The Censor's Burden

Hrishikesh Joshi

penultimate draft; forthcoming in *Noûs*

Censorship involves, inter alia, adopting a certain type of epistemic policy. While much has been written on the harms and benefits of free expression, and the associated rights thereof, the epistemic preconditions of justified censorship are relatively underexplored. In this paper, I argue that examining *intrapersonal* norms of how we ought to treat evidence that might come to us over time can shed light on *interpersonal* norms of evidence generation and sharing that are relevant in the context of censorship. The upshot is that justified censorship requires the censor to meet a very high epistemic burden regarding the target proposition(s)—importantly, one that exceeds knowledge.

1. Introduction

What does it take to justify a policy of censorship? Here I want to argue that for censorship to be justified, the censor must be in an unusually good epistemic position, and that this is highly unlikely to obtain in most cases of actual censorship. My argument will be Millian in spirit and will draw significantly from what Mill says in *On Liberty*, Chapter 2. However, my emphasis will be somewhat different, and I will eschew some of the assumptions he makes, which are not strictly necessary for a successful argument against censorship. Further, I don't want to claim that censorship will *never* be appropriate, as Mill often seems to suggest, but rather, more weakly, that the burden the censor must meet is higher than is commonly appreciated.

The argument builds on the observation that the generation of evidence regarding most claims is a process that occurs over time. And part of what it is to be an epistemically rational agent is to update one's beliefs in light of the new evidence as it is uncovered. Given that evidence comes to us over time, a policy of censorship requires the censor not only to be justified in thinking that *P* is true, but *also*, inter alia, to meet a high bar of justification for thinking that whatever evidence might be uncovered or presented to him in the future will either be non-existent, weak, or misleading.[1] It is this latter condition that will in practice be extremely difficult to meet regarding all but the most directly accessible propositions. The condition will not be met regarding complex empirical matters, which are often the objects of censorship.[2]

The basic point here is that the epistemic bar necessary for justifiably believing *P* is lower than the bar necessary for being justified in blocking future evidence regarding *P*. To use a simple example, suppose a researcher is studying the causal relationship between variables X and Y. She conducts a study with a reasonably large sample size which shows a statistically significant relationship between X and Y. She thereby concludes that X is a causal factor in bringing about Y.

[1] Framed this way, the view presented here need not deny closure (Luper 2016). The problem would be that if S is justified in believing *P*, then S is justified in believing that evidence against *P* is misleading—because *P* entails that evidence against *P* is misleading. But if we allow for degrees of justification, the thought would be that justification for believing that evidence against *P* is misleading must meet a *higher* bar such that it would license ignoring future evidence against *P*. This is highly plausible in the intrapersonal case, otherwise there would be license for a form of "ostrich epistemology"—get justification and hide!

[2] For a recent discussion of contemporary and historical cases that fit this bill, see Clark et. al. (2023).

Let's suppose she is justified in coming to this belief. This level of justification, however, is *not* enough for her to dismiss future evidence that might come to her. Ideally, she ought to hold her belief provisionally in the sense that she is open to future inquiries, which may show, for instance, that the result does not replicate in other samples or that there are confounding factors that undermine the case for the causal relationship between X and Y.

I am interested in a specific form of censorship, of the sort that Mill was primarily concerned with. The word 'censorship' may of course be used in other ways, but for our purposes, censorship involves the intentional act of either preventing some claim from being successfully communicated or significantly disincentivizing such communication. Such censorship is sensitive to the *content* of what is being said and is insensitive to whether the speaker is sincere. Thus, for example, you do not engage in the relevant kind of censorship when you ask students who are talking with each other during your lecture to please be quiet. Here, your intention is not in the first instance to prohibit a particular content from being communicated, but rather to avoid disruption of your class. Likewise, a social media site could eliminate "bots" posting spam, without thereby engaging in the form of censorship being discussed here. Similarly, we can imagine a fake news website which simply generates made-up stories which the operators of the website do not believe are true. Here again, the intention is not to suppress a specific content, but rather a specific type of activity—namely, deception. One can prohibit the use of verbal threats and slurs as well, without thereby engaging in censorship.[3]

That said, there are certain sorts of blocking of content that do not count as censorship. For example, an editor may reject a paper submission because it does not meet the standards of quality for their journal. This can be sensitive to the content of the paper and yet is not censorious in the relevant sense. We might contrast this with a case where an editor blocks the publication of an article or story for political considerations or to placate an advertiser. This latter case plausibly involves censorship. If that's right, then censorship involves a particular sort of *intent*. J.P. Messina (2023, 7) has recently offered a helpful characterization of censorial intent; for him, censorship involves "the attempted suppression of expressive content on the grounds that it is dangerous, threatening to the (moral, political, or religious) orthodoxy, or inimical to the material interests of the agent aiming to suppress it." I think this does a reasonable job at capturing the interesting cases and works well to fix ideas for the purposes of this paper. In what follows, I will distinguish the three sorts of censorial intent and treat them separately.[4]

The most interesting sort of censorship will involve stable *policies* or *norms*. Thus, a paradigm case of government censorship will involve a policy of imposing some sanction for disputing some claim *P*. Social norms can be censorious too, as when there are sufficiently strong social sanctions for asserting or disputing certain claims. Such policies or norms will require a high burden of justification because they hamper the generation and dissemination of evidence. One-off cases of censorship are less interesting for the purposes of this paper, but they too can have the above result insofar as they create a "chilling effect." Censors can be individuals, but more often, they are groups

[3] Much of the recent literature on justifying speech restrictions centers on slurs and other expressions, the function of which is not primarily to share evidence or express a sincerely held opinion—see for example, Waldron (2012) or McGowan (2019). My focus in this paper is primarily epistemic—both on the epistemic conditions the censor must meet in order to be justified in enacting censorship and the potential epistemic consequences of censorship. Because of this epistemic focus, the central speech acts I am concerned with are sincere assertions of propositions, or presentations of arguments and data analyses, and so on. In this way, much of what I have to say here doesn't rule out many of the regulations that these authors have argued for.

[4] Thanks to an anonymous referee for pressing me to be clearer about the nature of censorial intent.

or group agents.[5] Thus, censors can include government agencies, corporations, and a range of other institutions.[6] When censorship is enacted through social norms, it's not clear that a group agent is at play. Rather, plausibly, the censors here are the set of individuals who endorse and enforce the relevant norm.

Why might someone engage in an act of censorship? First, they might do it for malicious or self-interested reasons. Thus suppose the Great Leader wants to remain in power and believes that if certain evidence were to become widely disseminated (about the relative prosperity of the outside world, say), this would significantly increase the chances of him being dethroned. Here, he may engage in censorship simply to remain in power.

But someone may censor out of better intentions. One possibility is that they censor to protect the quality of our shared epistemic resources.[7] Or put differently, they might think that allowing certain opinions to be voiced would lead to a deterioration of our collective epistemic position, in some sense. This would be censorship, out of good intentions, for purely *epistemic* reasons. Censorship of this kind will be addressed in §2.

Alternatively, the censor might be driven by *non-epistemic* considerations. That is, they might believe that sharing certain sorts of arguments, putative evidence, opinions, etc., might cause more harm than good. Such a censor might target the sharing of evidence against *P*, even if he believes *P* is not true or is agnostic about *P*.[8] This sort of case will be addressed in §3. Of course, these two latter forms of censorship are not mutually exclusive—someone might engage in censorship for both epistemic and non-epistemic reasons. However, it will be helpful to treat them separately because different sorts of reasons are at play here and they call for different sorts of justification.

This paper is primarily an exercise in *ideal* epistemology.[9] Part of the general argument against censorship is that would-be censors are non-ideal agents, and thus will be susceptible to a

[5] When a group agent (List and Pettit 2011) engages in censorship, the relevant epistemic states will *group* beliefs; for a recent account of group belief, see Lackey (2021).

