Episteme (2025), 1–20
doi:10.1017/epi.2025.10082

episteme

ARTICLE

On the Perils of Engaging

Rory James Aird

University of Glasgow, Glasgow, UK
Email: r.aird.1@research.gla.ac.uk

(Received 18 October 2024; revised 8 August 2025; accepted 19 August 2025)

Abstract

Recent work in social epistemology has discussed obligations to engage with challenges to our beliefs like climate change denial or anti-vaccine sentiment, and the potential benefits and dangers for both the engager and the engaged from doing so. The spotlight being trained here, however, has elided a key issue: the possible risks from engaging for third-party observers, not merely the engager and the engaged. In this paper, I argue that not only are these risks an underappreciated aspect of engaging that should be discussed, but also their neglect is especially concerning as the potential negative epistemic fallout threatens to overwhelm any possible benefits that may be gained from engaging, regardless of how the engagement actually goes. I close by drawing out the theoretical and practical implications from this and sketch a few strategies to conceivably avoid said risks.

Key words: Objecting; engaging; epistemic duties; misinformation; negative epistemic effects

1. Introduction

What can one do when an interlocutor asserts that climate change is a hoax? What *ought* one to do when an interlocutor denies the Holocaust? Why should one engage with an anti-vaxxer? Why *shouldn't* one engage with a flat Earther?[1] These are pertinent and difficult questions, and ones that some philosophers have recently taken up in the social epistemology literature (McCormick 2023; McKenna 2023; Cassam 2019; Fantl 2018; Lackey 2020a, 2020b; Battaly 2018, 2021; McIntyre 2018, 2021). Naturally, there is disagreement between those working on these areas, but it is illuminating to highlight where the spotlight has been focused and the trends of agreement that can be flagged.

First, a key reason for engaging with those who believe and assert falsities is the hope that by doing so, one might be able to change their mind and get them to update from a false belief to a true one. As Quassim Cassam (2019) writes, 'If one can't be bothered to argue against conspiracy theories one can hardly complain if people end up believing them' (117). This thought is a driving force behind the literature (in particular those noted above) and understandably so; after all, however sceptical one might be of the

[1]To be clear, in this paper, I am focused on citizen/peer interactions in day-to-day circumstances and not the *scientific experts/lay people* interactions that characterise, for instance, Nancy Rosenblum's (2020) 'witnessing professionals' and their plausible duties to speak out and engage with climate change deniers.

prospects of this result, it's plausible that there is at least *a chance* of changing someone's mind for the better, and this chance is much higher conditioned on engaging than not.

It's unsurprising, then, that some view engaging with challenges to our beliefs as *obligatory*. John Stuart Mill's *On Liberty* (1859) is really the *locus classicus* of this thought: he argues that only by engaging with those we disagree with can we garner a myriad of positive effects – make sure that which we think is true really is so, stop commonly held beliefs from becoming unimpeachable 'dead dogmas', and gain further understanding of our own reasons for our beliefs.[2] He even thinks engaging with opinions that are false can be useful as they very often 'contain a portion of the truth', (118) which can only be appreciated and taken up via such engagement or discussion.[3] Cassam (2019) argues along similar lines, saying that if we don't engage with challenges to our beliefs, we lose our right to our justification of them and thus lose our knowledge, hence the obligation. The idea is that if one cannot respond to a counterargument to one's belief, then one simply is not justified in holding it, and so it no longer constitutes knowledge.[4] Jennifer Lackey's (2020b) duty to object is in a similar ballpark inasmuch as she takes it as an important factor to flag our disagreement to help prevent present bystanders from taking up the asserted falsity, although she downgrades the obligation somewhat to an *imperfect* duty where one must only engage sometimes, and how great this normative pressure is depends on a couple of different social and practical factors. McIntyre (2018, 2021) also acknowledges this point about public dismissal of claims helping to insulate against those coming across such claims for the first time getting the wrong impression – that is, that these are orthodox opinions.

The foregoing discussion essentially captures the key (epistemic) positives of engaging when it goes well. We can break the potential avenues of positive effects down into three areas: (i) positive effects for the engager (the one replying to the individual making the problematic claim), (ii) positive effects for the engaged (the one initiating the interaction by making the problematic claim), and (iii) positive effects for third parties or society as a whole. For (i), we have a gain in the understanding of one's own reasons for their beliefs, learning the whole truth about the topic at hand, the right to maintain our knowledge, and/or the fulfilment of an epistemic duty (Mill *et al.* 1859; Cassam 2019; Lackey 2020b). For (ii), we have the potential of their updating to a true belief (McCormick 2023). And for (iii), we have a general increase in the stock of true and secure beliefs in the epistemic environment (Mill *et al.* 1859; McKenna 2023) and an attempt to impede bystanders from taking up the false belief (Lackey 2020b; McIntyre 2018, 2021).

Not everyone is so optimistic about the overall prospects of engaging with challenges to our beliefs, however, and have turned to discussing the possible *negative* epistemic effects from doing so.[5] Jeremy Fantl (2018), for instance, is sceptical of this strong

[2]See, for instance, McKenna (2023, chap. 6), Fantl (2018, chap. 5), or Macleod (2021) for more on Mill's defence of an obligation to engage.

[3]While this may well be true in some (or even most) circumstances, I do not think it is particularly applicable to the cases (soon-to-be) employed in this paper. Indeed, Mill's thought here implies that there may be truth on both sides of, say, the Holocaust and its denial, which strikes me as rather dangerous. My thanks to an anonymous reviewer for urging me to discuss this.

[4]See Aikin (2010) for more on such a dialectical requirement for justification.

[5]For a discussion of potential costs from engaging in general (and indeed whether to even engage or not in the first place), see Paglieri (2013). While he does discuss 'negative consequences', he is not narrowly interested in epistemic ones as I am and is generally more focused on *practical* costs such as putting strain on one's relationship with the interlocutor at hand (156) or 'social image [and] self-esteem' (160). Nevertheless, it is interesting that we both employ a sort of consequentialist, expected value framework for looking at arguments and engagements with opinions contrary to our own.

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

Millian position of open-minded engagement. He argues that such engagement with that which we know to be false can in fact lead to *losses* of knowledge for the engager as their confidence may be distorted if they cannot respond effectively to their interlocutor. The thought is that even if one knows an argument against, say, the reality of the Holocaust is misleading, an inability to effectively respond to it and demonstrate precisely where and why it goes wrong can still adversely affect one's confidence in their knowledge of the Holocaust and perhaps even result in the loss of said knowledge for them. For this reason, then, he argues that one ought not engage open-mindedly with asserted falsehoods to try to insulate against this concern. He does think that we can perhaps engage *closed-mindedly* with false assertions – viz., engage without the possibility of changing our mind – but is even somewhat sceptical of the prospects of this strategy. McCormick (2023), Battaly (2021), and McIntyre (2020) agree with Fantl about the dangers of open-minded engagement but are more optimistic about the potential of closed-minded engagement, advocating for this strategy in engaging as a key method for changing the minds of those who believe things like climate change hoaxes or Holocaust denial. Once one is closed-minded to the possibility of changing their mind in response to discussion with a fringe believer, the thought goes, all possible risk from engaging has been eliminated (in terms of the engager's own knowledge at least), and the only way is up, inasmuch as there is now a chance of changing the fringe believer's mind to a true belief – or at least a chance of getting them to jettison their false one.

McKenna (2023) has a different worry with Millian and Cassamian open-minded engagement: he argues that some stand to lose a lot more than others by engaging open-mindedly with those that disagree with them – specifically, those who may suffer a *testimonial injustice* if they engage.[6] The idea is that a woman of colour may be shut down if she testifies to her male boss about, say, her experience of sexual harassment in the workplace and thus may end up losing her knowledge of what happened due to her not being respected as a knower. Thus, McKenna argues that she *cannot* have an obligation to engage with such a challenge to her beliefs.

Possible negative effects on the *engaged* from entering into a discussion have not been given much of a philosophical treatment – which isn't overly surprising. After all, the literature is generally focused around engaging with those who express false beliefs; thus, they're already in a bad situation, so how it might get worse for them doesn't seem especially important. Nevertheless, there is work in social psychology and political science that bears mentioning: backfire effects where being presented with counter-vailing evidence can make one more secure in their (false) belief (Cohen 2005; Nyhan and Reifler 2010, 2015), or polarisation which again can make their beliefs more extreme and secure (Sunstein 2002).

