SERRC
Social Epistemology
Review & Reply Collective

http://social-epistemology.com
ISSN: 2471-9560

Putting Hate Speech at Its Place: A Political-Epistemological Perspective

Massimiliano Badino, University of Verona, massimiliano.badino@univr.it

---

Badino, Massimiliano. 2024. "Putting Hate Speech at Its Place: A Political-Epistemological Perspective." *Social Epistemology Review and Reply Collective* 13 (2): 47–55.
https://wp.me/p1Bfg0-8Ag.

SERRC
Social Epistemology
Review & Reply Collective
13 (2): 47–55. 2024.
https://wp.me/p1Bfg0-8Ag

The pollution of our epistemic environments is a pressing problem. The largest share of our belief system is assembled through interactions with other agents mediated by digital epistemic environments such as social networks, forums, news websites, and the like. Recently, Neil Levy has argued that the emergence of bad beliefs is not so much the result of suboptimal belief-formation practices as the fact that such practices are applied within polluted environments (Levy 2021). Among the factors that make our digital lives difficult, hate speech is arguably the most troublesome. It has become so pervasive and arduous to eradicate that even large language models have been recently mobilized to face its threat (Vishwamitra et al. 2023; Kikkisetti et al. 2024).

However, what precisely constitutes hate speech? As is typical with social phenomena, the definition of hate speech is inherently intricate and multifaceted. Generally, hate speech is any expression encouraging or stirring up hatred towards a specific human group. To be more precise, I will adopt the analysis of Bhikhu Parekh (2012), who singles out three major characteristics of hate speech.

> First, hate speech is directed against a group of people "based on arbitrary and normatively irrelevant features" (Parekh 2012, 40). This is tantamount to saying that hate speech is *discriminatory*.
>
> Second, by ascribing qualities that are widely regarded as undesirable, hate speech stigmatizes the target group. In other words, it is *derogatory*.
>
> Third, hate speech expresses hostility toward the target group and insists on its marginalization or expulsion from society. Hate speech is implicitly or explicitly *exclusionary*.

Section 1: The Epistemic Case Against Hate Speech

My aim in this paper is to discuss the epistemic case against hate speech, namely whether there is an epistemic ground for banning hate speech from our epistemic environments in addition to the well-known political and ethical reasons. In my view, the problem of hate speech must be tackled from a political-epistemological perspective. In terms of its effect on society and its epistemic dynamics, hate speech is not essentially different from other political-epistemological phenomena, such as epistemic injustice. In a way, one can even state that it is the most extreme and detestable form of epistemic injustice, as Parekh's definition well elucidates.

Furthermore, arguments against the censorship of hate speech are typically political-epistemological in nature. For example, the rejection of censorship as a viable option against hate speech is typically justified as a defense of free speech. In turn, two types of arguments typically support free speech: (i) arguments appealing to the importance of free speech for democracy (Parekh 2012, 42-43; Post 2011), and; (ii) Mill-style arguments appealing to the necessity of taking into account all positions in our search for truth (Schauer 2012).

In other words, the case against hate speech concerns our belief-formation practices, and it is, in this sense, genuinely epistemic. But to the extent that these practices occur and manifest their effects in the social and political arena, they are inherently political. In the concluding remarks, I will show how the political-epistemological perspective helps us see the policy-oriented side of my epistemic argument. In Section 2, I analyze Wendy Xin's recent paper on hate speech (Xin 2023), which constitutes an excellent starting point. While I accept some of the key steps of her argument, I reject its general spirit, and I draw different consequences, which I develop into a novel argument in Section 3.

Section 2: Hate Speech: The Lesser of Two Evils?

In her recent article published on *Social Epistemology*, Wendy Xin formulates an epistemic argument to defend the claim that one should censor hate speech. Of course, epistemic arguments are not decisive in deciding whether censorship of hate speech is a viable option. As mentioned above, traditional arguments appeal preferably to moral or political considerations. According to Xin, these reasons would not be touched by the epistemic analysis of the issue: what remains to be decided is whether the epistemic perspective provides a *pro tanto* reason in favor of or against censorship of hate speech. As I have already stated in the previous section, I do not entirely agree with this point because I believe that hate speech is one of those phenomena that require a political-epistemological analysis, and it is impossible to separate out political and epistemic arguments neatly. I will, however, concede the point for the time being. I will return to it in the concluding remarks.

