



Social Epistemology

A Journal of Knowledge, Culture and Policy

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tsep20

On the Censorship of Conspiracy Theories

Fred Matthews

To cite this article: Fred Matthews (30 Jan 2025): On the Censorship of Conspiracy Theories, Social Epistemology, DOI: [10.1080/02691728.2025.2449612](https://doi.org/10.1080/02691728.2025.2449612)

To link to this article: <https://doi.org/10.1080/02691728.2025.2449612>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 30 Jan 2025.



Submit your article to this journal [↗](#)




View related articles [↗](#)



View Crossmark data [↗](#)

On the Censorship of Conspiracy Theories

Fred Matthews 

Department of Philosophy, University of Bristol, Bristol, UK

ABSTRACT

Is it permissible for the state to censor or suppress conspiracy theories, even within liberal democracies? According to a number of political and legal theorists, it is. In this paper, I will argue that the state may sometimes censor conspiracy theories, but it should be permitted to do so only after very strict conditions have been met. I shall first offer some brief thoughts about the definition of ‘conspiracy theory’. I will then critique one existing attempt to address this issue – namely Cibik and Hardoš’s public reason approach. Next, I shall outline my own proposal. I will argue that we should sometimes consider conspiracy theorising to be a form of discriminatory speech against vulnerable individuals and groups, and we can consider it to be a form of defamation in these cases. Consequently, the state may sometimes be permitted to enact civil laws to sanction such theorising. Finally, I will outline some of the conditions that should be met before state censorship can be considered.

ARTICLE HISTORY

Received 21 June 2024

Accepted 31 December 2024

KEYWORDS

Conspiracy theories; public reason; defamation; free speech

1. Introduction

Conspiracy theories (CTs) generate a significant amount of political controversy, and it appears that CTs are getting more bad publicity with each passing week. Although ‘conspiracy theories’ are not uniformly considered to be bad, the term is often considered to be pejorative (see Napolitano 2021, 82–84). Therefore, it is not surprising that there have been calls from some political philosophers (and others) to enact legislation to censor or at least discourage conspiracy theorising. The purpose of this article will be to consider the conditions under which it could be acceptable to enact such legislation. First, I will briefly outline what I take a conspiracy theory to be, at least for the purpose of the current discussion. Secondly, I will analyse one previous attempt to argue that CTs can sometimes legitimately be subjected to censorship: Cibik and Hardoš’s public reason approach put forward in their article ‘Conspiracy Theories and Reasonable Pluralism’ (2022).

Thirdly, I will put forward my own position regarding when it is acceptable for the state to enact legislation designed to censor or discourage CTs. I will argue that the state is sometimes justified in using civil laws (within a common law system) to discourage CTs that carry a serious risk of defamation against vulnerable groups (cf. Brison 1998, 313; Waldron 2012, *passim*; Xin 2024, 446–449). Lastly, I will argue that there are some stringent conditions that will need to be met before state intervention can be legitimately considered. There is an *epistemological* constraint and a *practical* constraint. The epistemological constraint says that we should be duly cautious about introducing legislation which is designed to censor or discourage a CT if the theory in question has some plausibility. The practical constraint says that there should be good reason to think that state regulation will be effective.

CONTACT Fred Matthews  fred.matthews@bristol.ac.uk  Department of Philosophy, University of Bristol, Cotham House, Bristol BS6 6JL, UK

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

In this article, I am going to be discussing whether or not state regulation of CTs is acceptable within liberal democracies, and will not primarily concentrate on non-liberal states. There are two reasons for this. First, much of the political controversy surrounding CTs (although by no means all) is occurring within liberal democratic states, particularly the US. Secondly, some of the most important attempts to argue for the censorship of conspiracy theories are found within liberal political theory. This does not mean that the arguments made here *all* rely on the framework of liberal democracy. Many of the considerations in this article will also be of interest to some non-liberal socialists, conservatives, Marxists, and anarchists, for instance. However, it is important to remember that some of the conclusions here might not apply if we were to assume the framework of one of these theories instead of liberalism.

When I write about 'liberal democracies', I am referring to states which are *democratic* in the sense that they have, roughly speaking, free and fair elections. This means that, at a minimum, there is universal suffrage for adults, that elections are not rigged, and that there is a genuine choice between at least two candidates or parties. These states are *liberal* in the sense that they make provision for civil liberties such as free speech and freedom of religion, and they have a separation of powers.¹ It is not easy to pin down exactly what counts as 'liberal theory', because liberal political philosophy is broad and varied. However, liberal political philosophy generally emphasizes the importance of individual freedom, equality before the law, human rights, and in many cases economic equality. Later, I shall discuss more specific versions of liberalism (such as the Rawlsian public reason liberalism defended by Cívik and Hardoš). However, for the most part, I will rely on the broad understanding of 'liberal democracy' and 'liberal theory' outlined here. I will now move on to a brief discussion of how we can understand the term 'conspiracy theory'.

2. 'Conspiracy Theory': An Uncertain Term

Much ink has already been spilled on how to define and conceptualise conspiracy theories, and there is no definition of CTs which can satisfy everyone. At the most basic level, the term 'conspiracy theory' refers to any theory which suggests that there is a conspiracy. To be more precise, it refers to theories (i.e. explanations) which state that some event has been brought about by two or more persons (conspirators) acting in secret, usually with the connotation that the conspirators have malevolent intentions. Clearly, on this basic definition, many conspiracy theories are undoubtedly true and well justified. The Watergate scandal is a very famous example of a proven conspiracy. Another well-known case is the proven claim that Lance Armstrong conspired with other members of the US Postal Service Cycling Team to cheat during the Tour de France by using banned doping substances – and possibly was, to some degree, protected by members of the sport's governing bodies.