This discussion captures the core strands of the literature so far in terms of possible *negative* epistemic effects from engaging. For the engager, they may lose their knowledge if they engage in an open-minded way and find that they do not have effective responses to the engaged's arguments, or they may suffer epistemic injustices depending on their social situation. For the engaged, they may become further entrenched in their problematic beliefs. This is the dark reflection of the potential *positive* effects I outlined above. Something interesting to note, however, is that there is one part that *hasn't* been discussed: potential negative epistemic effects on other people outwith the engager and engaged – third parties and/or society as a whole. It might be helpful to present this visually (Table 1):

[6]See also Terzian and Corbalán (2021) for more on this, specifically about vaccine hesitancy/denial.

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

Table 1. Engaging literature overview.

	Possible positive epistemic effects	Possible negative epistemic effects
Engager	Fulfil an epistemic duty (Lackey), learn the full truth and your own reasons better (Mill), maintain right to knowledge/ justification (Cassam)	Lose knowledge (Fantl, McCormick, Battaly, McIntyre), suffer epistemic injustice (McKenna)
Engaged	Take up true belief/get rid of false belief (All)	Polarisation (Sunstein), backfire effect (Nyhan and Reifler)
Third-party observers/ society	Increase stock of true beliefs in society (Mill), prevent bystanders from taking up false belief (Lackey), prevent confusion in others (McIntyre)	?

This paper aims to fill the question mark cell.[7] I do not want to fill this lacuna for mere posterity, however. I will in fact argue that it is especially concerning that this area has been so neglected in the literature as it has the potential to be by far the most problematic in terms of negative epistemic effects. So much so, in fact, that it could threaten to overwhelm any of the possible positives that we have so far discussed. Central to this idea is that much of the literature (in particular, the majority of those cited in the *Engaging literature overview* table) has elided key questions relating to *how* engagements can actually go; per the above, one could be forgiven for thinking that engaging is often simply a quick dismissal of 'that's wrong' which either works in changing the target's mind or doesn't, and then we all get on with the rest of our day. Engaging is often far more dialogical than that,[8] and this gives scope for engaging to go far more wrong than has been previously appreciated by many of these authors.

Specifically, I will here focus on engaging with what I am calling *controversial false assertions*. I will not precisely define these, but the class of assertions I have in mind are ones such as 'climate change is a hoax', 'the Holocaust didn't happen', or 'Covid vaccines are deadly'. These examples all overlap with Neil Levy's *bad beliefs*: 'a belief that (a) conflicts with the beliefs held by the relevant epistemic authorities and (b) held despite the widespread public availability either of the evidence that supports more accurate beliefs or of the knowledge that the relevant authorities believe as they do' (2021, xi). Nevertheless, I take my *controversial false assertions* to differ insofar as 'bad beliefs' for Levy are not always false unlike the cases I am interested in. He also generally takes those who believe bad beliefs to be rational which is not something I want to commit to, nor is it important for my purposes here.[9] Plausibly, the nearest comparison to my *controversial false assertions* is what Fantl calls 'controversial propositions', for which he gives examples of 'the theory of evolution, psychic phenomena, the efficacy and danger of vaccines, convoluted conspiracy theories, repugnant moral positions, the existence of God, and whether the Holocaust occurred' (2018, 28). As he is interested in engagers losing *knowledge* in relation to these subjects, he presumably takes the

[7]Fantl (2018) discusses an adjacent area in the final chapter of *The Limitations of the Open Mind*. Even so, he is mainly focused on (psychological and intrinsic) harms from inviting what he calls 'problematic speakers' to campuses. This is similar to a worry Levy (2019) raises about platforming problematic speakers creating misleading higher-order evidence about their credibility. These are quite different concerns from the sorts of cases I will discuss and the more (first-order) *epistemic* worries I'm interested in – in short, I am not thinking about platforming specifically, nor psychological or higher-order evidence problems.

[8]Again, see Paglieri (2013). Cf. Dutilh Novaes (2023) for a philosophical model of argumentation.

[9]Granted this diagnosis of rational does follow if you are working with a subjective Bayesianism framework as Levy is, but I do not want to commit to that either.

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

controversial propositions of interest to be false, thus aligning more closely with me than Levy did. On the other hand, his inclusion of 'psychic phenomena' and 'the existence of God' is certainly outwith the bounds of my discussion here, so I do not want to adopt this wholesale. Overall, I think the key point here is that I will follow *both* Levy and Fantl in not worrying about giving a precise conceptual analysis of my umbrella term for the examples employed in this paper. Instead, we can leave the general category at a more intuitive level, as the arguments and examples themselves are of greater importance than any definition, while noting that the exact characterisation likely lies somewhere around those of Levy and Fantl just discussed.[10]

Per the extant literature, one gets the sense that, if the engager is sufficiently closed-minded and any threats of epistemic injustice are accounted for, we have exhausted the ways in which engaging with or objecting to such controversial false assertions can go epistemically wrong.[11] All that's left now is either the engaged changes their mind, or they maintain their false belief – and in the latter case nothing has been lost anyway. I will show that there are in fact a variety of ways that such a discussion can go, where, for example, an engager can 'lose' the engagement, and that this can have wide-ranging negative epistemic effects not even primarily on the engager and the engaged but rather on *third-party observers* and *society at large*. What's more, the epistemic position prior to engaging with a challenge to our beliefs is such that it's very difficult for a potential engager to know how such a discussion might go, meaning that even the mere potential of a catastrophic engagement threatens to overwhelm any potential benefits that might be garnered in the good situation. In this sense, my argument here is certainly a pessimistic one insofar as I will argue that engaging with controversial false assertions is an extremely risky activity and one that we should be reticent about doing in a lot of ordinary circumstances. Nevertheless, all hope is not lost, as I will close with some suggestions of how to engage and plausibly avoid these problems.

2. Cases

I will now go through a few different cases, all of which have the same structure: someone putting forward a controversial false assertion, someone engaging with and objecting to this assertion, and an audience of third-party observers. The cases are primarily designed so as to make it easy to see the intuitions being pumped and the arguments being drawn from them, but they have the added virtue of being closely analogous to any such discussions that might take place *online* – that is, on social media. I take it that this is a further help as we might think that a lot of the time people are faced with controversial false assertions in these settings, where there is an audience of third-party observers, so any conclusions drawn here from the cases should apply more broadly than merely the exact in-person scenario discussed in the vignette. In particular, such conclusions are relevant to online arguments because the general set-up will often be very similar insofar as there is an audience not participating but bearing witness to the engagement (i.e., the users scrolling past such a discussion) and then the people actually engaged in the argument. Moreover, as my focus in this paper is primarily on third-party bystanders and their role whether online or in person is by stipulation non-participatory, then drawing analogies between the two is reasonable and legitimate.

[10]My thanks to an anonymous reviewer for urging me to give more detail on this.

[11]Again, my focus here is purely on the epistemic dimension as there are of course many *non-epistemic* things that can go wrong (e.g., violent reactions and non-epistemic issues in terms of respect that cannot be accounted for as cases of epistemic injustice). See Paglieri (2013) for more examples of potential non-epistemic costs.

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

Each situation will have a different outcome: I will consider a case where our protagonist (the engager) 'loses' the engagement, one where we might call it a stalemate, and one where the protagonist looks to be successful. In the first two, I will argue that we have good reason to be concerned about negative epistemic effects in these scenarios, while even the success case is not an unadulterated good as we may have initially thought. Finally, let's stipulate that these are what McCormick and McKenna might think of as 'dream scenarios' insofar as we will assume in each case that the engager is sufficiently closed-minded so as not to risk their own knowledge,[12] and any threats of testimonial injustice are not relevant. The idea is that even with all these positive measures secured, there are still large concerns about negative epistemic consequences. Now that we have all our pieces in place, let's begin with the first case.

2.1. Losing

> *Dinner*. Alex is out for dinner with some colleagues. During the mains, one colleague, Bob, starts espousing the view that climate change is a hoax. He presents some 'evidence' and generally puts forward the case of climate change denial well. Alex knows this view is false and so engages with Bob, objecting to his false assertion. Unfortunately, Bob has come rather well prepared and responds very effectively to Alex. He attacks her arguments cleverly, rejects her evidence, provides 'evidence' that she does not know how to refute, and generally seems to prove her wrong (at least from the perspective of the onlookers).