Xin's argument starts out with two important claims:

1. Hate speech cannot be formulated in a neutral way.

2. Censoring hate speech ends up creating an epistemic bubble.

The first claim means that no re-formulation or rephrasing of hate speech maintains the same content purged by the hatred it contains. Xin adopts a characterization of hate speech slightly different from mine. According to her, hate speech features: (i) discrimination against the target group; (ii) encouragement of hatred, and; (iii) harm to the dignity of the target group. The difference from my previous definition is not substantial, though.

The important point to make, nevertheless, is that hate speech possesses its defining features not only in virtue of its form but also in virtue of its content, i.e., the specific claim it conveys. If hate speech is inherently discriminatory, derogatory, and exclusionary, it is because what it says entails and justifies discriminatory, derogatory, and exclusionary behavior. Take Holocaust denial as an example. It conveys the view that the Holocaust never happened or happened at a much smaller scale than usually admitted, which is tantamount to saying that a specific group, the Jews—identified by normatively irrelevant features—is promoting a gigantic deceit for the purpose of gaining credit or making the rest of the world feel guilty. This is already discriminatory, and if we add the fact that the

SERRC
Social Epistemology
Review & Reply Collective
13 (2): 47–55. 2024.
https://wp.me/p1Bfg0-8Ag

Holocaust did, in fact, historically happen, it immediately becomes derogatory and exclusionary, regardless of how such a claim is expressed.

Once claim (1) is established, claim (2) follows suit as an almost immediate consequence. As there is no way to express hate speech in a neutral fashion, we cannot just censor the form, we must censor the content as well:

> Considering that the views expressed in (at least some) hate speech cannot simply be described by a neutral counterpart, censorship of hate speech inevitably leads to the omission of certain views. Censorship bubbles are thus formed: people who live in a community that censors hate speech also live in an epistemic bubble that excludes views expressed in hate speech (Xin 2023, 3).

Xin calls these "censorship bubbles," i.e., epistemic bubbles generated by the obliteration of the views contained in hate speech. Now comes the next step:

> 3. Hate speech also creates bubbles.

Hate speech creates an epistemic bubble in two ways. First, it silences the target group through its intimidating and discriminatory acts. Building on the fact that speech is a form of action, Xin argues that hate speech performs "locutionary silencing, illocutionary disablement, and perlocutionary frustration," which end up forming epistemic bubbles "because the audience is not exposed to certain views from the target group of hate speech." Thus, hate speech creates bubbles by filtering out the voices of the target group because it discourages them from expressing their own views.

One might argue that there is a substantial difference between hate bubbles and the classical examples of epistemic bubbles. In the latter, external views are filtered out by simple omission, while in the former, silencing occurs through violent and direct action. Hence, one might think that someone from the dominant group would stand up and protest against hate speech, thus bursting the bubble. This typically does not happen, according to Xin, which leads us to the second way hate bubbles work. Individuals from the dominant group are usually not culturally equipped to stand up in support of the target group:

> Because our social positions play an important role in shaping our experiences, it follows that minority groups might have certain knowledge that dominant social groups do not have, because the former have certain experience that the latter do not have. (...) The experience of minority group identities thus motivates one to gain certain knowledge that the experience of dominant group identities fails to motivate (Xin 2023, 6).

More importantly, hate speech is almost invariably associated with the identity of the group that expresses it, and individuals of the dominant group might be afraid to voice opinions

that undermine their belonging to the group. I will comment on both points later on, but for the time being, I will concede them.

It turns out that both censorship and hate speech generate epistemic bubbles because they are both mechanisms to select and filter out external views. The final step of Xin's argument amounts to proving the following claim:

4. Hate bubbles are more epistemically problematic than censorship bubbles.

Why is it so? The reason lies in an asymmetry in the behavior of hate bubbles and censorship bubbles. In brief, the former has a strong tendency to degenerate into echo chambers, while the latter do not. This asymmetric behavior originates from the effects that hate bubbles have on the target as well as the dominant group. Hate can easily turn into distrust and discredit, thus making people imprisoned in a bubble increasingly resistant to external views. On the contrary, censorship bubbles do not run this risk. Censorship does not mean a judgment on the hate speaker and can be removed at any moment. Censorship is directed against the label, not against the agent.

The conclusion follows swiftly now. Although both hate speech and its censorship pollute our epistemic environments, creating subcommunities that filter out certain views, censorship is still preferable because it does not degenerate into echo chambers the way hate speech does. In other words, we are epistemically better off censoring hate speech. Thus, this line of argument reinforces the moral and political reasons by showing that censoring hate speech is a good practice from an epistemic perspective.

In the next section, I will discuss some critical junctures in Xin's argument. While I agree with some of her key points, I will build on different steps and develop an entirely different argument to support what is basically the same conclusion.