[6] Legal scholars have extensively discussed cases where the censor is a government entity (Chemerinsky 2019). More recently, a lot of discussion has shifted to censorship by private, non-governmental entities such as social media corporations (Messina 2023). Much recent controversy centers around speech on campuses, and the ethics of (dis)inviting speakers. For recent philosophical defenses of limits on controversial speech within an academic setting, see Fantl (2018) and Simpson (2020).

[7] W.K. Clifford (1877, 292) emphasized that our epistemic resources are "common property" in an important sense. For him, this grounds the norm that individuals must only believe on the basis of sufficient evidence, because our beliefs don't merely concern ourselves. In a similar vein, it might be argued that censors can properly act to prohibit "epistemic pollution," as it were, of these resources. For a recent defense of the idea of an "epistemic commons" and its normative implications, see Joshi (2021).

[8] Indeed, this was the position of one of Mill's early critics, James Fitzjames Stephen, who wrote:

> [A]n opinion may be silenced without any assertion on the part of the person who silences it that it is false. It may be suppressed because it is true, or because it is doubtful whether it is true or false, and because it is not considered desirable that it should be discussed. In these cases there is obviously no assumption of infallibility in suppressing it (Stephen 1874, 41).

[9] For a useful characterization of the contrast between ideal and non-ideal epistemology, and the potential uses of the latter, see McKenna (2023).

range of self-serving or ideological biases.[10] Indeed, this is an important theme in Mill's own discussion and plausibly a crucial reason for why he defends free speech absolutism.[11]

Here, I want to bracket these non-ideal worries, however, and focus on a question in ideal theory: namely, what conditions must a censor meet if he is to be justified in censoring? Importantly, I will *not* assume that the general population, or the group who is being censored, is composed of ideal agents. These assumptions are thus the best case for justifying censorship—the would-be censors are assumed to be ideal epistemic agents (though not omniscient or infallible), while the agents who they censor, or on whose behalf they censor, are assumed to be non-ideal. If it turns out that even given these assumptions, the epistemic bar to be met for justifying censorship is very high, that will provide a strong presumption against censorship in general.

What sorts of idealizations are appropriate on behalf of the would-be censor? I want to be ecumenical here with respect to different theories of rationality and norms of reasoning. From a Bayesian perspective, we can assume they are probabilistically coherent and update their credences according to Bayes' Rule, perhaps with the addition that they have reasonable priors. Or more generally, we can suppose they follow the proper norms of reasoning (McHugh 2024), whatever they may be. We can assume that they have minimal propensities for motivated reasoning (Kunda 1990) or myside bias.[12] Further, we can add that they will not harbor irrational implicit biases or prejudices. In general, we may say, they do not possess the epistemic vices (Cassam 2016), whatever those might be.

However, there are certain dimensions along which idealization is not appropriate. Omniscience, for instance, would be too much. Rather, a useful stipulation might be that they have access to a set of evidence that can be reasonably expected given time and context. For example, would-be censors in Galileo's time cannot be expected to have satellite image evidence about the solar system. Similarly, a military commander during the Napoleonic Wars cannot be expected to have the evidence obscured by the fog of war. Furthermore, infallibility is not part of an appropriate idealization here—that is, sometimes we can be wrong about what we know and that's okay. Thus, there are cases where the would-be censor thinks they know something without actually knowing it. This sort of fallibility is mundane and need not display any epistemic vice. For example, suppose you park your car in the parking lot, with the appropriate permits and so on. Sitting in your office, you might believe the car is in the parking lot and take yourself to know this. However, out of a fluke of bad luck, your car was towed due to a mix-up from parking enforcement. Here, you take yourself to know something that you don't—but in doing so you do not display any epistemic vice. In what follows, I will assume the would-be censor can be fallible in this way.[13]

[10] The literature on such biases is vast, but for a useful overview of the relevant sort of bias—in particular "myside" bias—see Stanovich (2021). An important finding in this literature is that cognitive ability does not reduce this sort of bias (West, Meserve, and Stanovich 2012). See Kelly (2023) for a recent philosophical treatment of bias in general.

[11] Messina (2020) has recently interpreted the rationale for Mill's free speech absolutism along these lines. For a defense of the idea that Mill is a free speech absolutist, see Jacobson (2021). For a more qualified interpretation, see Brink (2013).

[12] There are some subtle issues here about how much it is appropriate to idealize away from myside bias. Some authors argue that this bias is a deep feature of human reasoning (Mercier and Sperber 2017). Regardless, we may say that from an ideal perspective, such a bias ought to be minimal, and the would-be censor is disposed to change their mind as appropriate when new evidence comes in, rather than digging in their heels.

[13] Thanks to an anonymous referee for pressing me to spell out the idealization in more detail.

2. Censorship on Epistemic Grounds

Suppose a censor believes that *P*. Further, he believes that whatever putative evidence could surface against *P* would either not be genuine evidence at all or that it would be weak or misleading. Suppose he is right. Is this enough to justify censoring arguments or putative evidence sharing against *P*?

Not quite. Even weak evidence is helpful. It can rationally affect how confident we ought to be about some proposition. Suppose I toss a coin which, for all I know, may or may not be fair. It comes out heads for 9 of the first 10 tosses—so I form a belief that it is not fair, with reasonably high confidence. When it is tossed for the 11th time, it comes out tails. This is useful evidence, though it doesn't settle the matter. I am still justified in believing the coin is not fair, but I should take my confidence down a notch. What's more, further coin tosses can lead to my doxastic attitudes becoming more *accurate*. So, in principle, allowing weak evidence to surface should be seen as an epistemic *benefit* not a cost.

Consider now the scenario where future evidence against *P* would be non-existent. That is, any putative counterevidence *E* would not in fact be evidence against *P*. However, not censoring putative evidence sharing against *P* would mean that some people might share that putative counterevidence which they mistakenly believe supports *P* but in fact it does not.

Even here, though, there is presumably a *prima facie* reason to allow people to voice their opinions, on grounds of protecting autonomy. Or it may be thought, as Shiffrin (2014) has argued, that free expression is necessary for individuals to develop as thinkers, and for that reason they ought to be allowed to make mistakes. Of course, people will disagree about how strong such reasons could be, and how they trade off against other reasons, for example, of preventing harm. But it's highly implausible that people should not be allowed to voice opinions or make arguments (even if they are false or unsound) if there is *no* cost to others.

So, from an epistemic point of view, there needs to be another supplementary assumption. This assumption would be something like the following: allowing people to share putative evidence against *P* would deteriorate our collective epistemic position because some significant proportion of people would take the wrong lessons from such discourse.[14]

To make things more precise, consider the following situation. Arthi would like to share a piece of evidence E_P that she believes counts against *P*. However, E_P, though it seems to count against *P* does not in fact count against *P*. Boris, however, is liable to draw the wrong lesson—he is liable to thinking that E_P does in fact count against *P*. Arthi's sharing E_P then leads to a deterioration of Boris's epistemic position. He may even lose knowledge that *P*. Further, there may also be higher-order effects to consider. Indeed, we adopt many of our beliefs via testimony from others. And the fact that others believe something (or make a particular inference) can often be a good reason for us to believe it too (or make that inference), assuming we appropriately trust them. The fact that they believe it is higher-order evidence for us (Levy 2021; 2022). Thus, Boris's coming to doubt *P* on the basis of E_P also has epistemic effects on others who trust him. The effect is compounded to the extent that Arthi has a large enough audience of this kind, i.e., if there are lots of Borises.[15]

[14] Similar issues have been discussed in the recent literature on "epistemic paternalism." Jackson (2022, 137) helpfully defines epistemic paternalism as "(i) intentionally and significantly interfering with someone's inquiry, (ii) without their consent, (iii) for their own epistemic good." The censor being discussed here has a goal analogous to condition (iii), though how exactly to cash out an individual's or group's epistemic good will remain an open question. For a canonical treatment of epistemic paternalism see Goldman (1991). See Ahlstrom-Vij (2013) for a more recent, detailed defense of such paternalism.