First, it does not strike me as at all controversial to say that this engagement did not go well – in fact, it's hard to imagine how it could have gone worse. Alex engaged because she knew Bob's assertion was false and perhaps hoped to change his mind, but instead found her own view attacked effectively and forcefully in such a way that she was incapable of addressing it. Now, we have already stipulated that Alex's engagement was a closed-minded one, so let's grant that her belief or knowledge that climate change is real remains unshaken. We also stated that there was no risk of testimonial injustices taking place either. Where then might we locate the negative consequences that seem intuitively endemic in this case? One might suppose that Bob will become even more secure in his false belief after such a discussion, but he presumably already believed it strongly enough to assert it to an audience, so this problem does not seem particularly pressing either. Our real concern, I propose, relates to the *third-party observers*.

If we assume that the others at dinner follow general societal trends, then let's say that most believe climate change is real (around 70%), some are unsure or suspended on the question (around 15%), and some are in agreement with Bob (also 15%) (Ballew *et al.* 2019; Tranter and Booth 2015). While we stipulated that *Alex* was closed-minded when she opted to engage with Bob, there was no such guarantee for the others present. Therefore, when Alex's arguments were effectively attacked, and her evidence rejected, and so on, we can easily predict that this could shake any of the third-party observers' beliefs or knowledge inasmuch as their justification for their belief or knowledge could

[12]One might argue that one can't simply *choose* to be closed-minded like this, and repetition and fluency effects (see, e.g., Levy (2017)) can result in false belief uptake or loss of knowledge irrespective of the engager's intentions. This may well be true, and if so it is merely more grist for my mill of the perils of engaging, but I nevertheless set aside such worries here as it's all the more impressive that my argument goes through *without* needing to wield this extra psychological weaponry.

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

be defeated by these (misleading) arguments.[13] In addition, those that were unsure or suspended now have good reason to come down on one side of the argument – after all, they just saw the position of climate change denial strongly come out on top in a public engagement.[14] What's more, this is all arguably *rational* for these parties. In fact, if we want to argue reasons *why* we ought to or can engage with such assertions to change the interlocutor's mind and/or prevent others from getting the wrong impression (in other words, *appeal to their rational capacities*), then there's no principled reason to suggest that the reverse can't be true if the discussion goes poorly – viz., these third parties losing their knowledge or even coming to believe the *opposite* of what we wanted when we engaged.[15]

What upshots can we take from this? Well, with this case we are already in somewhat uncharted territory insofar as none of the extant scholarship really discussed *how* an engagement could actually go, but it's quite clear that the foregoing literature does not really have the resources to explain this case anyway. McCormick's and McKenna's worries have been accounted for with closed-mindedness and no epistemic injustice, respectively. Lackey's duty to object is a deontic one, so presumably she's not overly concerned with actual consequences or outcomes. Mill's diagnosis here would likely be a rather odd one in that I suspect he would have to say that everyone should lower their credence in climate change's reality following such a discussion. Cassam almost seems to reject this sort of scenario outright as he seems to think it's rather easy to argue against people like Bob.[16] Fantl comes closest to saying *something* about such a case, but even he is more focused on the effects on the *engager* (Alex) – and we already stipulated that she won't suffer any negative epistemic effects here.

All of this is to say that I think it's rather clear that cases like *Dinner* highlight an important dimension of engaging with controversial false assertions that has clearly been

[13]One might suppose that audience members will all react differently to the exchange, so – without more detail on audience biases, attributions of authority, etc. – we have no way of knowing who 'won'. Moreover, someone's perception of who 'won' is different from finding the arguments convincing and/or having one's beliefs shaped by the debate. Starting with the latter point, while I agree that perceptions around winning and changing of minds *can* come apart, this will not be the standard case. This is borne out in empirical data; for example, Aalberg and Todal Jenssen 2007) found that 'the outcome of the debate in terms of who was perceived as winner made a difference [to beliefs around issues]'. (131) Or Wang et al. (2017), whose model predicts whom audiences ultimately vote as winning a debate, find that 'the inherent persuasive effect of argument content plays a crucial role in affecting the outcome of debates' (220) and that 'Our model also shows that winners use stronger arguments'. (229) Thus, coming back to the former objection, empirical data suggest we do have ways of knowing who won – and these ways relate specifically to the strength of arguments in play. Further empirical data concurs on this: Ettensperger et al. (2023) found that 'predispositions may be a strong filter but are far away from being the only determinative factors for the perception of the audience' (14), while Nwokora and Brown (2017), employing as a case study the Obama/McCain 2008 debate, found that when a debate win is convincing, it breaks partisan lines. Thus, we can know who won a debate (audiences are fairly homogenous on this), and who won will generally be determined by strength of argument, as my vignettes assume. My thanks to an anonymous referee for pressing me on this point.

[14]There's an interesting parallel here with a discussion Timothy Williamson has in his (2020) *Philosophical Method*: 'When two senior philosophers argue some issue out with each other in public, with prestige at stake, it is often clear that neither of them will ever persuade the other; even so, it is not a waste of time if there are uncommitted students in the audience, making up their own minds as to which of the two is having the better of the argument'. (41).

[15]For simplicity's sake, I'm not going to precisely discuss credences here, instead focusing mainly on belief, justification, and knowledge, but note that a lot of this discussion can equally be couched in terms of bystanders lowering their credence in climate change's reality and raising their credence in climate change being a hoax – clear negative epistemic effects.

[16]See the quote above on page 2. Also see Cassam (2019, chap. 5).

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

missed so far in the literature. Not only is it distinctly probable that third-party observers will suffer the sort of negative epistemic effects outlined above, but such effects are also far more concerning than the worries the literature has previously touched on just in terms of pure *volume*. Even if only two people in the dinner audience lose knowledge or inculcate a false belief, that's already more epistemic damage than if we only talk about potential effects on the engager, so this is clearly a problem that we must discuss and take seriously.

Nevertheless, my opponent might suggest that this situation where the engager undergoes such a dialectical trouncing like *Dinner* is rather rare.[17] Or, at the very least, there are certainly more ways a discussion can go than the way we just looked at. So, let's consider a different variation of *Dinner* and see what conclusions can be made.

2.2. Stalemate

> *Dinner**. Claire is out for dinner with some colleagues. During the mains, one colleague, Derek, starts espousing the view that climate change is a hoax. He presents some 'evidence' and generally puts forward the case of climate change denial well. Claire knows this view is false and so engages with Derek, objecting to his false assertion. They have a long back and forth, ending at a sort of impasse from the perspective of the third-party observers, where neither presents a slam-dunk argument or unanswerable piece of evidence (and neither concedes nor proclaims the other to be correct).[18]

Let's start with the obvious: this outcome is clearly not as bad as the previous case. Nevertheless, I propose that it's still not good enough and remains well placed to have negative epistemic effects on the third-party observers. To see why this is so, we have to consider the point of our engaging with false assertions. From our discussion so far, we can draw out three discrete but related elements: (i) to (hopefully) change the false assertor's mind, (ii) to display our own disagreement,[19] and (iii) to inhibit any third-party observers from taking up the false assertion. These are all connected by the central thread of doing everything we can to limit or prevent inculcation of false belief. The

[17]As it happens, I am rather sceptical of this thought. Studies have shown that climate change scepticism among US Republicans tends to *increase* with greater scientific literacy (Funk and Kennedy 2020); therefore, we can reasonably infer that often those asserting falsehoods about climate change can often have a lot of descriptive knowledge about climate science and related matters and thus may well be in a good position to argue their case effectively.

[18]One might think here that it is easy from the point-of-view of the universe to call Claire and Derek's discussion a tie (or indeed the others' wins or losses), but bystanders could draw their own conclusions about which side came across better and so think Claire 'won' and plausibly infer that climate change is clearly real. I think this is right, but obviously the reverse applies too: in a stalemate scenario, a bystander could think that Derek 'won' and draw the conclusion that climate change is indeed a hoax. Thus, let's assume these two situations cancel each other out. More generally, I will assume that bystanders' idiosyncrasies are not more frequently biased in either direction and so will set aside this worry of audiences not assessing evidence homogenously. On the other hand, however, this may be *too* charitable to my opponent insofar as remaining 'undecided' on a controversial issue (i.e., one on which there exists a broad societal consensus) suggests that the bystander already holds strong beliefs in favour of the false assertion. Thus, there is even more reason to think that a stalemate will favour the viewpoint of the controversial false assertion, and so the perils of engaging are even greater. Either way, my argument is supported. My thanks to two anonymous reviewers for urging discussion on this.