Section 3. Hate Speech and Epistemic Content

I will begin by noticing that Xin's argument has a sort of consequentialist flavor. In essence, she argues that censorship and hate speech are both suboptimal practices from the epistemic standpoint because they deprive us of certain views. However, the former is the lesser evil: it creates bubbles, but not of the "bad" type. All in all, applying censorship will leave us epistemically better off, or, more precisely, less worse off. Personally, I find this argumentative strategy dissatisfactory because my intuition is that hate speech is so epistemically rotten that its obliteration from our epistemic environments cannot merely be "somewhat less bad". Intuitively, hate speech does not enrich our epistemic life, and its disappearance would not impoverish it either, hence, a censorship policy of hate speech must have some merit in itself. In the following, I will try to develop this intuition into a more consistent line of argument, but, as a starting point, I wish to analyze some of the steps in Xin's article to show that it has other weak spots besides my personal dissatisfaction with the argumentative architecture.

SERRC
Social Epistemology
Review & Reply Collective
13 (2): 47–55. 2024.
https://wp.me/p1Bfg0-8Ag

Let us focus on claims (2) and (3) above, which are the hinges of Xin's argument. She holds that hate bubbles and censorship bubbles are both epistemic bubbles, but the former is worse because they tend to become echo chambers more easily. On the contrary, I want to stress that there are reasons to think that: (i) hate bubbles are very fragile epistemic bubbles, and; (ii) censorship bubbles already have built-in the key features of echo chambers.

Let's begin with the first point. I agree that hate speech causes the typical filter effect that constitutes an epistemic bubble: it tends to eliminate dissenting voices. However, we know that epistemic bubbles are fragile social structures (Nguyen 2020), and hate bubbles seem even more fragile for two reasons: (i) they usually contain a very tiny minority of people, and; (ii) they can be burst by people from the dominant group who stand up and defend the target group from the hate attacks. Xin lucidly recognizes this point. She is well aware that, potentially, hate bubbles could be easily isolated from society by the concerted action of the target and the dominant group. To respond to this objection, she argues that the defensive action of the dominant group would be ineffective because they do not share the same experiences as the target group and cannot speak for their identity. This argument seems to me a clear example of the perfectionist fallacy. Merely because members of the dominant group have not personally experienced the same life circumstances, it does not follow that they are incapable of effectively condemning hate speech as pernicious and damaging behavior.

Let's move on to point (ii): censorship bubbles are arguably very similar to echo chambers. Among the features of echo chambers, the most interesting for my argument is that they operate a dramatic redistribution of epistemic trust to the effect that sources inside the chamber acquire a disproportionate amount of credit, while sources outside the chamber are wholly discredited (Nguyen 2020). Does censorship produce a similar effect? Xin holds that this is not the case because censorship hits the label, not the holder:

> As such, we have no strong reason to believe that censorship bubbles have a similarly strong tendency to turn into echo chambers as hate bubbles do. Even though labelling something as hate speech also makes us assign less epistemic merit to it, such an effect might simply disappear when the label is removed. Moreover, censorship of hate speech does not necessarily turn the rest of us against people who express hateful views, because they are not necessarily demonised or dehumanised as a result of censorship (Xin 2023, 9).

I see a hidden assumption in Xin's claim, namely that echo chambers must operate through the strong emotional support of demonization or dehumanization of the opponents. But is it an inherent feature of echo chambers or rather an element contingent upon the specific social structure of the chamber itself? Strong feelings, such as demonization, are not necessary companions to epistemic discredit. It is true that some echo chambers work this way. I submit, however, that this is typical of small communities, which hold wildly divergent views and need strong emotional support to hold tight to their point against the majority. My point is that when discredit is directed against official sources, which are considered

trustworthy by the majority, such discredit must be sustained by a robust emotional overtone (oftentimes in the context of some grandiose and evil conspiracy) to be maintained.

When discredit comes from the majority, though, this emotional overtone is not required and almost seems exaggerated. This is precisely what happens with censorship. In my view, systematic and institutionalized censorship is the strongest form of epistemic discredit one can think of. Nothing, I believe, can better express discredit than the systematic removal of certain views from the public debate. Furthermore, censorship is indeed directed against the supporter of hate speech *qua* supporter of hate speech as a polluter of our epistemic life. True, censorship does not demonize hate speakers and leaves the door open for rendition, but this is simply because it does not need to. From a purely epistemic perspective, it seems to me that censorship bubbles work very similarly to echo chambers.

If these remarks are correct, Xin's consequentialist strategy has few chances to succeed. In the remainder of this section, I will develop an alternative to Xin's argument. Because of the limited space, my proposal will be necessarily sketchy, but the general line should be sufficiently clear.