[15] Thanks to an anonymous referee for raising this possibility.

Now, Catherine anticipates such a situation and thereby censors Arthi. In so doing, she prevents Boris's (and others') epistemic situation from deteriorating. It might be thought here that in general it can often be easy to know that a particular piece of evidence does not count in favor of doubting a proposition. And if so, then depending on the severity of the likely epistemic deterioration, this could justify censorship which takes the form of disallowing a particular sort of argument: namely that E_P is evidence against *P*. We can call this "narrow" censorship. What can justify such narrow censorship, from the purely epistemic perspective? For one, the censor must be sufficiently justified in thinking that E_P is in fact *not* evidence against *P*. Further, she must be justified in thinking that a significant portion of the relevant population is *gullible* with respect to *P* and E_P. Like Boris, they will take E_P to count against believing *P* even though it does not. I will discuss the gullibility assumption later on.

But for now, notice that when we share evidence in this way, sometimes the following can occur. Suppose Arthi claims that E_P is evidence against *P*. David, when he encounters this claim rightly sees that E_P is not evidence against *P*—however, he realizes that E_P supports some other claim, *Q*, which he had not antecedently believed, and so now comes to rightly believe *Q*. This is an epistemic upgrade for David, as well as for others who David might persuade. But Catherine's censorship prevents this. So, the censor must also be (rationally) confident that such upgrades will not occur or will be relatively insignificant.

Moreover, what a piece of evidence supports depends on one's background *total evidence* (Kelly 2008b). Thus, even if E_P may not count against *P* relative to Catherine's total evidence, it might do so relative to David's. And this might be a genuine and significant epistemic upgrade for David. Of course, if Catherine's total evidence is larger and better-quality than David's when it comes to issues like *P*, then the intervention might be appropriate from her end. (Compare: a novice might look at a low powered study which supports some conclusion *P* and come to believe *P*. But an expert might know that this is misleading because high powered studies and meta-analyses do not support *P*.) But what this brings out, again, is that Catherine must be in an extremely good epistemic position. If David's total evidence has important elements that Catherine is not privy to, then there's a risk of precluding an epistemic upgrade for David.

Usually though, censorship takes the form of disincentivizing the giving of any putative evidence against some claim *P*, which, from the censor's perspective, is strongly or decisively supported by the total evidence. This form of "wide" censorship requires an even stronger epistemic position. For the censor must now be warranted in thinking that the evidence in favor of *P* is so strong that any evidence *E* that could (in some relevant sense of 'could') be uncovered would not significantly affect the rationality of believing *P*. And as in the previous case, there needs to also be a rationally made gullibility assumption. Thinking that both these conditions are jointly satisfied requires one to be in an unusually strong epistemic position with respect to *P*, which I will argue is rarely the case—and almost never the case with respect to typical issues of public or scientific controversy. In what follows, I will focus on wide censorship, given that it is the more commonly implemented form.

2.1 Knowledge is Not Enough

How strong must the censor's epistemic position be? It might be thought that the requisite condition is knowledge. If the censor *knows* that *P*, then she is justified in blocking future putative evidence against *P*.[16]

[16] In the intrapersonal case, this sort of reasoning leads to the well-known Kripke-Harman paradox. Harman (1973, 148) characterizes it as follows:

Now, knowledge does seem to justify action in a unique way.[17] If I know that it's raining outside, that justifies my action of carrying an umbrella. Likewise, if you lack knowledge about certain things, that justifies taking precautions. Presumably, we buy health insurance because we don't know that we will not fall ill. But it's not obvious that knowledge can justify dogmatic future-oriented epistemic policies. Suppose I know that we are meeting at Applebee's tonight. That does not justify a policy of not checking to see if there's a message from you (suppose, as it turns out, there is a message from you but, to preserve knowledge, suppose it simply says you're running a few minutes late).[18] Plausibly, knowledge is consistent as well with norms requiring us to be open to potentially disconfirming future evidence.

The crucial point here is a distinction between the sorts of actions that knowledge justifies. If I know we are meeting at Applebee's, this justifies my getting into the car and driving there, for example. But this doesn't mean that very knowledge justifies disregarding future evidence—say, ignoring text messages. As John Biro (2022, 1) has recently put it:

> One's attitude to evidence is governed not by what one knows but by what one thinks one knows. Thinking that one knows something does not entail that it is true. Knowing *this*, one knows that there may be non-misleading evidence against what one thinks one knows and should be open to examining what purports to be evidence against it.

Another way to frame this point is to distinguish between subjective and objective oughts. It can often be the case that I objectively ought to do something that I subjectively ought not to do. To give a simple example, suppose there is a hidden button somewhere at the coffee shop where I am typing these words such that if it were to be pressed, that would end world poverty. In some sense I ought to look for and press the button—that is the objective sense. But in another sense, I ought not to look for this button, given that I have no reason to think such a button exists. This is the subjective sense, which is the ought relevant to action *guidance*. And moreover, it is the sense which invites critical appraisal: blame, praise, and so on.[19]

> If I know that *h* is true, I know that any evidence against *h* is evidence against something that is true: so I know that such evidence is misleading. But I should disregard evidence that I know is misleading. So, once I know that *h* is true, I am in a position to disregard any future evidence that seems to tell against *h*.

In general, this seems like bad reasoning, leading to, among other things, the implication that a teacher cannot give her students a surprise test. Harman's solution to the paradox involves denying that we can rationally disregard evidence on such grounds, because new evidence can destroy old knowledge. Harman (1973, 149) writes:

> Since I now know [my car is in the parking lot], I now know that any evidence that appears to indicate something else is misleading. That does not warrant me in simply disregarding any further evidence, since getting that further evidence can change what I know. In particular, after I get such further evidence I may no longer know that it is misleading.

For an alternative solution, based on indicative conditionals, see Sorensen (1988). In a response to Sorensen, Veber (2004) argues that in general, we are not in a position to disregard evidence even when we know it is misleading.

[17] See Hawthorne and Stanley (2008).

[18] This particular case may not be compelling to all readers. Some might get the intuition here that in such cases I don't really know we're meeting at Applebee's. However, the reader may substitute here cases they find more compelling as needed—see for example Harman's (1973) parking lot example referenced in the above footnote.

[19] This distinction has invited a large literature, but for an influential discussion see Smith (2010). Thanks to an anonymous referee for inviting me to say more here.

Part of the general point here is that while knowledge entails truth, we often do not have direct access to the truths that we know. And this is what would be required to justify an epistemic policy of disregarding future possible evidence. In other words, mere belief that one knows something does not justify disregarding future evidence—even if that belief is correct, i.e., one in fact possesses that knowledge. Though Mill doesn't put it in these terms, this is one way in which we can understand the "assumption of infallibility" that he is concerned with. Mill puts things in terms of certainty. "To refuse a hearing to an opinion," he writes, "because they are sure that it is false, is to assume that *their* certainty is the same thing as *absolute* certainty. All silencing of discussion is an assumption of infallibility" (Mill 1859, 22).

However, we can also give a similar gloss based on knowledge. An infallible creature would never be wrong about what she knows. Thus, whenever she thinks she knows *P*, she actually knows *P*. Perhaps different epistemic norms would apply to such agents. But fallible creatures are sometimes wrong about what they know—that is, sometimes they think they know *P* even when they do not. That we are fallible creatures, Mill hopes to convince the reader using inductive evidence—on both an individual and group level, we can acknowledge that we have been wrong about what we know in the past. It is thereby implausible to think that we now are infallible, though we have been fallible in the past.[20]

The important point for our purposes is that the admission of fallibility is consistent with possessing knowledge. Fallible agents can know things. Moreover, knowing that *P* is not sufficient for an epistemic policy of disregarding future evidence against *P*. Now, the censor does not quite disregard future evidence against *P*. Rather, she supplies a particular sort of incentive: she incentivizes others to avoid uncovering or sharing putative evidence that counts against *P*. However, this is structurally analogous to an individual agent avoiding future evidence against *P*—for instance, it is like me turning off my phone and not checking email once we have made plans to meet at Applebee's. If there is evidence that there has been a change of plans, it's unlikely to make its way to me.[21]

Now, some philosophers have worried that a certain kind of disposition towards future evidence can lead people to lose knowledge in a problematic way. For instance, Jeremy Fantl argues that open-minded engagement with certain arguments has this feature. Open-mindedness, on his characterization, involves, inter alia, being "willing to be significantly persuaded conditional on spending significant time with the argument, finding the steps compelling, and being unable to locate a flaw" (Fantl 2018, 12). Thus, for example, the average Eleatic might not be able to locate a flaw in Zeno's ingenious (but misleading) argument against motion. Such a person would lose knowledge, namely that things move, by open-mindedly engaging with Zeno's argument.