[19]See also the literature on the psychological function of disagreement in terms of signalling group membership or protecting ego/identity (Tappin et al. 2021).

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

problem is that when the truth (in this case, climate change's reality) goes head-to-head with a fringe falsity (climate change denial), it does not suffice to come out on equal pegging. Just as if Partick Thistle, a small Scottish football team not even in the top flight of the Scottish leagues, were to hold Real Madrid, historically one of the greatest teams ever and still currently one of the best in the world, to a 1-1 tie, fans of Partick Thistle would be well entitled to view this as essentially a win, so too can a deadlocked discussion come out on the side of the climate change denier.

To continue this analogy somewhat, suppose that each of Claire and Derek's 'goals' was an argument that they both responded to very effectively – so much so that any bystander whose belief was justified by this argument had such justification defeated. The problem here oddly comes from the fact that most people *do* believe in climate change; thus, if we equally apportion defeat from the two 'goals' to the bystanders, there are simply more people to lose justification (and thus knowledge) of climate change's *reality* than its denial. For instance, say there are twenty bystanders. Using our percentages from the previous section, that's fourteen believers, three suspended, and three non-believers. If we say half of each side have their justification defeated, that's seven people on the side of climate change who suffer loss of knowledge compared with only one or two who jettison their false belief. Thus, this dialectical tie looks to clearly favour Derek's side. And, once again, this is all presumably quite rational for everyone – there is no principled reason to think that belief in climate change is somehow more secure or less vulnerable to defeat than the opposing position; I am looking at this in the fairest way possible. Perhaps we could say that *neither* side will shift their beliefs at all, but if we go that far, then we have to wonder what the point is of ever engaging with anyone anyway.

So, this stalemate scenario is likewise a serious concern when we engage with controversial false assertions, carrying the potential for widespread negative epistemic effects which again hugely outstrip the possible benefits. Moreover, we don't even seem to have the possible comfort (as we did in the previous scenario) that this sort of situation is rather rare; in fact, a back-and-forth discussion that peters out at an impasse strikes me as a rather common outcome. For instance, a recent study found that 'the vast majority of [arguments on Facebook] (roughly 71%) ended in a neutral manner, without reaching any agreement' (Cionea *et al.* 2017, 444). Combined with general worries about arguments not changing minds (Gordon-Smith 2019) and polarization (Sunstein 2002), it is reasonable to assume an argument without resolution like *Dinner** is going to be fairly frequent. Let's now turn to the final variation of our *Dinner* cases, the 'success story', and see what can be drawn from it.

2.3. Winning?

> *Dinner***. Erica is out for dinner with some colleagues. During the mains, one colleague, Frank, starts espousing the view that climate change is a hoax. He presents some 'evidence' and generally puts forward the case of climate change denial well. Erica knows this view is false and so engages with Frank, objecting to his false assertion. Fortunately, Erica has come rather well prepared and responds very effectively to Frank. She attacks his arguments cleverly, rejects his evidence, provides evidence that he does not know how to refute, and generally seems to prove him wrong.

At last, we have arrived at the good scenario. Erica has 'won', the case for climate change has been forcefully shown, and so we might reasonably assume that positive epistemic effects will abound. At the very least, it seems that there are no obvious negative

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

epistemic effects here like there were in previous cases: none of the bystanders lose justification or knowledge, and even some or all of the non-believers jettison their false belief and take up the true belief of climate change's reality (although let's assume like Alex in the first case, her foil here in Frank does not change his mind either).[20] Despite all this, I will now put forward some suggestive comments about a lingering possibility of *some* problematic epistemic effects.

Climate change denial is a fringe position, so the mere fact that it was seriously engaged with looks to legitimize the discussion in that there was *something* to talk about. If I were to see two people with an interest in physics having a heated discussion about whether light can propagate through a vacuum with or without the presence of a luminiferous aether, I could well assume that this is an ongoing debate for which there is no settled answer.[21] Even if one side looks to come out on top and prove the other wrong, it doesn't necessitate that they are right, nor necessitate that this is a settled issue (as we saw in the original *Dinner* case). Were people in their day-to-day lives to hear *genuine* debate on climate change denial, see discourse on Covid vaccines being deadly, and witness dialogue on the Holocaust's truth – *even if* these sides are being proven wrong – the mere fact that there exists discussion goes a long way in normalising and legitimising the position.[22]

In a sense, the climate change denier has nothing to lose: these sorts of controversial false assertions thrive on engagement, and this is a known phenomenon.[23] We can draw some parallels here with actual strategies employed by tobacco (Oreskes and Conway 2011) and oil (McMullen 2022) companies in the past. Promoting a false case isn't about getting people to believe the *opposite of fact* (smoking *doesn't* cause cancer, or climate change *isn't* happening, respectively); rather, it is about sowing seeds of doubt, causing people to wonder and entertain possibilities they previously wouldn't have.[24] Of course, I already stipulated that we won't consider any losses of knowledge or the like in this case, but even just getting people talking about an issue that is a genuine matter of fact as though it might not be settled is, I think, still concerning – and indeed a strategy that was exactly employed by the aforementioned companies. I'm not claiming that when Frank (or Bob or Derek) espoused his false belief, it was part of some grand strategy in the way tobacco companies tried to fudge the link between smoking and health problems, but I propose that the general outcome is similar in all these cases in that it still somewhat increases the likelihood of begetting false beliefs in the epistemic environment. Plausibly,

[20]My opponent might be concerned about the idealisations/simplifications in play here in the vignettes and lack of discussion of background beliefs of the audience à la Beaver and Stanley (2023) who look at the difficulty of getting people to accept propositions that clash with their 'core ideology' (156). I want to be clear that, following the extant literature (i.e., those cited in the opening paragraph), my focus is on similar ordinary circumstances and general members of the laity who can and will change their minds in response to arguments, engagements, and objections, not, for example, those for whom climate change denial (or its reality) is a part of their 'core ideology' and thus would be unmoved irrespective of how the engagement goes.

[21]I would be wrong about this; the luminiferous aether hypothesis was conclusively discarded in 1887 due to the Michelson-Morley experiment. See Shankland (1964) for the full account of events.

[22]Theel et al. (2013) discusses how CBS over-represented climate change deniers by six times their actual representation in the scientific community. It is plausible that there is a causal connection between this and the fact that in the same year, only 42% of Americans thought there was a scientific consensus on climate change (Leiserowitz et al. 2013). For a recent philosophical treatment of this issue of 'balance' in news (and scientific) reporting, see Gerken (2020).

[23]The apocryphal P. T. Barnum quote, 'there's no such thing as bad publicity', may spring to mind. Also cf. 'rage farming' (Jong-Fast 2022; Schwemmer 2021).

[24]See Gerken (2022), especially 149-151. He calls this the *salient alternative effect*: 'Roughly, this is people's disinclination to accept ascriptions of knowledge in the face of contextually salient error-possibilities' (149-150). See also David Lewis' (1996) 'Elusive Knowledge' for further parallels.

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

the mere fact that something *is* being discussed can make the discussion (appear) legitimate and normalise it further.

Of course, even if I'm right about this, one might reply that the positive epistemic effects garnered by Erica's victory will still outweigh the minor negative effects proposed above. This could well be the case, but I don't think it's immediately obvious. If we assume that the other colleagues had no beliefs about climate change, then it does look like there will be a huge gain in the inculcation of true belief or even knowledge. Nonetheless, we already stipulated that the bystanders mainly *truly believe* or *know* that climate change is real. If this is the case, then an obvious question is: what positive epistemic effects *are* realised? Perhaps some achieve certainty? This seems unlikely: witnessing a creationist being dialectically trounced by a physicist doesn't strike me as something that would make certain the reality of the Big Bang to those who antecedently believe in it. Perhaps there are some small increases in confidence – either in the belief itself or the evidence for it. I think this is more plausible. And, of course, we must also allow that some (or all) of the deniers jettison their false belief and maybe even take up a true belief about climate change. Granting all this, does it outweigh the potential legitimising negative effects I outlined above? I have my doubts, but I don't have a definite answer to this. Either way, let's grant that there are indeed epistemic benefits from Erica's engagement triumph that outweigh any possible negative ones. Even allowing this, I will now show that it comes of little comfort to anyone potentially engaging with controversial false assertions.