My starting point is the conventional epistemic reason *not* to censor hate speech, i.e., epistemic libertarianism. This is the view that any opinion should be available in the marketplace of ideas because all opinions contribute to the epistemic goal of searching for truth. It's helpful to rephrase this famous argument in slightly more technical terms. Our epistemic life needs a certain dose of 'epistemic friction' generated by alternative views that challenge our opinions and force us to refine them. Epistemic friction is key to avoiding confirmation bias and seeing possibilities that would otherwise remain unnoticed. From this perspective, censorship is always an epistemically losing move.

In the abstract, epistemic libertarianism works very well; in concrete (i.e., from a political-epistemological perspective), it is disingenuous. It focuses on the process of acquiring true beliefs and avoiding false beliefs and fails to consider the epistemic agent that is supposed to perform such a process. More specifically, while epistemic friction is always beneficial—abstractly—for the process, this is not necessarily the case for the agent. This point has been excellently stressed by José Medina, who claims that knowledge requires acknowledgment (Medina 2013). According to Medina, excessive epistemic friction undermines our self-confidence and, derivatively, our ability to maintain knowledge and build upon it. Epistemic friction encourages volatility, while we also need stability and acknowledgment from our epistemic community.

But epistemic libertarianism is disingenuous in another respect. We live in extremely complex epistemic environments. We are bombarded by more sources and information than we can possibly manage and evaluate. Over the internet, we can easily access the most disparate views, theories, and conceptions. In such a situation, attention becomes a commodity and its economy a key epistemic factor (Smith and Archer 2020; Dutilh Novaes 2023). The necessity to modulate our attention in the belief-forming process makes epistemic libertarianism impractical and opens up a space for selection policies of the

SERRC
Social Epistemology
Review & Reply Collective
13 (2): 47–55. 2024.
https://wp.me/p1Bfg0-8Ag

epistemic candidates, i.e., doxastic items worth our scrutiny for their potential to contribute to the search for truth. Manuel De Pinedo and Neftalì Villanueva dubbed this process epistemic de-platforming (De Pinedo and Villanueva 2022). It is important to stress that epistemic de-platforming is more sophisticated than mere censorship:

> Epistemic de-platforming does not contain a set of recommendations to cancel every interlocutor that happens to disagree with the things that you know; it is simply an invitation to ponder whether that person is worth your time and attention in an epistemic environment in which both are severely limited (De Pinedo and Villanueva 2022, 125).

The underlying idea is that, given the social and political circumstances of our epistemic life, it is a rational policy to direct our limited resources toward serious epistemic candidates. Establishing the working criteria of this policy is a further important issue.[1] The relevant point for my argument, however, is that such a policy rests on rational grounds. For, my main claim is that if it is a rational policy in our epistemic environment to select serious epistemic candidates, then it is easy to argue that hate speech should be selected away. In other words, independently of the criteria adopted, hate speech cannot ever be taken as a serious epistemic candidate, hence, its censorship is justified.

Why cannot hate speech ever be an epistemic candidate? The main reason lies in Xin's claim (1) above: there is no way to formulate hate speech in a neutral fashion. The innermost motivation of hate speech is not to advance a view about the world that might contribute to our epistemic life but rather to attack a specific target group. In this sense, hate speech is not so much a statement about the world as an identity marker. Although it presents itself as a legitimate view about the historical course of events, the whole point of the Holocaust denial, and similar outrageously false affirmations, is not to submit a claim to the tribunal of reason but rather to mark belonging to a certain hate group.

As Helen Da Cruz has argued, the more outlandish a belief, the more identity-connotated (De Cruz 2020). The gist of holding very unpopular and egregious beliefs lies more in the fact that they act as a proxy for in-group recognition than in their content. But if hate speech is the expression of the identity of certain groups, then we don't lose anything by de-platforming them. If it is a rational practice to marginalize views that have no serious chances to be epistemic candidates, then, *a fortiori*, we are justified in de-platforming hate speech.

Section 4: Concluding Remarks

In the foregoing sections, I argue that Xin's argument has some weak points. In particular, I am dissatisfied with the consequentialist setup, and, even conceding it, I doubt that censorship bubbles are really substantially better than hate bubbles, at least in the terms that

[1] For example, De Pinedo and Villanueva argue that our epistemic community might provide us with good criteria to regulate our attention.