However, note that the epistemic rationales against censorship need not entail open-minded engagement of this form.[22] The latter involves a much more demanding requirement. Furthermore, there are plausible norms about how to weigh evidence in such cases that need not lead to loss of knowledge. In particular, we might think that "common sense" can provide us with evidence that

[20] Strictly speaking, the censor need not in fact *think* that she is infallible. However, for Mill, the act of censorship involves the *assumption* of infallibility—that is, it involves taking an action that only infallibility could justify. For a defense of this interpretation see Turner (2013).

[21] Indeed, a further, stronger claim is also plausible. Not only ought I not block future evidence from reaching me, but it can be rational for me to double-check the time/location as I head out. For discussion, see Woodard (2022).

[22] Moreover, weaker conceptions of open-mindedness need not have these consequences—see, for example, Kwong (2016) for an account of open-mindedness which requires engagement, but which is consistent with having firm beliefs.

defeats such revisionary arguments—so that the average Eleatic can reasonably retain his belief in motion even after encountering Zeno's arguments and being unable to find a flaw (Kelly 2008a; 2011).

Now it might be thought that if knowing that *P* is not enough to justify wide censorship with respect to *P*, perhaps knowing that one knows that *P* could be enough. But now suppose the would-be censor knows that she knows that *P*. As brought out in the Biro quote from earlier though, her attitude towards evidence relevant to the proposition *I know that* P should be governed not by the fact that she knows that proposition but by the fact that she thinks she knows it. Thus, she knows that there may be non-misleading evidence against the proposition *I know that* P. Now what could such evidence consist in? Unlike the proposition *P* itself, there are additional pieces of evidence that could be relevant. Thus, perhaps there is evidence that she formed the belief that *P* in an unreliable way and recognizing this could defeat her belief that she knows that *P*.[23] But note that whatever evidence would defeat the belief that *P* would *also* defeat the belief *I know that* P, given that knowledge is factive. Whatever would defeat my belief that my car is in the parking lot would also defeat my belief that I know that my car is in the parking lot. If the arguments above are right, then she must be open to examining evidence against *P* as well—even if she knows that she knows that *P*. So if knowledge is not enough, neither is higher-order knowledge.

2.2 Justification of Censorship Decays Over Time

Suppose at some point in time, the censor meets whatever epistemic condition is necessary to justify censorship. Assuming the preceding discussion is right, this has to go further than knowledge. In his critical discussion of the Millian project, David Lewis imagines a censor who has done his epistemic due diligence. He has considered arguments from both sides with maximum open-mindedness and has accumulated evidence that is enough to meet a very high bar. Lewis (1989, 160) writes:

> Our Inquisitor, if he takes Mill's word for this as he does on other matters, will not dare suppress heresy straightaway. First he must spend some time in free discussion with the heretics. Afterward, if they have not changed his mind, then he will deem himself justified in assuming the truth of his opinion for purposes of action; which he will do when he goes forward to suppress heresy, and burns his former partners in discussion at the stake.

Nishi Shah (2021) has recently argued that even such due diligence cannot justify censorship, on a close reading of Mill. Shah's thought, roughly, is that epistemic justification requires certain dispositions on behalf of the agent—namely dispositions to be open to future evidence as it may arise. The very act of censorship *demonstrates* that such dispositions are absent—thus, censorship is a uniquely self-undermining act. In other words, to be justified in censoring arguments against *P*, one must be justified in believing *P*—but this requires one to be open to future evidence. Any act of censorship reveals that one lacks this disposition, and thereby is not justified in believing *P* at the outset. This offers a natural way of reading Mill's (1859, 24) remarks on what justification requires, as for example when he writes: "Complete liberty of contradicting and disproving our opinion, is the very condition which justifies us in assuming its truth for purposes of action; and on no other terms can a being with human faculties have any rational assurance of being right."

Shah's diagnosis is something like this. On the modern picture of evidence and justification, the focus is on particular *beliefs*, and so epistemologists are interested primarily in the question of whether a particular *belief* held by a person is justified. And here, a natural and influential idea is that a belief is justified insofar as it's properly based on the available evidence. On Shah's reading of Mill,

[23] In particular, this is if one is attracted to reliabilism as a theory of knowledge (Goldman and Beddor 2021). But the reader may substitute the analogous conditions depending on their favored theory.

the primary locus of justification is the *person*, not the *belief*. We can then be interested in the question: in what way must a particular belief be held by a person in order for that *person* to be justified in holding that belief? And here, the Millian thought is roughly that open-mindedness and sensitivity to future evidence are virtues/dispositions that a person must have in order to be justified in believing something.

But suppose we reject this picture. Suppose that my future-directed orientations are not relevant to justification—what matters simply is whether I form the belief *P* on the basis of sufficiently good evidence.[24] Even in this case, I argue, the censor faces an extremely difficult epistemic task.

First, consider indexical propositions of the form: *It is raining*. Here, even a maximally good evidential position will not justify blocking future evidence. Suppose my perception is highly reliable and I look out the window and see that it is raining. This does not justify an epistemic policy of avoiding future evidence for the simple reason that it might stop raining. Moreover, as time goes on, my justification will decay if I shield myself from relevant evidence. Thus, imagine that I go into the basement for a few hours—assuming I don't check the weather from there in some way, presumably I lose justification for thinking that it is raining, even though I had perfectly good justification a few hours ago while looking out the window.

Thus, the more interesting case of censorship involves claims that are meant to hold across time. These might involve claims about scientific principles or historical facts. Suppose our censor then has at time t_0 accumulated sufficiently good evidence, such that he can be rationally confident that there is no available evidence out there that can defeat *P* or cast significant enough doubt on *P*. Let's grant, on this basis, that at time t_0 the censor is justified in blocking arguments against *P*, for the sake of the epistemic good of others.

Such a situation, though, is *unstable* in an important sense. For, consider the scenario at a sufficiently distant future time, t_1. Here, there is a two-fold epistemic worry. First, between t_0 and t_1, evidence might have been uncovered such that it genuinely counts against *P* but is not shared. The problem is that there is a significant epistemic possibility that—insofar as the censorship has been successful—we are operating on a biased subset of the total available evidence. This itself provides a defeater for *P*.[25]

A second worry is that censorship also undermines incentives to *uncover* evidence, for the following reasons. First, discovering new evidence often involves reasoning with others. Scientists, for example, are not usefully thought of as solitary thinkers discovering and analyzing new evidence by themselves—rather, they are embedded in communities and rely on extensive communication to generate ideas, test hypotheses, etc.[26] Effective censorship can disrupt these channels of communication and thus hamper the generation of evidence. Second, many professional rewards center around the sharing of evidence—for example, by publication in journals, presenting talks, and so on. By blocking these possibilities, effective censorship can dramatically reduce the incentives to conduct research the conclusions of which cannot be published or shared in other ways.

To use Mill's own example, consider Newtonian physics. Mill (1859, 26) writes, "If even the Newtonian philosophy were not permitted to be questioned, mankind could not feel as complete assurance of its truth as they now do." Suppose that in 1859, when Mill wrote *On Liberty*, censors had succeeded in stopping future disputations of the theory. The thought then is that someone

[24] We can assume, for instance, evidentialism about justification as defended in Feldman and Conee (1985).