2.4. Stocktake

So where are we? I began by noting a lacuna in the engaging with false assertions literature related to potential negative epistemic effects *outwith* the engager and the engaged. I connected this to another underappreciated aspect of the literature: namely, the myriad of ways an engagement can go. From there, I then looked at three different outcomes from an engagement – a loss from the champion of truth, an impasse, and a win. I then argued that each of these three options are not made equal insofar as a loss is far more negative epistemically speaking than a win is positive. Moreover, a stalemate is not a neutral outcome but in fact can in some sense be considered a win for the false assertor. Clearly, none of this is particularly good news – if any readers are harbouring Millian sympathies about the marketplace of ideas, the arguments above suggest some scepticism to say the least.

With all this in mind, let's throw one more plausible idea into the mix: there is no real way for a prospective engager to know how an engagement will go beforehand. Therefore, pre-engagement, the epistemic risks from doing so look to far outweigh the potential rewards. Consider the following expected value calculation (Table 2):

Table 2. Expected epistemic value of engaging.

Outcome	Probability	Epistemic (Dis)value
Loss	0.2	−10
Stalemate	0.7	−5
Win	0.1	3

The expected epistemic value here works out at -5.2, meaning that the balance is well in the negative on average when we opt to engage with the sorts of controversial assertions I have been discussing throughout. Of course, the numbers here are somewhat

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

arbitrary, but they do track with the arguments laid out in the previous subsections. We already covered in section 2.1. how disastrous a loss would likely be in terms of potential knowledge lost/justification defeated from the bystanders who previously believed in the reality of climate change and the false belief inculcated in those who were suspended on the question, so there are good grounds for a weighty disvalue sum far more negative than the win is positive. We also discussed how a stalemate situation, while clearly not as pernicious as the loss scenario, still looked to overwhelmingly favour the denier due to, for instance, the way justificatory defeat would be apportioned among bystanders; thus, it is also more epistemically concerning than the win is decisive. In fact, giving the win a plus-value is potentially somewhat charitable given the legitimation worries, so the expected value is plausibly even worse than outlined here. The probabilities stem first from the communications study cited earlier that suggested that 71% of such discussions end at an impasse (Cionea *et al.* 2017). I then put losing as nominally more likely due to the data that show climate change scepticism *increased* with greater scientific literacy among Republicans (Funk and Kennedy 2020), but even with the two being equally likely *or* winning having a higher probability (which I think would be mistaken), the expected epistemic value is still majorly weighted in the negative.

In fact, supposing we put together a hugely *optimistic* calculation, the expected value remains negative (Table 3):

Table 3. Optimistic expected (epistemic) value of engaging.

Outcome	Probability	Epistemic (Dis)value
Loss	0.33	−3
Stalemate	0.33	−2
Win	0.33	3

Here we get -0.63 which is evidently close to neutral but recall that the set-up in the cases above was supposed to be the sort of 'ideal' scenarios according to the others in the literature (closed-minded engagers, sans any epistemic injustices). Moreover, from what was argued above, the numbers here look rather implausible; there's nothing to suggest that each outcome is equally likely, and I argued that the negative effects in the losing case are far greater than the positives in the winning case. So, putting together an unrealistically optimistic scenario still works out as overall negative, even when we eliminate the concerns that have taken precedence in the extant literature!

All this to say that there are clearly more dimensions to engaging with certain sorts of assertions than have been previously recognised by those working in this area, and we have good reason to be more reticent about objecting and engaging (and any obligations therein) than was before appreciated.

2.5. Online

Recall earlier when I noted that it's a virtue of the *Dinner* cases that they are relevantly analogous to possible online settings insofar as there are two interlocutors and an audience. The reason why I view this as a boon is because an arena where controversial false assertions are certainly prevalent is, of course, online – specifically on social media sites (Vosoughi *et al.* 2018; Suarez-Lledo and Alvarez-Galvez 2021; Y. Wang *et al.* 2019). Readers themselves may even consider how many times they have observed or even participated in such political or medical arguments on, for example, Facebook or

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

Twitter. Importantly, then the arguments above about the epistemic dangers for bystanders *are* relevant to such online scenarios. Moreover, not only do they apply *mutatis mutandis*, but I will now argue that the concerns are likely exacerbated online compared to in person. Specifically, and staying in lockstep with what has been discussed so far in this paper, the focus in this subsection is the (even greater) difficulty one is faced with in securing the *winning scenario* (*Dinner***) in an online setting. Nevertheless, it is also important to note that the entire online situation is often worse due to the platforms themselves being designed and indeed optimised in a way that amplifies hateful, incendiary, and violent content (Munn 2020).[25]

Let's consider one final case, plausibly more common than any of the *Dinner* cases outlined above:

> *Social Media*. Greta is scrolling on social media when she comes across a long thread purporting to show that the Holocaust is a hoax. Greta knows that this is false and so replies to the post, objecting to its claims. The author never responds to her, however. In fact, the only tangible effects Greta's response really had was that she amplified engagement on the post (she increased the comment counter and caused some of her own social media followers to click on the thread).

Above, we granted that if an engagement is demonstrably successful inasmuch as the objector seems to strongly prove the false assertor wrong, then it looks like it will have a positive epistemic impact. Moreover, it seemed to be the only scenario that could ensure overall positive epistemic effects. The problem here in *Social Media* is that whether this is true or not, the normative pressure for the false assertor to respond to an engagement or objection does not exist in the same way that it does in person. Therefore, even if we stipulated that Greta *would* be very successful if the Holocaust denier were to respond, the assertor can simply ignore her and thus entirely prevent the possibility of this kind of *dialogical*, *engagement* victory for her. Crucially, this does not reflect as poorly on the denier the same way it would if the discussion were in person.

To explain further, suppose *Social Media* were face-to-face in the manner of the *Dinner* cases. Were Greta to engage with the denier and they were to simply shut down and make no response, this would look much like an engagement victory for Greta – that is, her immediate objection or response was so effective that her interlocutor had no rejoinder. Therefore, this is plausibly a situation in which our overall positive epistemic effects are realised. However, the same reasoning arguably does not apply in the online version. When the Holocaust denier does not respond to an objection, an equivalent inference of their inability to respond does not seem to be as legitimately drawn because of the distinct dearth of normative pressure to engage with objectors when compared with in-person; therefore, they plausibly don't even have to take the *risk* of being dialectically trumped by anyone. Admittedly, a lack of response could be interpreted as the denier having no good answer for Greta, but at the very least it will not *always* be interpreted so – and perhaps not even *mostly*. Thus, when combined with the fact that an engagement *victory* is already the most unlikely scenario granted that one even gets the *opportunity* for a dialogue, there are legitimate worries that the good scenario of *Dinner*** will be extremely rare online. And so, a regular outcome from this sort of online engagement is just exposing the thread to more people,[26] and it is difficult to see how this could possibly be overall more epistemically positive than negative.

[25]My thanks to an anonymous referee for suggesting I note this.

[26]Cf. Saul (2021, 147–48). She calls this 'amplification'.

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

Neither does it seem like we can just ignore all such controversial false assertions online.[27] Were everyone who knows their falsity to simply scroll past when they come across them, it would likely prove problematic; the most abhorrent of claims would only be accompanied with widespread agreement. It is facile to predict negative epistemic consequences from this. Hence, controversial false assertions are plausibly even more problematic online than in any other arena, and their commonality means there isn't even the comfort one might have had earlier that these situations don't arise particularly often.

3. Upshots

The foregoing conclusions have all been rather pessimistic. I have suggested that there are far more epistemic dangers with engaging than have been appreciated thus far in the literature, and so have called for caution, even when the worries that other scholars have pointed out are accounted for. In this final section, then, I will use these conclusions to yield some upshots – both theoretical and practical. The theoretical upshots will mainly focus on the problems my arguments might generate for certain philosophers who have considered obligations to engage or object in consequentialist terms; the practical upshots will sketch some ways of potentially engaging that avoid the problems outlined above and also suggest that some epistemically paternalistic practices may be the way forward.