The term 'conspiracy theory' is, however, often considered to be pejorative in nature, and to indicate that a theory has serious epistemic defects. Neither of the now-proven theories mentioned above would usually be considered 'conspiracy theories'. However, the term is not always used pejoratively, and there is little agreement about why something should be defined as a 'conspiracy theory' in the abusive sense of the term (e.g. which epistemic defects it needs to have). Some philosophers have attempted to clarify the matter. Some have argued for a neutral, minimalist definition of CTs (Dentith 2014), while others have argued that 'conspiracy theories' are intrinsically resistant to counter-evidence, in contrast to credible theories about conspiracies (e.g. Napolitano 2021). Some have argued that we can define CTs as theories about conspiracies that conflict with some official account (e.g. Coady 2019). While this undoubtedly captures *something* about the way in which people commonly talk about conspiracy theories, it is also imprecise because official sources sometimes *themselves* promote explanations that intuitively seem to be 'conspiracy theories' (Shields 2022). A different approach is found in Cassam (2019). He argues that 'Conspiracy Theories' generally have the following features: they are speculative, contrarian, esoteric, amateurish, premodern,

and are a form of political propaganda.² While there can be little doubt that some conspiracy theories have these features, this appears to be far too broad to apply even to all unwarranted CTs (see Dentith 2023). For instance, not all unwarranted or irrational 'conspiracy theories' are put forward by amateurs (see Shields 2022, 331).

What is the best approach for an article like this, in which there is insufficient space to enter the complex debate about the conceptualisation of conspiracy theories? I think it is important to understand that, as with terms such as 'democracy' and 'knowledge', there is no single definition that is going to capture all the important aspects of the term 'conspiracy theory'. It should be regarded as an 'essentially contested concept': no single analysis or definition can resolve questions surrounding the essence of 'conspiracy theories', because no such analysis can possibly capture the term's intrinsic complexity, value connotations, and multifaceted usage (Gallie 1956).

With that in mind, though, we can still generalise the kinds of conspiracy theories that will be discussed in this paper, and we can outline the sorts of conspiracy theories to which the arguments in this paper are *more likely* to apply. I take it as obvious that it would be implausible for any state to censor CTs when they are defined in accordance with the minimalist understanding. Just because a theory posits that there is a conspiracy, this does not mean that it requires state censorship or intervention. I shall, therefore, inevitably have to focus on a subset of CTs (or perhaps several subsets) when discussing the issue of whether the state can legitimately suppress or censor CTs. Let us, therefore, have a closer look at what kinds of CTs will be covered by the arguments in this paper.

The CTs discussed here will, of course, be theories which state that an event has been brought about by two or more persons (conspirators) acting in secret, usually with the connotation that the conspirators have malevolent intentions. These theories will often be in conflict with official sources, at least in Western societies. This is because, as we shall see below, one of the general criteria for deciding whether a CT is a legitimate candidate for state intervention is a lack of plausibility. But how are we supposed to tell whether or not a CT is genuinely implausible? Much of the time, it is necessary to look to the testimony of others, as we cannot possibly investigate all the CTs that have some *prima facie* plausibility. In some cases, we need to rely on 'official' sources of information – such as public health institutions, academics, and government agencies – because (for one thing) nobody has the knowledge, time, or ability to assess all the relevant facts and information. However, it is also important to recognise that not all CTs that run contrary to official stories in Western societies are irrational, and not all 'official' sources will necessarily agree with each other.

There is another feature that many of the CTs discussed in this article will have: they will generally be resistant to counter-evidence. That is, they will have a tendency to explain away or absorb possible sources of counter-evidence. As we shall see in section three, CTs that carry a serious risk of defamation are likely to have this feature, because they will probably involve multiplying accusations of conspiracy. As I will also discuss in section three, there is a philosophical debate about whether CTs that appear to be evidentially self-insulating necessarily suffer from epistemic faults simply because of this. Nonetheless, as we shall see, it is still important for the ethical arguments in this article that some CTs at least *appear* to be evidentially self-insulating.

Therefore, as a general rule, we can say that the conspiracy theories discussed in this article will have the following features. They will be theories (explanations) which posit that some event or series of events has been brought about by at least two persons acting in secret. They will also often, but not always, be theories that conflict with the 'official' account in Western societies. Lastly, they will often have a tendency to resist counter-evidence, and will sometimes generate a large number of conspiracy accusations to support themselves. I will elaborate on these features in later sections, but it is important to realise that not all CTs have these characteristics. The CTs with these attributes are simply those that are likely to be relevant in light of the arguments that I will make in this paper. I shall now turn to an analysis of the public reason approach to censoring CTs.

3. Cíbik and Hardoš on Conspiracy Theories: The Public Reason Approach

This section will be dedicated to analysing Matej Cíbik and Pavol Hardoš's article 'Conspiracy Theories and Reasonable Pluralism' (2022). Their article argues for the permissibility of the censorship of some conspiracy theories. I will contend that their proposal has an important flaw, although they are correct to suggest that the epistemic defects of conspiracy theories cannot alone justify using state intervention to counter them. Here is a summary of Cíbik and Hardoš's argument: first, they state that some conspiracy theories are false and irrational (2022, 447). Secondly, the epistemic defects of CTs alone cannot justify censorship or regulation. Thirdly, the criterion for deciding whether state action is appropriate depends on whether or not a conspiracy theory has morally 'unreasonable' content by the standards of public reason liberalism; if a conspiracy theory is unreasonable, censorship may be appropriate and justified in some cases.

Considering that Cíbik and Hardoš repudiate the idea that epistemic defects can be used as a sufficient justification for censorship, it would be unnecessary to analyse their take on the epistemology of conspiracy theories here (although their article does include a substantial discussion of this). This section will therefore concentrate only on the claim that we can use the criterion of ethical reasonableness to decide whether or not a CT can legitimately be subjected to state intervention. It is important to analyse this public reason approach, not least because of the ongoing popularity of public reason liberalism within political philosophy.