Xin uses to compare them. Furthermore, I accept Xin's claim (1), from which, however, I draw a different conclusion. While Xin sees (1) as the reason to state that censoring hate speech always generates an epistemic bubble, I regard it as tantamount to saying that hate speech is not a serious epistemic candidate because it is not supposed to contribute to our search for truth. Rather, hate speech is a marker for identity (a point also alluded to by Xin), hence, its censorship is not epistemically detrimental. If one adds that hate speech pollutes our epistemic environment, it is easy to conclude that censorship is even recommendable.

As much for the epistemic side of the story. Hate speech is a much larger issue, though. While censorship is a legitimate and epistemically sound defensive strategy, from a political-epistemological perspective, one probably needs to take a more constructive course of action. Hate speech often presents itself as a legitimate claim, whereas in fact, it is a manifestation of belonging. But political epistemology teaches us that identity and belonging have a role to play in the belief-forming process. It is at this juncture that, perhaps, we can find a space for what Davids, building on Judith Butler's work, calls the re-signification of hate speech (Davids 2018). While an epistemological analysis stops at justifying the censorship of hate speech as a polluting and fundamentally parasitic practice, a larger, political-epistemological perspective should invite us to find a more constructing way to deal with this phenomenon in the context of a democratic society.

References

Almagro Holgado, Manuel, Llanos Navarro Laespada, and Manuel De Pinedo García. 2021. "Is Testimonial Injustice Epistemic? Let Me Count the Ways." *Hypatia* 36 (4): 657–75. https://doi.org/10.1017/hyp.2021.56.

Davids, Nuraan. 2018. "On the (In)Tolerance of Hate Speech: Does It Have Legitimacy in a Democracy?" *Ethics and Education* 1–13. https://doi.org/10.1080/17449642.2018.1477036.

De Cruz, Helen. 2020. "Believing to Belong: Addressing the Novice-Expert Problem in Polarized Scientific Communication." *Social Epistemology* 34 (5): 440–452. https://doi.org/10.1080/02691728.2020.1739778.

De Pinedo, Manuel, and Neftalì Villanueva. 2022. "Epistemic De-Platforming." In *The Political Turn in Analytic Philosophy*, edited by David Bordonaba Plou, Víctor Fernández Castro, and José Ramón Torices, 105–34. Amsterdam: De Gruyter.

Dutilh Novaes, Catarina. 2023. "The (Higher-Order) Evidential Significance of Attention and Trust—Comments on Levy's Bad Beliefs." *Philosophical Psychology* 36 (4): 792–807. https://doi.org/10.1080/09515089.2023.2174845.

Kikkisetti, Dhanush, Raza Ul Mustafa, Wendy Melillo, Roberto Corizzo, Zois Boukouvalas, Jeff Gill, and Nathalie Japkowicz. 2024. "Using LLMs to Discover Emerging Coded Antisemitic Hate-Speech in Extremist Social Media." *arXiv*. http://arxiv.org/abs/2401.10841.

Levy, Neil. 2021. *Bad Beliefs: Why They Happen to Good People*. Oxford: Oxford University Press.

Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Oxford: Oxford University Press.

13 (2): 47–55. 2024.
https://wp.me/p1Bfg0-8Ag

Nguyen, C. Thi. 2020. "Echo Chambers and Epistemic Bubbles." *Episteme* 17 (2): 141–161. https://doi.org/10.1017/epi.2018.32.

Parekh, Bhikhu. 2012. "Is There a Case for Banning Hate Speech?" In *The Content and Context of Hate Speech: Rethinking Regulation and Responses*, edited by Michael Herz and Péter Molnár, 37–56. Cambridge: Cambridge University Press.

Post, Robert. 2011. "Participatory Democracy and Free Speech." *Virginia Law Review* 97 (3): 477–489.

Schauer, Frederick. 2012. "Social Epistemology, Holocaust Denial, and the Post-Millian Calculus." In *The Content and Context of Hate Speech: Rethinking Regulation and Responses*, edited by Michael Herz and Péter Molnár, 129–143. Cambridge: Cambridge University Press.

Smith, Leonie and Alfred Archer. 2020. "Epistemic Injustice and the Attention Economy." *Ethical Theory and Moral Practice* 23 (5): 777–95. https://doi.org/10.1007/s10677-020-10123-x.

Vishwamitra, Nishant, Keyan Guo, Farhan Tajwar Romit, Isabelle Ondracek, Long Cheng, Ziming Zhao, and Hongxin Hu. 2023. "Moderating New Waves of Online Hate with Chain-of-Thought Reasoning in Large Language Models." *arXiv*. http://arxiv.org/abs/2312.15099.

Xin, Wendy. 2023. "Censorship Bubbles Vs Hate Bubbles." *Social Epistemology* 17 (2–3): 1–12. https://doi.org/10.1080/02691728.2023.2274324.