3.1. Theoretical

Throughout, I have been particularly concerned with the consequences (and potential or expected consequences) from engaging with or objecting to controversial false assertions. I have shown that, on balance, the outcomes are epistemically negative. Therefore, anyone who grounds any obligations or duties to engage or object in *consequences* looks to be in trouble. The natural question then is: does anyone do this? McKenna, for one, specifically appeals to an (epistemic) consequentialist framework when discussing obligations to engage.[28] McIntyre (2018, 2020, 2021) also appeals to consequentialism when putting forward arguments as to why we should engage with controversial false assertions. Unsurprisingly, so does one of the original progenitors of consequentialism, Mill. The rest don't explicitly state how they are thinking about this exact area, but their discussions are all replete with allusions and references to consequences, so these arguments are certainly ones they have to take seriously. Lackey is presumably the only one immune to these arguments due to the fact that her duty to object is a deontic one, so consequences are immaterial. Nevertheless, in later work when discussing the duty to object, Lackey appeals to a Singer-esque umbrella principle,[29] which certainly appears to be consequentialist in nature, so it's possible that even she must take the arguments in this paper into consideration.

In any case, all the authors I highlighted in the introduction are, to varying degrees of strength, interested in consequences and clearly believed that aside from a few potential

[27]McIntyre (2020) also says much the same: 'providing no response to misinformation was the worst thing one could do; with no rebuttal message, subjects were most likely to be swayed toward false beliefs'. (220) McIntyre draws on Schmid and Betsch (2019) for this conclusion.

[28]See McKenna (2023), especially p.105.

[29]She calls this epistemic umbrella principle *Interpersonal Epistemic Duties*: 'If it is in our power to prevent something epistemically bad from happening through very little effort on our part, we ought, epistemically, to do it'. (Lackey 2020a, 287). While I don't think this is necessarily consequentialist, it can definitely be read as so (and its Singer origins give some grounds for viewing it this way as well).

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

problems an objector can have when engaging, there was not much else to be overly concerned about. My arguments have shown that this is not true, and there are certainly further potential areas of epistemic danger outwith the dichotomy of the engager and their interlocutor.

3.2. Practical

Finally, let's turn to the practical upshots from the arguments and conclusions throughout. First, for the online side of the engagement problems I've outlined, it looks like while there may be concerns with *individual* responses to controversial false assertions, there seem to be obvious, *institutional* responses that are in fact already implemented in the real world: epistemically paternalistic policies of deplatforming and censorship.[30] On social media (in the past, at least – things are less clear since new management took over Twitter/X), if one spreads controversial false assertions like climate change denial, their posts will be removed, and if they continue to do so repeatedly, they will be deplatformed (indefinitely banned from the site). So, the thought goes, this sort of solution eliminates problems I have outlined above with engaging by simply cutting the problem off at the root; if people can't even *encounter* such controversial false assertions, no one has to engage with them as there is nothing to engage with, and so people's epistemologies cannot be adversely affected by them. Moreover, empirical data strongly support this claim that such practices are effective in reducing the reach of bad actors and their assertions (Rauchfleisch and Kaiser 2021; Innes and Innes 2021; Jhaver *et al.* 2021).

Nevertheless, this is not a perfect solution. For a start, there are basic worries about granting large corporations and governments such epistemically paternalistic powers as the scope for abuse is evidently concerning (Goldman 1991, 127; Aird 2023). Second, even if implemented, such policies are not infallible in that some controversial false assertions will slip through the cracks (for a time, at least). In such cases, the problems outlined in this paper again arise in that it is a live question of what one should do in cases where one comes across such an assertion that is yet to be removed. Third, there are historical examples of (attempted) censorship of books or films simply making them *more* popular (e.g., *Lady Chatterley's Lover* (Baksi 2019) or *A Clockwork Orange* (Brew 2019)). Fourth, there is some empirical work that suggests that deplatforming may increase polarisation and further entrench people in the views of the deplatformed individual as they follow them to a different site with more lax moderation rules and fall deeper into an echo chamber (Ali *et al.* 2021). Finally, even if we were to grant that such epistemically paternalistic policies *are* effective in our online scenarios, this gives us no solution to our *Dinner* cases insofar as it's not immediately obvious if we could 'deplatform' Bob, Derek, or Frank from the dinner table in any meaningful way.

This final point is an interesting one and worth exploring. After all, despite the potential problems with deplatforming and censorship discussed above, it does look like a far more effective strategy in general for combatting controversial false assertions than simply engaging and disagreeing.[31] Are there in-person analogues for deplatforming that we could apply to *Dinner* and similar cases? Prima facie, it doesn't look like it, but, in

[30]For the key works on epistemic paternalism, see Goldman (1991) and Ahlstrom-Vij (2013). For epistemic paternalism specifically connected to deplatforming and censorship (and the potential problems therein), see Aird (2023).

[31]Particularly if one thinks that they will lose the engagement à la *Dinner* or *Dinner**. Interestingly, Paglieri (2013, sec. 3) suggests that *likelihood of winning* is the main predictor for prospective engagers commit to arguing or not, so it is plausible that speakers are capable of identifying this with some reliability.

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

formal settings, they *do* exist. For instance, think about the practice of 'striking from the record' in courtrooms. The Federal Rules of Civil Procedure state: 'The court may strike from a pleading an insufficient defence or any redundant, immaterial, impertinent, or scandalous matter' (United States Courts). This is essentially the deplatforming and/or censorship of certain ideas, and, say, climate change denial presumably comes under 'redundant', 'impertinent', or 'scandalous'. Again, however, this is a practice in a formal setting, so how to apply it to our *Dinner* cases isn't especially clear (Alex putting forward a motion to strike after Bob's espousals doesn't look like a very effective strategy). Nonetheless, there is something in the idea of striking from the record that is useful: the explicit recommendation that a statement or assertion is so outrageous that it does not even deserve any consideration or response; in fact, we should expunge it and move on as though it did not exist. Capturing this phenomenon in our informal cases is no mean feat and likely beyond the scope of this paper, but I think a potential method is that of *interrupting* or *closing the conversation*. For instance, Bob begins stating his climate change denial and Alex immediately interrupts and moves the conversation on to something else. This is clearly different from the serious engagement of objecting and arguably avoids our engagement problems as Bob's assertion is cut short before it is added to the conversational record and properly discussed. Of course, there are again caveats here in that Bob might just not allow himself to be interrupted or the conversation to be closed, and then we seem to be back to where we started with our engagement problems, but as a first gloss these conversation closers have encouraging attributes.

There is a further strategy that is quite obvious. I have throughout been focused on the effects on those outwith the engager and the engaged – the third-party observers. Thus, if we can remove this audience, then our only worries are the ones that have been discussed in the literature already (avoiding open-mindedness and preventing any testimonial injustices). Therefore, if one of these sorts of discussions about controversial false assertions is unavoidable, a good strategy to avoid the problems I have discussed is to try to have the conversation in a one-on-one setting.[32] In such a case, there is no worry about deleterious effects on third-party observers – because there simply aren't any. Of course, somehow ensuring the conversation is sans bystanders is a difficult task, but it's an important factor to keep in mind.

A final alternative that strays away from the sort of dialogical, pure engagement of objecting and arguing with someone (essentially, telling the person that they are *wrong*) that characterised the *Dinner* cases above is instead to employ narratives (Whitmarsh and Corner 2017) and avoid using certain terms when discussing the controversial false assertion at hand. For example, Arbuckle *et al.* (2014) found that 'emphasising terminology and narratives that focus on adaptation to weather variability rather than climate change may be better received and more effective when working with farmers' (515). So, the idea is that the *reality* of the matter is still under discussion (e.g., genuine extreme weather events caused by climate change), but it is couched in terms that avoid the charged, politicised language of, say, 'climate change' or 'global warming', and is not phrased in a way that is necessarily *objecting to* or even *disagreeing with* the engaged. Not only does the aforementioned empirical data suggest that this is an effective engagement strategy simpliciter, but also the relevance here is that this kind of engaging plausibly avoids the worries outlined in this paper. This is because the engager can avoid getting into an argument like the protagonists in the *Dinner* cases, and thus this more cautious

Thus, we have reason to think that finding strategies to employ when one believes they will be unsuccessful in the dialogue would be useful.

[32]McCormick (2023) also notes this.

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

engaging should not carry with it the same possibility of dialectical failure or stalemate that plagued those examples.[33]

The practical suggestions above are mere sketches, and the exact effectiveness of them is not something that I can guarantee. Nevertheless, as a starting point to potentially avoid the engagement issues I have discussed throughout, I think these show some promise and ought to be explored further.