[25] The idea here is that awareness of unpossessed evidence can provide us with defeaters. For a detailed defense of this point see Ballantyne (2015; 2019).

[26] For helpful discussion on this, drawing on recent work in cognitive science and psychology, see Mercier and Sperber (2017) and Sloman and Fernbach (2017).

could not now be as confident about the truth of Newtonian physics as they would have been in 1859—for, future evidence generation and sharing against the theory would have been disincentivized. Thus, a theory like Einstein's General Theory of relativity would not have made its way to us. This possibility itself defeats our justification for believing in Newtonian physics now, under the censorious regime.

Hence, the epistemic position of the censor (along with everyone else) deteriorates from t_0 to t_1. There are two noteworthy implications of this fact. First, even if the censor was justified, *ex hypothesi*, in blocking putative evidence against *P*, at time t_0, he is no longer justified in doing so. Second, the censor's original goal—of improving the epistemic condition of others—has been frustrated. In an important sense, the epistemic condition of others has *deteriorated*, for while they originally had a justified belief (in Newtonian physics, say) they no longer are justified in holding that belief. That is because there is now a significant possibility of undiscovered or unshared evidence, and that possibility makes it rational to doubt Newtonian physics.[27]

2.3 Predicting Future Evidence

To justify censorship as a rationally stable policy, then, the censor must not only have compelling evidence for *P*, at time t_0, he must also be able to predict, with high-enough rationally justified certainty, that epistemically significant counterevidence will not emerge over time. One problem, however, is that evidence generation occurs within *complex* social systems, with many strategically interacting individuals and institutions.[28] When it comes to such systems, it is difficult to predict their evolution, and particularly, how interventions within a particular area might affect behavior in another area.

These difficulties are illustrated in some of Philip Tetlock's (2005; 2015) work on forecasting. Even experts find it hard to be reliable at predicting the behavior of complex social and political systems. Furthermore, the more accurate predictors display more open-mindedness and intellectual humility, and often give probabilistic judgments rather than yes or no answers. But this cuts against the rationality of censorship. The more someone is uncertain about what future evidence might be uncovered and how it might bear on *P*, the less they can rationally block that evidence from surfacing.

2.4 Gullibility

As discussed earlier, the justification of censorship also requires the assumption of gullibility on the part of the general population. The basic worry is that if putative evidence against *P* is allowed to be shared, then a significant portion of the population will draw the wrong lesson—they might lower credence in *P*, suspend judgment, or come to believe not-*P*, even though this is not warranted.

Part of the challenge for establishing gullibility of this sort will involve grappling with recent work in cognitive psychology which suggests that we are for the most part epistemically vigilant (Sperber et al. 2010). A core idea here is that gullibility is maladaptive; because competitors often have incentives to deceive us at our expense, gullible traits would have been selected against within human populations. Further, attempts at mass persuasion based on false information often fail. In part this is because we have dispositions to engage in "plausibility checking" (Mercier 2020) and we critically assess the information and arguments presented to us, in light of our background beliefs.

[27] For a recent analysis of doubt in terms of significant epistemic possibility, see Moon (2018).

[28] See Holland (2014) for a brief overview of complexity.

Nonetheless, there might be cases in which experts and novices process evidence in different ways. That is, it might take some significant expertise to recognize that some putative evidence E_P against *P* is misleading. Hence, while from a novice perspective, it might look like E_P counts against *P*, a genuine expert in the subject matter can see that E_P does not in fact count against *P*.

How might we deal with this issue? At first glance, it might seem that censorship can be a useful tool here for epistemic improvement with respect to the novices. However, if the preceding discussion is on the right track, then censorship is an extremely *blunt* tool. For, effective censorship will not only prevent such epistemic downgrades caused by misleading but seemingly convincing evidence, but it will also prevent genuine and compelling evidence against *P* from surfacing if it exists. A minor bruise on one's foot is better dealt with using some anti-septic and a band-aid, rather than amputation. So, if there are such solutions, we should use them.

One such possible solution is "evidential preemption" by the relevant experts (Begby 2021). The basic idea here is that an expert can preempt misleading evidence by saying something like: you will encounter misleading evidence E_P but that should not lead you to doubt *P*. Insofar as the preemptor is a genuine expert and recognized by the novice to be so, the latter can rationally discount E_P even if she does not understand the reasons *why* that evidence is misleading.

This, of course, is not a radical idea, and is in fact quite commonplace. Suppose you get an email from an unknown author who purports to have a new revolutionary theory of quantum mechanics and time. Out of curiosity, you read on. However, because of the technicality of the material and your novicehood with respect to physics, you find some of the claims to be seemingly supported by the presented evidence. This is typically not enough to change your beliefs though—if you're really interested in the issue, you might talk with a friend who is a physicist or philosopher of physics, and if she points out that that evidence is misleading and already accounted for in the best theories, you will (rationally) simply dismiss the arguments of the email. Of course, I don't mean to suggest that such preemption will work all the time. But I want to suggest that such preemption is potentially one tool we might use among others.[29]

3. Censorship on Grounds of Harm

In many interesting cases, the rationale for censorship is the prevention of harm.[30] As a general matter, it is possible that censorship regarding some issues might promote overall well-being.[31] And if promoting well-being is one of our goals then it's not obvious why censorship in some cases is not legitimate—whenever we are optimizing with respect to multiple goals, there will be trade-offs. There presumably will be some harms done by the censors, particularly to those whose speech is

[29] Part of the challenge here might be the "illusory truth" effect, wherein claims that are often repeated are perceived to be true even if they are false; for a recent overview, see Brashier and Marsh (2020). Depending on the strength of this effect in particular cases, it might be difficult for experts to get out ahead and preempt poor evidence consistently. The implications of the illusory truth effect for public discourse are interesting in their own right, and I don't have the space to do justice to these issues here. However, I want to note that it's not obvious that the presence of this effect supports censorship—in fact, it might do the opposite. For, if only a partial subset of the total evidence is allowed to be discussed, then claims made on the basis of that partial subset will be oft repeated (especially if the issue is significant), leading people to be more confident in them than is warranted. Thanks to an anonymous referee for bringing this point to my attention.

[30] For a recent defense of limiting scientific inquiry on these grounds, see the Nature Editorial titled "Science must respect the dignity and rights of all humans" (2022).

[31] Stanley Fish (1994), for example, has influentially argued that complete freedom of speech is bound to conflict with certain other goals we might have, and for that reason there are appropriate limits. For critical discussion of Fish's argument, see Jacobson (2004).

curtailed, but it's possible that these are outweighed by the harms that are prevented by the censorship.[32]

Now, one might think that as a matter of right, states ought not to restrict expressions of opinion, even if these expressions cause harm. One might thus endorse what Scanlon calls the "Millian Principle," against government censorship:

> There are certain harms which, although they would not occur but for certain acts of expression, nonetheless cannot be taken as part of a justification for legal restrictions on these acts. These harms are: (a) harms to certain individuals which consist in their coming to have false beliefs as a result of those acts of expression; (b) harmful consequences of acts performed as a result of those acts of expression, where the connection between the acts of expression and the subsequent harmful acts consists merely in the fact that the act of expression led the agents to believe (or increased their tendency to believe) these acts to be worth performing. (Scanlon 1972, 213)

However, note that Scanlon's argument trades on a certain view of the proper relationship between citizens and the state. In particular, a state should have only those powers which citizens could allow while still seeing themselves as equal, autonomous, and rational agents (Scanlon 1972, 215). Construed as an argument for limits on state power, however, the view leaves open much by way of censorship by private agents.

Moreover, suppose someone accepts some other moral theory which does not assign an independent right to free expression. Or perhaps they think there is a prima facie duty to avoid censorship, for reasons of preserving autonomy, say, but the duty can be overridden where some threshold of anticipated harm is met.[33] What might the case against censorship look like for such a person?

