton University Press.
- Sunstein C (2017) *#Republic*. Princeton: Princeton University Press.
- Sunstein C and Vermeule A (2009) Conspiracy Theories: Causes and Cures. *Journal of Political Philosophy* 17 (2): 202–227.
- Talisse R (2019) *Overdoing Democracy: Why We Must Put Politics in Its Place*. Oxford: Oxford University Press.
- Tonkovic M, Dumancic F, Jelic M, et al. (2021) Who Believes in Conspiracy Theories in Croatia? *Frontiers in Psychology* 12: 643568.
- Uscinski J (2020) *Conspiracy Theories: A Primer*. New York: Rowman & Littlefield.
- van Prooijen J-W, Ligthart J, Rosema S, et al. (2021) The Entertainment Value of Conspiracy Theories. *British Journal of Psychology* 113: 25–48.
- Vardi N (2021) Covid's Forgotten Hero. *Forbes*. Available at: <https://www.forbes.com/sites/nathanvardi/2021/08/17/covids-forgotten-hero-the-untold-story-of-the-scientist-whose-breakthrough-made-the-vaccines-possible/> (accessed 2 June 2022).
- Walter A and Drochon H (2022) Conspiracy Thinking in Europe and America: A Comparative Study. *Political Studies* 70 (2): 483–501.
- Wood T and Porter E (2020) The Elusive Backfire Effect. *Political Behavior* 41: 135–163.
- Worland J (2019) Donald Trump Called Climate Change a Hoax. Now He's Awkwardly Boasting About Fighting It. *Time*. Available at: <https://time.com/5622374/donald-trump-climate-change-hoax-event> (accessed 2 June 2022).

Author Biography

Maxime Lepoutre is a Lecturer in Political Theory at the University of Reading. Prior to joining Reading, he worked as a Prize Research Fellow at the University of Oxford, and obtained a PhD in philosophy at the University of Cambridge. In 2019, he received the Ernest Barker Prize for Best Dissertation in Political Theory. His research centres on political philosophy, legal theory and social epistemology. He has written extensively about democracy, with emphasis on issues surrounding hate speech, disinformation, group identity and the function of political anger.