4. Conclusion

I began with a broad overview of the current literature on engaging with challenges to our beliefs, noting a lacuna relating to the possible epistemic effects outwith merely the engager and the engaged. I then went on to discuss three versions of the same case, each with a different outcome, where the highly pernicious nature of certain engagements was shown. Overall, I proposed that the possible effects on third-party observers when it comes to engaging with controversial false assertions are so wide-ranging that they swamp any of the benefits previously discussed in the literature. Alone, these are novel arguments and conclusions, but I drew on them to suggest that this should give pause and make us reconsider many of the obligations and duties to object or engage put forward thus far. Finally, I closed with some suggestions on how to possibly avoid these engagement worries that plagued vanilla engagement and objecting, looking at some epistemically paternalistic policies among other, more individualistic, responses.[34]

References

Aalberg T. and Todal Jenssen A. (2007). 'Do Television Debates in Multiparty Systems Affect Viewers? A Quasi- Experimental Study with First-Time Voters.' *Scandinavian Political Studies* **30**(1), 115–35. https://doi.org/10.1111/j.1467-9477.2007.00175.x.

Ahlstrom-Vij K. (2013). *Epistemic Paternalism*. London: Palgrave Macmillan UK. https://doi.org/10.1057/9781137313171.

Aikin S. (2010). *Epistemology and the Regress Problem*. Vol. **25**. New York: Routledge. https://doi.org/10.4324/9780203833247

Aird R. (2023). 'A Puzzle of Epistemic Paternalism.' *Philosophical Psychology* **36**(5), 1011–29. https://doi.org/10.1080/09515089.2022.2146490.

Ali S., Saeed H.M., Aldreabi E. et al. (2021). 'Understanding the Effect of Deplatforming on Social Networks.' 13th ACM Web Science Conference 2021 (New York, NY, USA), WebSci '21, June 22, 187–95. https://doi.org/10.1145/3447535.3462637.

[33]On the other hand, there is considerable scepticism in science communication about this idea of simply presenting facts to the public being a successful engagement strategy for understanding science – for more on the deficit model and its flaws, see, among many others, Miller (2001), Nisbet and Scheufele (2009), and Simis et al. (2016). Nevertheless, these issues are more focused on science communicator/layperson interactions which are at least somewhat different from the more ordinary peer-to-peer discussions that I have been thinking about here. For a philosophical treatment and scepticism that these issues are even ones of information deficit, also see Levy (2021, 24–27). My thanks to an anonymous reviewer for urging me to discuss this.

[34]Special thanks to Ross Patrizio for critical support and advice. I am also indebted to Adam Carter, Emma Gordon, Tim Kearl, and an anonymous reviewer for extremely helpful comments on numerous different versions of this paper. Thanks also to audiences at the Scottish Epistemology Early Career Researchers (SEECRs) WiP, Understanding Value XI at the University of Sheffield, the University of Glasgow PG Seminar, and the Society for Applied Philosophy Annual Conference 2024 at the University of Oxford. Final thanks to Peter Djikhuizen for luminiferous aether. This work was generously supported by the Scottish Graduate School for Arts and Humanities and the Royal Institute of Philosophy.

Arbuckle J.G., Hobbs J., Loy A., Morton L.W., Prokopy L.S. and Tyndall J. (2014). 'Understanding Corn Belt Farmer Perspectives on Climate Change to Inform Engagement Strategies for Adaptation and Mitigation.' *Journal of Soil and Water Conservation* **69**(6), 505–16. https://doi.org/10.2489/jswc.69.6.505.

Baksi C. (2019). 'Lady Chatterley's Legal Case: How the Book Changed the Meaning of Obscene.' Law. The Guardian, August 1. https://www.theguardian.com/law/2019/aug/01/lady-chatterleys-legal-case-how-the-book-changed-the-meaning-of-obscene.

Ballew M.T., Leiserowitz A., Roser-Renouf C. et al. (2019). 'Climate Change in the American Mind: Data, Tools, and Trends.' *Environment: Science and Policy for Sustainable Development* **61**(3), 4–18.

Battaly H. (2018). 'Can Closed-Mindedness Be an Intellectual Virtue?.' *Royal Institute of Philosophy Supplement* **84**(November), 23–45. https://doi.org/10.1017/S135824611800053X.

Battaly H. (2021). 'Engaging Closed-Mindedly with Your Polluted Media Feed.' In M. Hannon and J. de Ridder (eds), *The Routledge Handbook of Political Epistemology*, 312–324, 1st ed. Oxon: Routledge. https://doi.org/10.4324/9780429326769-38.

Beaver D. and Stanley J. (2023). The Politics of Language. Princeton: Princeton University Press. https://doi.org/10.1515/9780691242743.

Brew S. (2019). 'The Myth of A Clockwork Orange's Ban.' Den of Geek, December 19. https://www.denofgeek.com/movies/the-myth-of-a-clockwork-orange-s-ban/.

Cassam Q. (2019). *Vices of the Mind: From the Intellectual to the Political*. 1st ed. Oxford: Oxford University Press.

Cionea I.A., Piercy C.W. and Carpenter C.J. (2017). 'A Profile of Arguing Behaviors on Facebook.' *Computers in Human Behavior* **76**(November), 438–49. https://doi.org/10.1016/j.chb.2017.08.009.

Cohen D.H. (2005). 'Arguments That Backfire.' In D. Hitchcock and D. Farr (eds), *The Uses of Argument*, 58–65, Ontario Society for the Study of Argumentation.

Novaes D., Catarina. (2023). 'VII—Can Arguments Change Minds?' Proceedings of the Aristotelian Society **123**(2), 173–98. https://doi.org/10.1093/arisoc/aoad006.

Ettensperger F., Waldvogel T., Wagschal U. and Weishaupt S. (2023). 'How to Convince in a Televised Debate: The Application of Machine Learning to Analyze Why Viewers Changed Their Winner Perception during the 2021 German Chancellor Discussion.' *Humanities and Social Sciences Communications* **10**(1), 546. https://doi.org/10.1057/s41599-023-02047-5.

Fantl J. (2018). *The Limitations of the Open Mind*. Oxford: Oxford University Press.

Funk C. and Kennedy B. (2020). 'For Earth Day 2020, How Americans See Climate Change and the Environment in 7 Charts.' *Pew Research Center*. https://pewrsr.ch/3saE1wX.

Gerken M. (2020). 'How to Balance Balanced Reporting and Reliable Reporting.' *Philosophical Studies* **177**(10), 3117–42. https://doi.org/10.1007/s11098-019-01362-5.

Gerken M. (2022). *Scientific Testimony: Its Roles in Science and Society*. Oxford: Oxford University Press.

Goldman A.I. (1991). 'Epistemic Paternalism: Communication Control in Law and Society.' *The Journal of Philosophy* **88**(3), 113. https://doi.org/10.2307/2026984.

Gordon-Smith E. (2019). Stop Being Reasonable: How We Really Change Our Minds. Sydney: PublicAffairs.

Innes H. and Innes M. (2021). 'De-Platforming Disinformation: Conspiracy Theories and Their Control.' *Information, Communication & Society* **26**(6), 1262–1280. https://doi.org/10.1080/1369118X.2021.1994631.

Jhaver S., Boylston C., Yang D. and Bruckman A. (2021). 'Evaluating the Effectiveness of Deplatforming as a Moderation Strategy on Twitter.' Proceedings of the ACM on Human-Computer Interaction **5**(CSCW2): 381:1–381:30. https://doi.org/10.1145/3479525.

Jong-Fast M. (2022). 'Owning the Libs Is the Only GOP Platform.' The Atlantic, January 12. https://newsletters.theatlantic.com/wait-what/61df26769a2cf50020163abc/owning-the-libs-is-the-only-gop-platform/.

Lackey J. (2020a). 'Epistemic Duties Regarding Others.' In K. McCain and S Stapleford (eds), *Epistemic Duties*, 281–295. New York: Routledge.

Lackey J. (2020b). 'The Duty to Object.' *Philosophy and Phenomenological Research* **101**(1), 35–60. https://doi.org/10.1111/phpr.12563.

Leiserowitz A., Maibach E.W., Roser-Renouf C., Feinberg G. and Howe P. (2013). 'Climate Change in the American Mind: Americans' Global Warming Beliefs and Attitudes in April 2013.' Available at SSRN 2298705.

Levy N. (2017). 'The Bad News about Fake News.' *Social Epistemology Review and Reply Collective* **6**(8), 20–36.