If we wish to assess their proposal, we first need to understand what Cíbik and Hardoš are referring to when they talk about 'reasonable' and 'unreasonable' people or doctrines. According to public reason liberalism, the most influential form of which was put forward by John Rawls, the ideal liberal state has legitimacy and stability only if its basic structure (and its most fundamental policies) can be justified to all 'reasonable' citizens in terms that they can harmonise with their comprehensive doctrines. These 'reasonable comprehensive doctrines' put forward different, and very often incompatible, positions on religion, ethics, metaphysics, and epistemology. However, there is no requirement to justify the liberal state or its key policies to 'unreasonable' citizens in terms that *they* can accept. One of the key problems for public reason liberals is that it is very difficult to offer clear and non-circular definitions of the terms 'reasonableness' and 'unreasonableness'. If the definition of 'reasonableness' is too broad, then it is unclear if public reason liberalism can really justify *anything*. If the definition is too narrow, there is a worry that public reason liberalism will exclude everyone who is not a political liberal, and thus that it will become question-begging (see Enoch 2015, 122; Mittendorf 2023, 4). To put it simply, because Cíbik and Hardoš subscribe to public reason liberalism, their article inherits this problem, at least as far as the issue of CTs is concerned. Let us see how this manifests itself. At first, Cíbik and Hardoš appear to endorse quite a broad definition of reasonableness, such as that put forward by Martha Nussbaum (Cíbik and Hardoš 2022, 454–455). According to this definition, one qualifies as a reasonable citizen if one respects other citizens as equals, and believes that society should be construed as a fair system of social cooperation (Nussbaum 2011, 33; cf. Rawls 2005, 483). Cíbik and Hardoš appeal to Jonathan Quong's list of criteria regarding who qualifies as an *unreasonable* person. Quong says that an unreasonable person 'rejects at least one, but usually several of the following: (1) that political society should be a fair system of social cooperation for mutual benefit, (2) that citizens are free and equal, and (3) the fact of reasonable pluralism' (Cíbik and Hardoš 2022, 455; Quong 2004, 315).

From this understanding of reasonableness, Cíbik and Hardoš start to build a case for the censorship of (some) conspiracy theories. As they write, 'consider the differences between two conspiracy theories. One claims that 30 Jewish bankers rule the world, the other that this power is held by a giant pipe-smoking rabbit. We hold that the former is clearly ethically unreasonable in the Quong-Nussbaum sense of the word, although the latter is not' (2022, 456). Put simply, conspiracy

theories that contain ethically unreasonable content can be subjected to state censorship (at least if there is good evidence that it will work), while ‘harmless’ CTs that do not contain unreasonable content should not be.

One immediate issue is that, if such a broad definition of reasonableness is used, it remains unclear how many conspiracy theories could actually be construed as ‘unreasonable’. Take Quong’s definition, for instance. Which political ideologies deny that, on some level, society should be a fair system of social cooperation for mutual benefit? Which theories would say that citizens are not free and equal in some important sense? Which theories deny the fact of reasonable pluralism *tout court*? It is very challenging to think of any political theories that would count as unreasonable because of Quong’s first criterion. The second criterion will catch more ideologies: obviously the Nazis do not believe that citizens are free and equal, nor do some religious fundamentalists, nor do authoritarian conservatives in the tradition of, say, Joseph de Maistre (although some fascists might claim to). But most others would. For instance, Stalinists and Maoists, despite supporting a system that we might term ‘hard authoritarianism’, would say that they believe in both freedom and equality. In fact, they would claim that they support Stalinism or Maoism precisely *because* a strong workers’ state led by a professional vanguard is the only way to eventually realise the values of freedom and equality in a stateless, classless communist society. Much of the ‘freedom’ and ‘equality’ in liberal capitalist states is, according to them, an illusion. Who, then, would fall foul of Quong’s third criterion? This all depends on how one defines ‘reasonable pluralism’. If a relatively broad definition of this term is used, then arguably most ideologies would pass this test. Most liberals, socialists, conservatives, anarchists, and communists believe that reasonable pluralism exists on some level. Again, even Stalinists and Maoists might not deny it. They would argue that a large degree of pluralism – larger than that permitted by supporters of public reason liberalism – will be allowed to flourish in the communist future. It is simply that, in order to prevent the resurgence of the dispossessed bourgeoisie, restrictions on reasonable pluralism are temporarily necessary.

Therefore, we can conclude that, if a broad definition of ‘reasonableness’ is used, hardly any CTs will be suitable candidates for state censorship. Is this what Cibik and Hardoš intended? This is improbable, for the following reasons. First, there are passages in their article in which it is implied that their conclusion is more wide-ranging than this (Cibik and Hardoš, 462–463). Secondly, and more importantly, public reason liberalism probably cannot use such a broad definition of ‘reasonableness’ and remain theoretically tenable. Cibik and Hardoš are writing as if, within public reason liberalism, virtually nobody counts as ‘unreasonable’ except Nazis, psychopaths, and maybe some other extremists. This, however, is not likely to be the case, because public reason liberalism of the sort that Cibik and Hardoš support cannot even get off the ground unless a narrower definition of reasonableness is utilised. Recall that, according to public reason liberalism, the liberal state and its fundamental policies should be justified to all reasonable citizens in terms that are harmonious with these citizens’ comprehensive doctrines. If we define the concept of reasonableness such that only the above-named and unapologetic fanatics are excluded from it, we will not be able to justify the liberal state and its policies in this way. As David Enoch puts it, ‘Can public-reason theorists retreat to a pre-theoretical understanding of the reasonable, perhaps so that you qualify as reasonable if your reasoning mechanisms are functioning well, or if you’re willing to listen to other views, and so on? They can, of course, but the price would be anarchism ... among those who are reasonable in just some very thin sense of this kind, *everything* is controversial’ (Enoch 2015, 122, italics original).

It would appear, therefore, inevitable that public reason liberalism will have to use a narrower definition of reasonableness. Many proponents of public reason liberalism have done this explicitly, including Jonathan Quong, Andrew Lister, and David Estlund, amongst others (see Estlund 2008; Quong 2011). Usually, the definition of reasonableness ends up being very narrow, particularly because public reason liberals often wish to make room for certain principles and policies that many seemingly reasonable people will reject. Narrowing the scope of ‘reasonableness’ is one way in which they can do this. As a general rule, even proponents of public reason liberalism who wish to be relatively accommodating exclude at least all *non-liberals* from their definition, and probably some in

the liberal tradition as well (for instance, comprehensive liberals such as Joseph Raz). The fact that proponents of public reason liberalism generally use a narrower definition of reasonableness is, I think, broadly accepted by both sides of the debate. Even when proponents of public reason liberalism respond to the objection that their definition of reasonableness is too narrow, they usually contend that it is *acceptable* to use a narrow definition, rather than deny that the definition is in fact narrow.