3.1 Justification and Action

In section §2.2, it was argued that censorship degrades justification over time. In a situation where an enforced norm of censorship has persisted for some significant time with respect to *P*, we can be less sure of whether *P* is in fact true. However, this degrades the extent to which we can act rationally on the basis of *P*.

An interesting case here is one in which the censor himself is agnostic as to whether *P*, or perhaps even believes *P* is a "noble lie." In this case, the censor's actions themselves do not presume the truth of *P*, and hence we might think the degradation of justification as to whether *P* does not affect the rationality of the censor's actions—namely of disincentivizing the giving of arguments against *P*.

However, if the censor is truly successful, the general population will believe *P* even though they are not justified in believing *P*. In particular, they might have lots of first-order evidence in favor of *P*, and little or no first-order evidence against *P* (given that the censor has been successful). Nonetheless, because they are non-ideal epistemic agents, they may not properly notice or weigh the higher-order evidence against *P*, namely that they are likely operating on a biased subset of the total evidence, and the experts to whom they defer do not have truth-tracking incentives.

Furthermore, the subject matter of claims regarding which censorship is exercised is likely to be *practically significant*. Censorship, historically, has been applied with respect to religious, political, or social scientific claims which bear significantly on how people should act, which norms they should endorse, and what policies they should support. Thus, we don't observe censorship with respect to,

[32] For a recent discussion of the potential harms of silencing, see Cohen (2020).

[33] Cf. W. D. Ross (1930).

say, how many blades of grass there are in a particular backyard. But this means that insofar as censorship is successful, many people will be acting as if *P* is true, on matters of great practical significance, without being justified in thinking that *P* is true.

I want to simply note here that this is one potential cost of censorship that must be tallied in the final moral cost-benefit analysis. On a wide range of moral views, it is a moral cost to do something practically significant without being justified in believing those things that would rationalize it. For example, it would be wrong for a shipowner to send people on a voyage with his ship if he is not justified in believing it is seaworthy (Clifford 1877). Likewise, it is wrong for a doctor to prescribe a drug to a patient if she hasn't taken adequate stock of the side-effects.

That said, it is a familiar point that sometimes, having certain beliefs can be practically beneficial even if those beliefs are not fully warranted. Thus, for instance it might be that an interviewee's having a more positive view of their own credentials and abilities than is strictly warranted by the evidence can increase their chances of getting the job. Similarly, all other things equal, an underdog team which somewhat irrationally thinks it can win is more likely to actually win than a team which forms an accurate perception of the matchup.[34] Epistemic and practical rationality can sometimes come apart.

In this vein, Mill imagines the following type of justification for censorship. With respect to a certain class of our beliefs:

> The claims of an opinion to be protected from public attack are rested not so much on its truth, as on its importance to society. There are, it is alleged, certain beliefs, so useful, not to say indispensable to well-being, that it is as much the duty of governments to uphold those beliefs, as to protect any other of the interests of society. (Mill 1859, 27)

3.2 The Harm Thesis

As Mill notes though, this sort of justification pushes the relevant analysis up one level. For the censor to act justifiably in this regard, he must be rationally confident of the claim that *the belief in P is helpful to society*, or conversely, *denying P is harmful to society* even if he need not be confident of *P* in itself. Mill (1859, 27) writes, "The usefulness of an opinion is itself matter of opinion: as disputable, as open to discussion, and requiring discussion as much, as the opinion itself."

This observation presents several challenges. First, it is often difficult to tell which views are helpful or harmful to society, irrespective of their truth. Of course, we might have various hunches, but rigorously establishing such theses about harm is difficult in the case of society in general.[35] In the case of establishing what sorts of beliefs may help individuals, irrespective of their truth, we might conduct statistical analyses based on large enough samples of data or find ways to conduct randomized controlled experiments. However, it is difficult if not impossible to use such methods when it comes to entire societies.

Second, even if it is rigorously established that belief in *P* is beneficial to society at some time t_0, it won't do to simply assume that it is similarly beneficial at some sufficiently distant future time t_1. Plausibly, which views are beneficial for society to hold will depend on a range of contingent factors—including technology, resources, competition, and so on. The "noble lies" we might appropriately tell during wartime could be very different from those we might tell during times of peace. Myths that may have had important functions within hunter-gatherer conditions might not have similarly beneficial functions now. The upshot here is that propositions like *belief in P is beneficial*

[34] See James (1896) for the classic discussion on this point.

[35] For a recent critique along these lines see Clark et. al. (2023).

to society must constantly be tested for plausibility, and it's hard to see how such testing can be done under conditions of censorship.

In particular, we would have to construct an ethos where, while there are strong disincentives to share evidence against *P*, there is robust freedom to discuss arguments for and against whether belief in *P* is beneficial. Now, as mentioned before, there are beliefs where some level of inaccuracy is beneficial—some examples include the optimism bias (Sharot 2011) and the placebo effect.[36] However, note that there's presumably a degree of inaccuracy beyond which it's no longer beneficial to have that belief. No amount of optimism bias can get me to defeat Floyd Mayweather in a boxing match. Taking arsenic as a placebo will not cure someone's cancer. This is one way to make precise for our purposes Mill's (1859, 27) claim that "the truth of an opinion is part of its utility."

So, in some sense there must be an appropriate band, as it were, where the disconnect of the belief from reality is not too large, in which censorship may serve some good practical purposes. Thus, even in the best case, the censorship cannot be absolute—it cannot disincentivize people from giving evidence that *P is sufficiently far away from reality*. Even here then, appropriate censorship would be very limited. Furthermore, in a collection of diverse agents, there is the possibility of a kind of cascade effect. Suppose there are bits of evidence E_1 to E_n, which taken together would support thinking that P is disconnected enough from reality (like the proposition that I can defeat Mayweather in a boxing match). But each piece of evidence only marginally makes the case. Now suppose these pieces of evidence are dispersed across agents A_1 to A_n. What is the censor to do here? Here's one policy she may have: censor those statements that do not show that P is sufficiently disconnected from reality. But that would mean censoring each agent, A_1 through A_n, because none of their individual arguments is enough to make the case by itself. But these actions would, taken together, lead to a situation where P is sufficiently disconnected from reality but wrongly thought to be beneficial.

3.3 Culpable Ignorance

An underexplored potential consequence of censorship, particularly regarding matters of practical significance, is action out of culpable ignorance. In the standard sort of case of culpable ignorance, an agent can be blameworthy even if she makes the best choice in light of the evidence she has at time t_0, because she fails to acquire some important evidence at a previous time t_{-1} (Smith 1983; Rosen 2003). Thus imagine a doctor prescribing some remedy to a patient, which is the best given her current evidence. However, had she done the assigned reading while in medical school, she would have known that this remedy has very bad side effects for patients of this particular sort, as compared with some alternative remedy which is nearly as effective. Here, the doctor acts out of ignorance and is plausibly blameworthy for the act of prescribing the remedy with bad side effects.

However, censorship has important similarities with this case because it constitutes a method of blocking future evidence from surfacing. This feature can put the censor in a structurally analogous situation to the doctor above. Suppose that at time t_0, some action Φ is best supported by the evidence the censor has. However, had she not disincentivized evidence generation by beginning a policy of censorship at t_{-1}, the best available evidence would have recommended a different action Ψ. It seems that the censor is now plausibly blameworthy for the harms caused by Φ-ing rather than Ψ-ing. This possibility of culpable ignorance is part of the normative burden a censor must take on.

[36] For a recent review of the placebo effect in the context of adult depression, see Jones et. al. (2021).

Furthermore, there is also the issue of responsibility for the actions of *others* who are ignorant due to the censor's actions, but not culpably so. Thus, suppose there is compelling evidence that Mustard committed the murder. However, Plum hides this evidence from the authorities, which leads the courts to convict Scarlett, which is, let's suppose, rational given what evidence they have. Here, Plum is blameworthy for the wrongful conviction of Scarlett because he blocks the evidence pointing to Mustard.