Levy N. (2019). 'No-Platforming and Higher-Order Evidence, or Anti-Anti-No-Platforming.' *Journal of the American Philosophical Association* **5**(4), 487–502. https://doi.org/10.1017/apa.2019.29.

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

Levy N. (2021). *Bad Beliefs: Why They Happen to Good People*. 1st ed. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780192895325.001.0001.

Lewis D. (1996). 'Elusive Knowledge.' *Australasian Journal of Philosophy* **74**(4), 549–67. https://doi.org/10.1080/00048409612347521.

Macleod C. (2021). 'Mill on the Liberty of Thought and Discussion.' In A. Stone and F. Schauer (eds), *The Oxford Handbook of Freedom of Speech*, 2–19. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198827580.013.1.

McCormick M.S. (2023). 'Engaging with "Fringe" Beliefs: Why, When, and How.' *Episteme* June **13**, 1–16. https://doi.org/10.1017/epi.2023.33.

McIntyre L. (2018). *Post-Truth*. Cambridge, Massachusetts: MIT Press.

McIntyre L. (2020). 'Science Denial, Polarisation, and Arrogance.' In A. Tanesini and M. P. Lynch (eds), *Polarisation, Arrogance, and Dogmatism*, 193–211. Oxon: Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780429291395-16/science-denial-polarisation-arrogance-lee-mcintyre.

McIntyre L. (2021). *How to Talk to a Science Denier: Conversations with Flat Earthers, Climate Deniers, and Others Who Defy Reason*. Cambridge, Massachusetts: MIT Press.

McKenna R. (2023). *Non-Ideal Epistemology*. Oxford: Oxford University Press.

McMullen J. (2022). 'The Audacious PR Plot That Seeded Doubt about Climate Change.' Science & Environment. BBC News, July 22. https://www.bbc.com/news/science-environment-62225696.

Mill J.S., Bromwich D., Kateb G. and Bethke Elshtain J. (1859). *On Liberty. Rethinking the Western Tradition*. New Haven and London: Yale University Press.

Miller S. (2001). 'Public Understanding of Science at the Crossroads.' *Public Understanding of Science* **10**(1), 115–20. https://doi.org/10.3109/a036859.

Munn L. (2020). 'Angry by Design: Toxic Communication and Technical Architectures.' *Humanities and Social Sciences Communications* **7**(1), 1–11. https://doi.org/10.1057/s41599-020-00550-7.

Nisbet M.C. and Scheufele A.D. (2009). 'What's next for Science Communication? Promising Directions and Lingering Distractions.' *American Journal of Botany* **96**(10), 1767–78. https://doi.org/10.3732/ajb.0900041.

Nwokora Z. and Brown L.M. (2017). 'Narratives of a Race: How the Media Judged a Presidential Debate.' *American Politics Research* **45**(1), 33–62. https://doi.org/10.1177/1532673X15614891.

Nyhan B. and Reifler J. (2010). 'When Corrections Fail: The Persistence of Political Misperceptions.' *Political Behavior* **32**(2), 303–30. https://doi.org/10.1007/s11109-010-9112-2.

Nyhan B. and Reifler J. (2015). 'Does Correcting Myths about the Flu Vaccine Work? An Experimental Evaluation of the Effects of Corrective Information.' *Vaccine* **33**(3), 459–64.

Oreskes N. and Conway E.M. (2011). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. New York: Bloomsbury Publishing USA.

Paglieri F. (2013). 'Choosing to Argue: Towards a Theory of Argumentative Decisions.' *Journal of Pragmatics*, Biases and constraints in communication: Argumentation, persuasion and manipulation vol. **59**(December), 153–63. https://doi.org/10.1016/j.pragma.2013.07.010.

Rauchfleisch A. and Kaiser J. (2021). 'Deplatforming the Far-Right: An Analysis of YouTube and BitChute.' SSRN Scholarly Paper No. 3867818. Rochester, NY, June 15. https://doi.org/10.2139/ssrn.3867818.

Rosenblum N.L. (2020). 'Introduction: Paths to Witnessing, Ethics of Speaking Out.' *Daedalus* **149**(4), 6–24. https://doi.org/10.1162/daed_e_01813.

Saul J. (2021). 'Someone Is Wrong on the Internet: Is There an Obligation to Correct False and Oppressive Speech on Social Media?' In Alison MacKenzie, Jennifer Rose, and Ibrar Bhatt (eds), The Epistemology of Deceit in a Postdigital Era: Dupery by Design, Postdigital Science and Education, 139–157. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-72154-1_8.

Schmid P. and Betsch C. (2019). 'Effective Strategies for Rebutting Science Denialism in Public Discussions.' *Nature Human Behaviour* **3**(9), 931–39.

Schwemmer C. (2021). 'The Limited Influence of Right-Wing Movements on Social Media User Engagement.' *Social Media + Society* **7**(3), 20563051211041650. https://doi.org/10.1177/20563051211041650.

Shankland R.S (1964). 'Michelson-Morley Experiment.' *American Journal of Physics* **32**(1), 16–35. https://doi.org/10.1119/1.1970063.

Simis M.J., Madden H., Cacciatore M.A. and Yeo S.K. (2016). 'The Lure of Rationality: Why Does the Deficit Model Persist in Science Communication?.' *Public Understanding of Science* **25**(4), 400–414. https://doi.org/10.1177/0963662516629749.

Suarez-Lledo V. and Alvarez-Galvez J. (2021). 'Prevalence of Health Misinformation on Social Media: Systematic Review.' *Journal of Medical Internet Research* **23**(1), e17187. https://doi.org/10.2196/17187.

https://doi.org/10.1017/epi.2025.10082 Published online by Cambridge University Press

Sunstein C.R. (2002). 'The Law of Group Polarization.' *Journal of Political Philosophy* **10**(2), 175–95. https://doi.org/10.1111/1467-9760.00148

Tappin B.M., Pennycook G. and Rand D.G. (2021). 'Rethinking the Link between Cognitive Sophistication and Politically Motivated Reasoning.' *Journal of Experimental Psychology. General* **150**(6), 1095–114. https://doi.org/10.1037/xge0000974.

Terzian G. and Inés Corbalán M. (2021). 'Our Epistemic Duties in Scenarios of Vaccine Mistrust.' *International Journal of Philosophical Studies* **29**(4), 613–40. https://doi.org/10.1080/09672559.2021.1997399.

Theel S., Greenberg M. and Robbins D. (2013). 'Media Sowed Doubt in Coverage of UN Climate Report.' Media Matters for Americans. October 10, 2013.

Tranter B. and Booth K. (2015). 'Scepticism in a Changing Climate: A Cross-National Study.' *Global Environmental Change* **33**, 154–64.

United States Courts. n.d. 'Federal Rules of Civil Procedure.' Accessed 27 April 2023. https://www.uscourts.gov/rules-policies/current-rules-practice-procedure/federal-rules-civil-procedure.

Vosoughi S., Roy D. and Aral S. (2018). 'The Spread of True and False News Online.' *Science* **359**(6380), 1146–51. https://doi.org/10.1126/science.aap9559.

Wang L., Beauchamp N., Shugars S. and Qin K. (2017). 'Winning on the Merits: The Joint Effects of Content and Style on Debate Outcomes.' *Transactions of the Association for Computational Linguistics* **5**(July), 219–32. https://doi.org/10.1162/tacl_a_00057.

Wang Y., McKee M., Torbica A. and Stuckler D. (2019). 'Systematic Literature Review on the Spread of Health-Related Misinformation on Social Media.' *Social Science & Medicine* **240**(November), 112552. https://doi.org/10.1016/j.socscimed.2019.112552.

Whitmarsh L. and Corner A. (2017). 'Tools for a New Climate Conversation: A Mixed-Methods Study of Language for Public Engagement across the Political Spectrum.' *Global Environmental Change* **42**(January), 122–35. https://doi.org/10.1016/j.gloenvcha.2016.12.008.

Williamson T. (2020). *Philosophical Method: A Very Short Introduction*. Oxford: Oxford University Press.

Rory Aird is a PhD student and member of the COGITO Epistemology Research Centre at the University of Glasgow. He works mainly on political and social epistemology. More information can be found at: www.roryairdphilosophy.com. He can also be reached at r.aird.1@research.gla.ac.uk.

Cite this article: Aird R.J. (2025). "On the Perils of Engaging." *Episteme* 1–20. https://doi.org/10.1017/epi.2025.10082