Given this, we can conclude that Cívik and Hardoš's public reason approach will necessitate a narrower understanding of 'reasonableness' than they explicitly acknowledge in their article. What would happen if they simply conceded this point? This would lead to clearly unacceptable, and seemingly illiberal, conclusions. Let us say, for instance, that all non-Rawlsians (including many liberals) are 'unreasonable'. If we were to accept Cívik and Hardoš's view that the state can sometimes legitimately censor unreasonable views, we would have to accept the conclusion that all non-Rawlsian views can, in some situations at least, be legitimately subjected to censorship. This view is at odds with the fundamental liberal commitment to free speech, because this would lead to the conclusion that almost *everything* can legitimately be censored. This position would potentially be even more extreme than Cass Sunstein and Adrian Vermeule's claim that the state can legitimately undermine 'demonstrably false' conspiracy theories that embody 'crippled epistemologies' (Sunstein and Vermeule 2009, 204). Sunstein and Vermeule suggest that the state can do this by secretly infiltrating groups of conspiracy theorists (particularly online), and posing as conspiracy theorists in order to feed these groups information intended to undermine the CTs from within. Cívik and Hardoš correctly criticise Sunstein and Vermeule for using an epistemic standard for judging whether CTs should be subject to censorship, as this sets the bar too low for state intervention, at least if we are writing from a liberal perspective. As Cívik and Hardoš write, 'if the state were to adopt an ambition to constrain crippled epistemologies, then its efforts would be much more wide-ranging than Sunstein and Vermeule recognise. Their proposal sets such a low bar to government intervention that it would encompass the subversion of religious doctrines, astrology and other less-than-fully verified beliefs held by its citizens' (2022, 458). However, if we define 'reasonableness' too narrowly, Cívik and Hardoš's own proposal runs into the very same problem. They implicitly set the bar for censorship far too low – perhaps even lower than Sunstein and Vermeule set it.

Here is the problem for the public reason view put forward by Cívik and Hardoš. On the broad definition of 'reasonableness' ostensibly endorsed by Cívik and Hardoš, most CTs will count as reasonable. It will not normally be clear that a particular CT is *unreasonable*, except in some very obvious cases. It seems as though, for the most part, only far-right conspiracy theories could sometimes legitimately be subjected to censorship. And, probably, not even all of these would count, particularly because conspiracy theorists could (and often already do) use coded language to disguise the morally unreasonable aspects of their CTs. More fundamentally, public reason liberalism as a whole almost certainly cannot remain tenable if such a broad definition of reasonableness is used. If we use a narrower definition of 'reasonableness', however, that will lead to the conclusion that a wide range of views can legitimately be subjected to free speech restrictions. This is not what Cívik and Hardoš intended, of course, and it appears at odds with the fundamental liberal commitment to protecting free speech. Even if Cívik and Hardoš could somehow find a 'Goldilocks zone' – i.e. a definition of reasonableness which is neither too broad nor too narrow – a large number of people would still presumably face state censorship. For instance, let us say that all citizens who are broadly liberal in their outlook (even comprehensive liberals) count as reasonable, but all non-liberal citizens are unreasonable. This would still be in tension with the liberal commitment to free speech, because it would mean that all Marxists, anarchists, non-liberal socialists, non-liberal conservatives, and some communitarians could legitimately be subjected to free speech restrictions (cf. Mittendorf 2023, 4). Moreover, as the writings of leading public reason liberals should make clear, the search for this 'Goldilocks zone' has so far been unsuccessful. As I noted above, most public reason liberals respond to the criticism that their definition of reasonableness is too narrow by claiming that this is not, in fact,

a problem. They do not usually respond by denying that their definition is narrow (see Lister 2018, 82–83).

The considerations raised here should not be taken to mean that public reason liberalism is false or indefensible. In fact, public reason liberals could respond relatively straightforwardly. All they need to do is argue that ethical unreasonableness should not, in itself, determine whether something should be subjected to state intervention. However, if public reason liberals were to opt for this path, it would defeat Cívik and Hardoś's account, because it would mean that the concept of reasonableness should not be used to determine whether a CT can be subjected to state censorship.

4. Conspiracy Theories, Harm, and Defamation

Considering the failure of using the concept of 'reasonableness' as a criterion for deciding whether or not the state can legitimately censor CTs, we should look to other concepts. It is first worth noting that, even if all other arguments fail, almost nobody believes that we should have a right to say whatever we want. Freedom of expression has limits, even for the strongest defenders of free speech. Liberals would generally agree, for instance, that speech which carries a high risk of violence or a threat to other people's basic rights can legitimately face censorship (this is a version of the famous 'Harm Principle'). For instance, even Ronald Dworkin – usually regarded as one of the most uncompromising liberal defenders of free speech – says that free speech can be restricted when there is a serious risk of harm to others (see Levin 2009, 357–358). Because this is relatively uncontroversial, I will not discuss this issue further. I will take it for granted that, sometimes, forms of speech which carry a significant risk of violence or harm to people's basic rights can legitimately be subjected to censorship. What I wish to do here is outline the additional, and somewhat less obvious, ways in which it might sometimes be acceptable to censor CTs. One of my aims here is also to find a feature more *distinctive* to CTs that might make them legitimate subjects of state censorship. Is it possible to do so?

I will argue that certain CTs contain content that is defamatory, particularly against groups that are societally vulnerable (e.g. at a high risk of racism or economic repression). Under some circumstances, it is acceptable to use civil laws (under a common law system) to enact sanctions against such conspiracy theories. Clearly, defamatory content cannot possibly be said to be distinctive to conspiracy theories. However, I will argue that certain conspiracy theories have a marked tendency to generate accusations that are defamatory. This will apply, in particular, to CTs that have one of the features that I mentioned earlier – namely a tendency to become 'evidentially self-insulating'.