Likewise, suppose that some set of policies Φ is rationalized by the subset of evidence that the censor allows. However, if the full set of evidence—i.e., the total evidence that would be available had there been no censorship—were considered, the appropriate set of policies would have been Ψ.[37] The harms caused by adopting Φ rather than Ψ, then, are plausibly at least partially attributable to the censor. This is an important moral risk the censor takes on when he enacts censorship on practically significant matters.

Of course, the censor might luck out if the stars align in the following way: the exact set of policies that are warranted given the total set of evidence are also warranted given the subset of evidence the censor allows in public policy deliberation. This would constitute a sort of moral luck.[38] However, precisely because censors are usually interested in practically significant evidence the stars are unlikely to align in such a way.

4. Implications and Cases

I have been arguing that epistemically, censorship is akin to disregarding future evidence and thus suspect for that reason. It might be wondered though, what guidance this observation can provide. Surely, some instances of censorship are justified—and moreover, it's easy to generate cases. Suppose an evil demon threatens the destruction of the world (and we are justified in believing this is a real threat, not a hallucination and so on) unless a true but mundane claim about yesterday's weather is censored. Here, the right action, it seems to me, is clearly to censor this claim. The harm is stipulated to occur, and the censored claim is not even significant. Since it's implausible that an anti-censorship principle will always trump other normative considerations, there will be cases where censorship is justified. At any rate, from a dialectical standpoint, I do not need to persuade those who antecedently believe in such a strong anti-censorship principle.

But we might also wonder if there is something more useful to be said that could provide guidance is more messy, real-world cases. The main thing to notice is that inquiry is overwhelmingly a *collective* project. This is part of the core insight in modern work in social epistemology (cf. Hardwig 1991; Goldman 1999). Scientific discoveries, for example, take place within a context of many different inquirers gathering and analyzing evidence, proposing alternate hypotheses and so on (Mercier and Sperber 2017). And sound public policy decisions must be made by incorporating evidence that is dispersed across society (Anderson 2006). For these processes to work well, the incentives faced by the different inquirers must be sufficiently aligned—and this is what censorship disrupts.

In light of these observations, censorship is more likely to be justified the more *localized* its potential effects are. For example, if I start a mystery novel book club with a handful of people and forbid the discussion of quantum physics there, I am not in any significant way disrupting the process of scientific inquiry into quantum physics. Second, there are specific contexts where one agent has such a level of evidential superiority relative to others that the relevant inquiry is not in any

[37] For several historical examples of this phenomenon, see Sunstein (2019).

[38] Cf. Williams (1976) and Nagel (1979).

important sense collective. This can be the case in certain contexts, as when you are talking to your 5-year-old about eating vegetables or when a schoolteacher addresses her 2nd grade math course. But this condition is unlikely to hold when it comes to, for example, scientific censorship (cf. Clark et al. 2023).[39] Third, the justification of censorship depends on the epistemic (and practical) significance of the relevant proposition.[40] Censoring claims about the number of blades of grass in your backyard is thus different in this sense from censorship about claims regarding the age of the earth or evolutionary science or economic policy. From this point of view, the sort of censorship defended in Plato's *Republic*, for instance, would be a paradigm case of the unjustified kind.[41] Even supposing that the rulers could be justified in acting as they do with regards to various matters at the outset, their actions would, over time, lose justification insofar as they assumed the truth of the orthodoxy in defense of which the censorship was instituted. For, belief in that orthodoxy would itself be less and less justified over time for reasons explored in §2.2.[42]

5. Conclusion

Discussion around freedom of expression has focused on the rights or interests of speakers and hearers on one hand, and the harms that speech might cause on the other. If we treat the question abstractly, the main exercise seems to be to weigh these different moral considerations and then claim either that speech ought to be protected or that limitations are legitimate.

But censorship is enacted *by* somebody. So then there arises a question: what position must *that agent* (or agents) be in so that her actions are justified? Though this is part of Mill's focus, the question is relatively neglected in modern discussions of the topic. In this paper, I have argued that the censor must be in an unusually good epistemic position for her censorship to be justified. And I have tried to make this case by drawing on a separate set of literatures that examine norms of how moral and epistemic agents ought to collect and process their evidence over time.

Now, it is a familiar point in non-ideal theorizing that any person to whom we might give coercive powers is but human and thus is prone to moral and epistemic error.[43] Power corrupts. However, this paper has aimed to give a sort of best-case scenario for the would-be censor. The censor has been idealized, while the general population has not. The only limitations on the censor I have assumed are lack of omniscience and fallibility. If the arguments presented here are correct, then even given this idealization, the censor faces an unusually difficult epistemic task, and one that requires an especially strong evidential position. Furthermore, the censor takes on *moral risks* for the actions of others who will act out of ignorance due to the censor's policy. The censor's burden, then, is greater than it might seem at first blush.

[39] For a range of different cases where scientific censorship has been defended, see the Clark et al. paper.

[40] For a helpful discussion of significance in the scientific context, see Kitcher (2001).

[41] Plato (1997).

[42] Thanks to the referees for pressing me to say more about how the view might handle different cases.

[43] The literature here is vast, but for a recent philosophical treatment of non-ideal theory as it applies to state action, see Freiman (2017).

References

Ahlstrom-Vij, Kristoffer. 2013. *Epistemic Paternalism: A Defence*. London: Palgrave Macmillan.

Anderson, Elizabeth. 2006. “The Epistemology of Democracy.” *Episteme* 3 (1–2): 8–22. https://doi.org/10.3366/epi.2006.3.1-2.8.

Ballantyne, Nathan. 2015. “The Significance of Unpossessed Evidence.” *The Philosophical Quarterly* 65 (260): 315–35.

———. 2019. *Knowing Our Limits*. New York: Oxford University Press.

Begby, Endre. 2021. “Evidential Preemption.” *Philosophy and Phenomenological Research* 102 (3): 515–30.

Biro, John. 2022. “‘Dogmatism’ and Dogmatism.” *Episteme*, 1–5. https://doi.org/10.1017/epi.2022.15.

Brashier, Nadia, and Elizabeth Marsh. 2020. “Judging Truth.” *Annual Review of Psychology* 71. https://doi.org/10.1146/annurev-psych-010419-050807.

Brink, David. 2013. *Mill’s Progressive Principles*. New York: Oxford University Press.

Cassam, Quassim. 2016. “Vice Epistemology.” *The Monist* 99 (2): 159–80.

Chemerinsky, Erwin. 2019. *Constitutional Law*. 6th Edition. Philadelphia: Wolters Kluwer.

Clark, Cory J., Lee Jussim, Komi Frey, Sean T. Stevens, Musa al-Gharbi, Karl Aquino, J. Michael Bailey, et al. 2023. “Prosocial Motives Underlie Scientific Censorship by Scientists: A Perspective and Research Agenda.” *Proceedings of the National Academy of Sciences* 120 (48): e2301642120. https://doi.org/10.1073/pnas.2301642120.

Clifford, W.K. 1877. “The Ethics of Belief.” *Contemporary Review* 29:289–309.

Cohen, Andrew Jason. 2020. “Harms of Silence: From Pierre Bayle to De-Platforming.” *Social Philosophy and Policy* 37 (2): 114–31.

Editorial. 2022. “Science Must Respect the Dignity and Rights of All Humans.” *Nature Human Behavior* 6:1029–31.

Fantl, Jeremy. 2018. *The Limitations of the Open Mind*. Oxford, New York: Oxford University Press.

Feldman, Richard, and Earl Conee. 1985. “Evidentialism.” *Philosophical Studies* 48 (1): 15–34.

Fish, Stanley. 1994. *There’s No Such Thing As Free Speech, and It’s a Good Thing, Too*. New York: Oxford University Press.

Freiman, Christopher. 2017. *Unequivocal Justice*. New York: Routledge.