To see why this is the case, we can turn briefly to a familiar objection to the practice of conspiracy theorising: the idea that 'unwarranted conspiracy theories' constitute 'degenerating research programmes'. This is related to the idea that conspiracy theories are unfalsifiable, and therefore should be considered disreputable – this is perhaps the most commonly encountered objection to CTs (Keeley 1999, 121). A full discussion of this epistemological issue is impossible within the bounds of this article, so I will attempt to be brief and outline the features of this concept that are the most morally salient. As I mentioned in the introduction, one familiar aspect of conspiracy theorising that many people find intuitively implausible is the tendency for conspiracy theories to become evidentially *self-insulating* (Napolitano 2021; cf. Clarke 2002). That is, if there is a lack of evidence for a conspiracy theory, or some evidence against it, this will often be interpreted as evidence *for* the conspiracy theory in question (or at least the counter-evidence will be written off as intrinsically unreliable). For instance, many conspiracy theorists will argue that evidence against the CT has been fabricated by the conspirators, or that we are unable to trust supposed experts because they are themselves being successfully deceived by the conspirators. This, in turn, leads to the CT becoming a 'degenerating research programme' because it is closed to all counter-evidence, and cannot make any fruitful predictions.

Some philosophers have argued against the idea that we should consider conspiracy theories to be epistemically faulty simply because they have the appearance of being evidentially self-insulating

theories, or degenerating research programmes. As Brian Keeley writes, 'Unfalsifiability is only a reasonable criterion in cases where we do not have reason to believe that there are powerful agents seeking to steer our investigation away from the truth of the matter' (Keeley 1999, 121). Lee Basham adds that 'While falsifiability is an appropriate criterion in the case of natural science, a standard that includes such a criterion [for assessing conspiracy theories] would rule out our initial inquiry into even authentic conspiracies like those involving President Richard Nixon or U.S. Army Colonel Oliver North' (Basham 2003, 93). However, for the purpose of this article, it is unnecessary to engage in a complex discussion of this epistemological issue. Regardless of whether the concept of falsifiability (or some variant thereof) is a good epistemic criterion to use for assessing CTs, it is undeniable that some CTs do have a tendency to generate further accusations of conspiracy in order to explain apparent counter-evidence. As Patrick Stokes has argued, while this might not constitute an *epistemic* fault, it does still have important *ethical* consequences. Stokes contends that, if a theory leads to multiplying accusations of conspiracy, we can consider that to be a 'moral cost' of the theory in question. To say that a person (or a group of people) has participated in a conspiracy is to make an accusation which, all things being equal, it would be morally better not to make (Stokes 2018). Conspiracy theories which lead to multiplying accusations of conspiracy are more ethically problematic in this respect than other sorts of theories. Of course, there might be good reasons to override these 'moral costs' of conspiracy theorising, and Stokes's concerns are clearly defeasible. Nonetheless, when making a decision about whether to engage in conspiracy theorising, we should consider these 'moral costs of conspiracy theory'. While Stokes's argument is intended to be ethical, rather than political, we can extend it so that it can sometimes be used as a justification for state intervention.

We can formulate a general criterion for deciding whether a conspiracy theory can legitimately be subjected to state intervention. A conspiracy theory can potentially be censored if it contains defamatory content against either individuals, or against vulnerable groups in society (much as any form of speech that is defamatory can legitimately face state censorship). We can also conclude that some conspiracy theories, because of their epistemic features, will have a tendency to generate a large number of conspiracy accusations, and therefore we are somewhat more likely to be justified in censoring CTs than most other forms of speech on this basis.

Which prominent conspiracy theories multiply accusations of conspiracy, as outlined above? There is not, of course, a definitive list of such theories – that would be impossible. It is, however, possible to give examples of how this could come about. Let us take a conspiracy theory about 9/11 which implicates Mossad as having foreknowledge. It is apparent that such a CT could have antisemitic connotations, and so we might think that it is already a good candidate for meeting the criteria above. However, the problem is sharpened if we consider the way in which the number of conspiracy accusations might increase if the theory is challenged. If little confirming evidence is found to support the idea that Mossad had foreknowledge of 9/11, or if apparent counter-evidence is offered (e.g. that Mossad did not have a strong enough motive to withhold foreknowledge of 9/11), many conspiracy theories will explain this counter-evidence away by saying that it has been falsified by the conspirators. They might alternatively argue that those who are presenting the counter-evidence are being successfully deceived by the conspirators. Of course, this would raise the question of who the conspirators are in this case. Not only could this theory lead to defamatory statements against individuals; it could also lead to potentially calumnious assertions about vulnerable groups. In this case, American Jews would probably be the group that is at the greatest risk of harmful discourse. It is important to consider that this group is already at a high risk of religious/ethnic hate crimes and intolerance. According to one estimate, American Jews are the victims of approximately 60% of all religiously motivated hate crimes in the United States (see Matza 2023).

Another example would be the claim that a large number of Muslims in New York celebrated during 9/11. This assertion has been made by influential figures, such as Donald Trump (who has a large support base), and thus it cannot be written off as a marginal case. Fact-checking organisations have repeatedly debunked this claim using 'normal' journalistic

methods (see Cassella 2015). Because of this, the claim already arguably constitutes defamation towards a vulnerable group in American society. However, once again, the problem is sharpened if we consider the fact that, in order to 'save' this theory from debunking, proponents might have to multiply their accusations of conspiracy. This could potentially lead to a number of additional calumnious assertions about a vulnerable group – e.g. that Muslims in the US are being protected by corrupt media or are being afforded 'special treatment by the establishment'. This CT could, therefore, be similar to the example of antisemitism outlined above.

Of course, these are only illustrations of how CTs could lead to these problems. Not all conspiracy theories are *bound* to do this, and it is not always wrong to suggest that some apparent counter-evidence is unreliable, or has been planted by conspirators. Nonetheless, as these examples show, there are some conspiracy theories that will probably become morally objectionable for the reasons outlined above.

However, it is not enough simply to show that some CTs are likely to contain defamatory content, and to assert that it is acceptable to enact laws against such defamatory content. We also need to provide a justification for believing that it is in fact acceptable to limit freedom of speech by censoring defamation. While it is common, in contemporary societies, including contemporary liberal democracies, to have laws that prohibit slander and libel, it is not self-evident that this sort of restriction on free speech has a good philosophical basis. Some free speech advocates would fundamentally object to laws against defamation (even civil laws). In the case of defamation laws, they might say, two rights are being balanced against each other: the right to free speech and the right to reputation (Barendt 2007, 198). Defenders of free speech might challenge the idea that the right to free expression – which they regard as absolutely fundamental for human freedom – can be balanced by considerations about 'reputation', which appear far less important, and perhaps even seem to be a relic from a bygone era (see Nicholas 1962, 215–18). Moreover, my argument suggests that not only individuals, but also vulnerable *groups*, should be permitted to take legal action when they suspect defamation. This might be seen as more controversial than the claim that individuals can legitimately take action against defamation; after all, some will be inclined to believe that individuals can be harmed in this way, whereas groups cannot.

Also, even if there is a good basis for saying that there should be legal sanctions for defamatory speech, it is not clear what *kinds* of legal sanctions should be utilised in these cases. It is, however, relatively easy to deal with this problem, at least for the purpose of the current discussion. In the passages that follow, I will assume the framework of the common law system that currently exists in many Western countries. I shall contend that conspiracy theories which contain defamatory content can be subject to 'civil laws' within a common law system, although they should not be criminalised in the same way as, for instance, speech that carries a high risk of violence. The main reason for this is that, as a matter of principle, legal sanctions should generally be in proportion to the offence committed. To introduce criminal penalties for defamation seems excessive, particularly considering that the harm in these cases is more specifically directed towards an individual or a group, rather than to individuals *and* society as a whole (as is supposed to be the case with criminal offences).³ So, taking all this into consideration, what can we say in defence of the claim that defamation is, at least in some cases, beyond the bounds of protected free speech?

A full discussion of this issue is not possible here, but I would argue that the considerations above do not land a decisive blow against defamation laws within a liberal democracy. Clearly, very few liberals believe in unrestricted free speech, and, as with any important policy decision, we should aim to strike the right balance between different rights and responsibilities. While we can, and should, certainly argue about the specifics of libel laws, there are good reasons to think that we should not give free speech absolute priority here. One important consideration is that libel laws are sometimes necessary for the preservation of political equality: substantive (instead of merely formal) equality of status, and fair equality of opportunity, amongst citizens (Anderson 1999). As Eric Barendt has written,

Some people, particularly those of a sensitive disposition, may be deterred from entering public life if they can never protect their reputation . . . In the absence of any legal restraints, the media would be free to say what they like about politicians and other public figures, and this would almost certainly encourage a concentration on scandal in preference to a serious discussion of political and social issues . . . in the long term the credibility of the press and other media would suffer, since the public would have no guarantee that their reports were likely to be accurate (Barendt 2007, 202).

It only requires a little imagination to see that always prioritising the right to free speech over the right to reputation could have negative, and potentially disastrous, consequences. As David van Mill has written in a general discussion of free speech, 'I have already suggested that all societies do (correctly) make some speech more costly than others. If the reader doubts this, it might be worth considering what life would be like with no sanctions on libellous statements, child pornography, advertising content, and releasing state secrets. The list could go on' (van Mill 2023, §1).

This analysis also shows why defamation that targets groups, particularly vulnerable groups, can be just as damaging as defamation directed towards individuals (and sometimes more so). False or misleading claims made about religious and ethnic minorities, for instance, could lead to broader discrimination against these groups. In many countries, the law already says that some organisations (such as corporations) can sue for defamation, and so the concept of 'group defamation' is not completely unfamiliar to us. It might be objected that, in many cases, these defamation laws benefit the rich, powerful, and famous, rather than the vulnerable. The answer to this, however, is not to do away with the concept of 'group libel' or 'collective libel' altogether, but to make sure that it primarily serves to protect the vulnerable in society – whether they are at risk because of economic status, religion, sexuality, ethnicity, gender, or beliefs.

The issue of defamation and free speech is multifaceted and complex (see Barendt 2007; Milo 2008). I cannot, therefore, claim to have provided anything like a full analysis. However, I have put forward some reasons to think that, in principle, we can legitimately enact civil defamation laws to protect the vulnerable.

5. The Conditions for State Intervention

Even if this is correct, it is clear that there should be some stringent restrictions regarding when it is acceptable to censor defamatory conspiracy theories. As with any other curtailment of free speech, we not only have to take into consideration the question of whether it is theoretically acceptable to enact censorship. We also have to decide whether it is an effective policy decision, and we should also consider whether restrictions on free speech might have other undesirable consequences. We must also, of course, always bear in mind that any power to restrict free speech could be abused. That is why I wish to propose two constraints on the censorship of defamatory conspiracy theories: an epistemological constraint and a practical constraint.⁴

The epistemological constraint says that, for a conspiracy theory to become a candidate for being censored, the theory in question should often lack plausibility. That is, the theory should be unable to enter the 'marketplace of ideas'. This is not to be confused with the claim that epistemic implausibility is a *sufficient* criterion for something to be worthy of censorship. I already considered and rejected this idea in the discussion of Cibik and Hardoš's public reason approach. However, it is clear that epistemic implausibility is often an important condition for the censorship of conspiracy theories. We usually want to avoid enacting laws against conspiracy theories that might actually be true, or which might at least contain a significant portion of the truth. Such conspiracy theories might be in the public interest, of course. The question of how much weight we wish to give to the 'epistemic constraint' will be context-dependent. When it comes to the question of defamation, however, the epistemic constraint will carry a lot of weight. For something to qualify as 'defamation', it must, by definition, be untrue (or probably so). The claim that Mossad had foreknowledge of 9/11 (and that evidence for this is being suppressed) cannot count as 'defamatory' unless we can show that it is probably false.

As this analysis shows, it is important to think carefully about whether a CT is epistemically plausible before censorship is considered. However, this raises the complex question of how we, as a society, are supposed to decide which conspiracy theories are plausible enough to enter the marketplace of ideas. This problem is exacerbated by the fact that there is no body of 'experts' that can be appealed to in the case of conspiracy theories. As M Dentith has argued, while there are experts on certain aspects of particular CTs, there are no people who can claim to be experts on conspiracy theories *per se* (Dentith 2018, 196). How, then, are the courts and other authorities supposed to settle these matters? My preference would be to err on the side of caution, and to apply only epistemic criteria that really do command widespread acceptance. For instance, in order for CTs such as Holocaust denial or Rwandan genocide denial to be true, the basic standards of evidence used in disciplines such as political science, history, philosophy, and forensic science would have to be wrong. These CTs violate widely accepted epistemic norms, which have been adopted for good reason over a long period of time. On the other hand, it is not so clear that even highly controversial conspiracy theories, such as those about the assassinations of JFK and Robert F. Kennedy, stand in violation of widely accepted epistemic norms in the same way. It would be impossible to provide a definitive list of theories that are 'plausible', of course, but we do have some general epistemic standards that we can draw upon to help us. This does not necessarily mean that everyone always applies these epistemic norms consistently, or that all people necessarily *think* that these epistemic norms are important. The key point is that almost all people make a tacit commitment to some of these epistemic standards, even if some do not realise that such standards apply to their favoured conspiracy theories (or other theories or beliefs).