Goldman, Alvin. 1991. “Epistemic Paternalism: Communication Control in Law and Society.” *Journal of Philosophy* 88 (3): 113–31.

———. 1999. *Knowledge in a Social World*. Oxford: Oxford University Press.

Goldman, Alvin, and Bob Beddor. 2021. “Reliabilist Epistemology.” *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/reliabilism/.

Hardwig, John. 1991. “The Role of Trust in Knowledge.” *The Journal of Philosophy* 88 (12): 693–708.

Harman, Gilbert. 1973. *Thought*. Princeton, NJ: Princeton University Press.

Hawthorne, John, and Jason Stanley. 2008. “Knowledge and Action.” *Journal of Philosophy* 105 (10): 571–90.

Holland, John. 2014. *Complexity: A Very Short Introduction*. New York: Oxford University Press.

Jackson, Elizabeth. 2022. “What’s Epistemic about Epistemic Paternalism?” In *Epistemic Autonomy*, edited by Jonathan Matheson and Kirk Lougheed. New York: Routledge.

Jacobson, Daniel. 2004. “The Academic Betrayal of Free Speech.” *Social Philosophy and Policy* 21 (2). https://doi.org/10.1017/S0265052504212031.

———. 2021. “A Defense of Mill’s Argument for the ‘Practical Inseparability’ of the Liberties of Conscience (and the Absolutism It Entails).” *Social Philosophy and Policy* 37 (2): 9–30. https://doi.org/10.1017/S0265052521000029.

James, William. 1896. *The Will to Believe*. New York: Longmans, Green, and Co.

Jones, Brett D. M., Lais B Razza, and Cory R Weissman. 2021. "Magnitude of the Placebo Response Across Treatment Modalities Used for Treatment-Resistant Depression in Adults: A Systematic Review and Meta-Analysis." *JAMA Network Open* 4 (9): e2125531. https://doi.org/10.1001/jamanetworkopen.2021.25531.

Joshi, Hrishikesh. 2021. *Why It's OK to Speak Your Mind*. New York: Routledge.

Kelly, Thomas. 2008a. "Common Sense as Evidence: Against Revisionary Ontology and Skepticism." *Midwest Studies in Philosophy* 32 (1): 53–78.

———. 2008b. "Disagreement, Dogmatism, and Belief Polarization." *The Journal of Philosophy* 105 (10): 611–33.

———. 2011. "Following the Argument Where It Leads." *Philosophical Studies* 154 (1): 105–24.

———. 2023. *Bias: A Philosophical Study*. New York: Oxford University Press.

Kitcher, Philip. 2001. *Science, Truth, and Democracy*. New York: Oxford University Press.

Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108 (3): 480–98.

Kwong, Jack M. C. 2016. "Open-Mindedness as Engagement." *Southern Journal of Philosophy* 54 (1): 70–86.

Lackey, Jennifer. 2021. *The Epistemology of Groups*. New York: Oxford University Press.

Levy, Neil. 2021. "Virtue Signalling Is Virtuous." *Synthese* 198:9545–62. https://doi.org/10.1007/s11229-020-02653-9.

———. 2022. *Bad Beliefs: Why They Happen to Good People*. New York: Oxford University Press.

Lewis, David. 1989. "Mill and Milquetoast." *Australasian Journal of Philosophy* 67 (2): 152–71.

List, Christian, and Philip Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. New York: Oxford University Press.

Luper, Steven. 2016. "Epistemic Closure." *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/closure-epistemic/.

McGowan, Mary Kate. 2019. *Just Words: On Speech and Hidden Harm*. New York: Oxford University Press.

McHugh, Conor. 2024. "Norms of Reasoning." *Philosophy Compass* 19 (7). https://doi.org/10.1111/phc3.13008.

McKenna, Robin. 2023. *Non-Ideal Epistemology*. New York: Oxford University Press.

Mercier, Hugo. 2020. *Not Born Yesterday: The Science of Who We Trust and What We Believe*. Princeton, NJ: Princeton University Press.

Mercier, Hugo, and Dan Sperber. 2017. *The Enigma of Reason*. Cambridge, MA: Harvard University Press.

Messina, James P. 2020. "Freedom of Expression and the Liberalism of Fear: A Defense of the Darker Mill." *Philosophers' Imprint* 20 (34): 1–17.

———. 2023. *Private Censorship*. New York: Oxford University Press.

Mill, John Stuart. 1859. "On Liberty." In *On Liberty and Other Essays (2008)*, edited by John Gray. Oxford, New York: Oxford University Press.

Moon, Andrew. 2018. "The Nature of Doubt and a New Puzzle about Belief, Doubt, and Confidence." *Synthese* 195:1827–48.

Nagel, Thomas. 1979. *Mortal Questions*. Cambridge, UK: Cambridge University Press.

Plato. 1997. *Plato: Complete Works*. Edited by John Cooper. Indianapolis, IN: Hackett Publishing.

Rosen, Gideon. 2003. "Culpability and Ignorance." *Proceedings of the Aristotelian Society* 103:61–84.

Ross, W. D. 1930. *The Right and the Good*. Oxford: Oxford University Press.

Scanlon, Thomas. 1972. "A Theory of Freedom of Expression." *Philosophy & Public Affairs* 1 (2): 204–26.

Shah, Nishi. 2021. "Why Censorship Is Self-Undermining: John Stuart Mill's Neglected Argument for Free Speech." *Aristotelian Society Supplementary Volume* 95 (1): 71–96.

Sharot, Tali. 2011. "The Optimism Bias." *Current Biology* 21 (23): R941–45.

Shiffrin, Seana. 2014. *Speech Matters: On Lying, Morality, and the Law*. Princeton, NJ: Princeton University Press.

Simpson, Robert Mark. 2020. "The Relation between Academic Freedom and Free Speech." *Ethics* 130 (3): 287–319.

Sloman, Steven A., and Philip Fernbach. 2017. *The Knowledge Illusion: Why We Never Think Alone*. New York: Riverhead Books.

Smith, Holly. 1983. "Culpable Ignorance." *The Philosophical Review* 92 (4): 543–71.

———. 2010. "Subjective Rightness." *Social Philosophy and Policy* 27 (2): 64–110.

Sorensen, Roy. 1988. "Dogmatism, Junk Knowledge, and Conditionals." *The Philosophical Quarterly* 38:433–54.

Sperber, Dan, Fabrice Clement, Christophe Heintz, Oliver Mascaro, Hugo Mercier, Gloria Origgi, and Dierdre Wilson. 2010. "Epistemic Vigilance." *Mind & Language* 25 (4): 359–93.

Stanovich, Keith E. 2021. *The Bias That Divides Us*. Cambridge, MA: MIT Press.

Stephen, James Fitzjames. 1874. *Liberty, Equality, Fraternity*. 2nd ed. London: Smith, Elder, & Co.

Sunstein, Cass. 2019. *Conformity*. New York: New York University Press.

Tetlock, Philip E. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.

Tetlock, Philip E, and Dan Gardner. 2015. *Superforecasting: The Art and Science of Prediction*. New York: Crown.

Turner, Piers Norris. 2013. "Authority, Progress, and the 'Assumption of Infallibility' in On Liberty." *Journal of the History of Philosophy* 51 (1): 93–117.

Veber, Michael. 2004. "What Do You Do with Misleading Evidence?" *The Philosophical Quarterly* 54 (217): 557–69.

Waldron, Jeremy. 2012. *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press.

West, Richard F, Russell J Meserve, and Keith E Stanovich. 2012. "Cognitive Sophistication Does Not Attenuate the Bias Blind Spot." *Journal of Personality and Social Psychology* 103 (3): 506–19.

Williams, Bernard. 1976. "Moral Luck." *Proceedings of the Aristotelian Society, Supplementary Volume L*, 115–35.

Woodard, Elise. 2022. "Why Double-Check?" *Episteme*. https://doi.org/10.1017/epi.2022.22.