To take one of the examples used previously, we could imagine that a group of Jewish Americans launched a defamation case alleging that American Jews were being defamed by the theory that Mossad was involved in 9/11 (and that this is being covered up by the US government). They would have to show that, on the balance of probabilities, such a conspiracy theory is false according to the epistemic standards that we all share. They would also have to show that the conspiracy theory has damaging effects on the group (a connection that might be difficult to prove or provide evidence for). It is an open question whether this could successfully be done, but this example gives some sense of how such a case could play out in practice.

Finally, we must consider the 'practical constraint': we should only consider enacting censorship if there is good evidence to suggest that it will be effective. State intervention that is likely to be ineffective, or which might lead to other problems, should of course be avoided. It is barely necessary to give examples here: state intervention has often been ineffective or harmful. The 'War on Drugs' and the 'War on Terror' are just two examples (see Costs of War Project 2021; Tonry 1994). There are also, for instance, serious doubts about the ability of states to outlaw pornography, at least without highly draconian measures to regulate the internet far more than they already do. Once again, there is no straightforward method for deciding when intervention will be effective, and when it will not be. This is another reason for being sceptical about many proposals for censorship. Just because some forms of discourse might appear to be defamatory, for instance, that does not mean that state intervention will be justified; we might reasonably think that the proposed 'cure' will be worse than the disease, in some cases.

After all these conditions have been put forward, it might seem relatively unusual for state intervention against conspiracy theories to be legitimate. After all, demonstrating that there is a causal connection between a conspiracy theory and violence or discrimination will be a difficult endeavour. It will be even more challenging to make the case that intervention will be effective, because in the majority of cases, there will be many variables to take into consideration. Take a theory like the 'Great Replacement' theory. It is reasonable to think that intervention against this theory could become counter-productive, not least because any attempt to censor this CT might affect freedom on the internet. There is even reason to doubt whether it would be possible to regulate such theories out of existence, at least while remaining within a liberal democratic framework. Therefore, we might think that, if we take the proposals in this article

seriously, relatively few CTs will become legitimate candidates for censorship. While this is true, it does not have to be seen as a problem. This article has sought to investigate when (and whether) the censorship of CTs is justified within liberalism. We have still succeeded in outlining the conditions that should pertain in the cases where state intervention is, in fact, justified (even if these are relatively unusual). Therefore, it would be wrong to assume that the conditions in this section are a problem for my proposal in general.

6. Conclusion

This article has analysed the issue of censoring conspiracy theories within liberalism. As well as criticising one of the existing approaches, I have put forward my own set of criteria for deciding whether a CT should be considered a good candidate for censorship. I provided some brief thoughts about what would count as an adequate definition of ‘conspiracy theory’ for my purpose here. I settled on a definition that marks out the class of conspiracy theories I am interested in, namely ones that (often) run contrary to official sources and which are frequently resistant to counter-evidence. I then put forward a critique of Cibik and Hardoš’s public reason approach to the censorship of CTs, and I suggested that the criterion of ‘reasonableness’ within public reason liberalism is not a good standard for deciding whether a CT may be subjected to state censorship.

After this, I outlined my own proposal. I argued that, in some cases, we might consider conspiracy theorising to constitute discriminatory speech against vulnerable individuals and groups, and we can legitimately classify such theories as a form of defamation. Consequently, the state may sometimes be permitted to enact civil laws in order to sanction such theorising. Finally, I contended that a number of conditions should be met before state censorship can be legitimately considered. These conditions probably mean that it will be relatively unusual for the censorship of CTs to be acceptable, but there are likely to be a number of important exceptions to this.

Notes

1. Existing liberal democratic states do not always protect fundamental rights and freedoms, and there are arguably many non-democratic aspects to all states. Therefore, questions about the degree to which a state is *really* a liberal democracy in the sense given here will sometimes be open to debate.
2. In his book *Conspiracy Theories* (2019), Cassam distinguishes between Conspiracy Theories (upper case C, upper case T) and conspiracy theories (lower case c, lower case t). According to Cassam, the former have the dubious features that I have mentioned, whereas the latter are simply theories or explanations involving conspiracies, many of which Cassam considers to be true and/or well justified.
3. This is a matter of degree. Some civil wrongs can, at least indirectly, have harmful effects on society as a whole, while some criminal offences may have relatively little overall impact. All we can say is that civil wrongs *usually* have *more* of a harmful impact on private individuals and groups than on society as a whole.
4. Although it is not something that I explicitly argue for in this article, I take it that these constraints often also apply to censorship in cases where there is a serious risk of violence. However, this section deals with the issue of defamation.

Acknowledgments

I would like to thank the participants at the 2nd International Conference on the Philosophy of Conspiracy Theory for comments on an earlier draft of this paper, and I would also like to thank members of the University of Bristol’s Postgraduate Work in Progress Seminar for their comments. I thank two anonymous referees for providing valuable advice that helped me to improve the paper. Finally, I would like to thank Seiriol Morgan and Martin Sticker for their help.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

I have no funding to declare for this article.

Notes on contributor

Fred Matthews is studying for a PhD in philosophy at the University of Bristol, specialising in moral and political philosophy. He previously completed an MPhil in political theory at the University of Oxford, and a BA in philosophy at the University of East Anglia. His research interests include general political philosophy, liberalism, applied epistemology, and environmental ethics.

ORCID

Fred Matthews  <http://orcid.org/0000-0002-7981-4482>

References

- Anderson, Elizabeth S. 1999. "What is the Point of Equality?" *Ethics* 109 (2): 287–337. <https://doi.org/10.1086/233897>.
- Barendt, Eric. 2007. *Freedom of Speech*. Oxford: Oxford University Press.
- Basham, Lee. 2003. "Malevolent Global Conspiracy." *Journal of Social Philosophy* 34 (1): 91–103. <https://doi.org/10.1111/1467-9833.00167>.
- Brisson, Susan. 1998. "The Autonomy Defense of Free Speech." *Ethics* 108 (2): 312–339. <https://doi.org/10.1086/233807>.
- Cassam, Quassim. 2019. *Conspiracy Theories*. Cambridge: Polity Press.
- Cassella, Megan. 2015. "Trump: 'I was 100 Percent Right' on 9/11 Muslim Remarks." *Reuters*, November 30. Accessed October 29, 2024. <https://www.reuters.com/article/world/trump-i-was-100-percent-right-on-911-muslim-remarks-idUSKBN0T058/>.
- Cibik, Matej, and Pavol Hardoš. 2022. "Conspiracy Theories and Reasonable Pluralism." *European Journal of Political Theory* 21 (3): 445–465. <https://doi.org/10.1177/1474885119899232>.
- Clarke, Steven. 2002. "Conspiracy Theories and Conspiracy Theorizing." *Philosophy of the Social Sciences* 32 (2): 131–150. <https://doi.org/10.1177/004931032002001>.
- Coady, David. 2019. "Conspiracy Theories and Official Stories." In *Conspiracy Theories: The Philosophical Debate*, edited by David Coady, 115–127. Abingdon: Routledge.
- Costs of War Project. 2021. "Costs of the 20-Year War on Terror: \$8 Trillion and 900,000 Deaths." *News from Brown*, September 1, 2021. Accessed January 20, 2024. <https://www.brown.edu/news/2021-09-01/costsofwar>.
- Dentith, M. 2014. *The Philosophy of Conspiracy Theories*. New York: Springer.
- Dentith, M. 2018. "Expertise and Conspiracy Theories." *Social Epistemology* 32 (3): 196–208. <https://doi.org/10.1080/02691728.2018.1440021>.
- Dentith, M. 2023. "Some Conspiracy Theories." *Social Epistemology* 37 (4): 522–534. <https://doi.org/10.1080/02691728.2023.2173539>.
- Enoch, David. 2015. "Against Public Reason." In *Oxford Studies in Political Philosophy, Volume 1*, edited by David Sobel, Peter Vallentyne, and Steven Wall, 112–142. Oxford: Oxford University Press.
- Estlund, David M. 2008. *Democratic Authority: A Philosophical Framework*. Princeton: Princeton University Press.
- Gallie, Walter Bryce. 1956. "IX.—Essentially Contested Concepts." *Proceedings of the Aristotelian Society* 56 (1): 167–198. <https://doi.org/10.1093/aristotelian/56.1.167>.
- Keeley, Brian L. 1999. "Of Conspiracy Theories." *The Journal of Philosophy* 96 (3): 109–126. <https://doi.org/10.2307/2564659>.
- Levin, Abigail. 2009. "Pornography, Hate Speech, and Their Challenge to Dworkin's Egalitarian Liberalism." *Public Affairs Quarterly* 23 (4): 357–373.
- Lister, Andrew. 2018. "The Coherence of Public Reason." *Journal of Moral Philosophy* 15 (1): 64–84. <https://doi.org/10.1163/17455243-00002431>.
- Matza, Max. 2023. "FBI Director Warns Antisemitism in US Reaching 'Historic levels.'" *BBC*, October 31, 2023. Accessed January 20, 2024. <https://www.bbc.co.uk/news/world-us-canada-67281042>.
- Milo, Dario. 2008. *Defamation and Freedom of Speech*. Oxford: Oxford University Press.
- Mittendorf, Will. 2023. "Conspiracy Theories and Democratic Legitimacy." *Social Epistemology* 37 (4): 481–493. <https://doi.org/10.1080/02691728.2023.2172700>.
- Napolitano, M. Giulia. 2021. "Conspiracy Theories and Evidential Self-Insulation." In *The Epistemology of Fake News*, edited by Sven Bernecker, K. Flowerree Amy, and Thomas Grundmann, 82–105. Oxford: Oxford University Press.
- Nicholas, Barry. 1962. *An Introduction to Roman Law*. Oxford: Clarendon Press.

- Nussbaum, Martha. 2011. "Perfectionist Liberalism and Political Liberalism." *Philosophy & Public Affairs* 39 (1): 3–45. <https://doi.org/10.1111/j.1088-4963.2011.01200.x>.
- Quong, Jonathan. 2004. "The Rights of Unreasonable Citizens." *The Journal of Political Philosophy* 12 (3): 314–335. <https://doi.org/10.1111/j.1467-9760.2004.00202.x>.
- Quong, Jonathan. 2011. *Liberalism without Perfection*. Oxford: Oxford University Press.
- Rawls, John. 2005. *Political Liberalism: Expanded Edition*. New York: Columbia University Press.
- Shields, Matthew. 2022. "Rethinking Conspiracy Theories." *Synthese* 200 (331). Accessed January 20, 2024. <https://doi.org/10.1007/s11229-022-03811-x>.
- Stokes, Patrick. 2018. "On Some Moral Costs of Conspiracy Theory." In *Taking Conspiracy Theories Seriously*, edited by MRX. Dentith, 189–202. London: Rowman & Littlefield.
- Sunstein, Cass, and Adrian Vermeule. 2009. "Conspiracy Theories: Causes and Cures." *The Journal of Political Philosophy* 17 (2): 202–227. <https://doi.org/10.1111/j.1467-9760.2008.00325.x>.
- Tonry, Michael. 1994. "Race and the War on Drugs." *University of Chicago Legal Forum* 25:25–82. <https://heinonline.org/HOL/P?h=hein.journals/uchclf1994&i=29>.
- van Mill, David. 2023. "Freedom of Speech." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman. Stanford: Stanford University. <https://plato.stanford.edu/archives/win2023/entries/freedom-speech/>.
- Waldron, Jeremy. 2012. *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press.
- Xin, Wendy. 2024. "Censorship Bubbles Vs Hate Bubbles." *Social Epistemology* 38 (4): 446–457. <https://doi.org/10.1080/02691728.2023.2274324>.