

September 18, 2022

Semester of Summer '22

Words: 2360 of 6000

LMU Munich

Faculty for Philosophy,

Theory of Science
and Study of Religion

Simulating free speech

An agent based modeling approach

PD Dr. habil. Jürgen Landes

Module: P3 Analysis, reflection and critique of
basic problems in theoretical philosophy

Seminar: Social Epistemology

Cedric Tonye-Djon

Master: Theoretical Philosophy

Student Number: 12387354

cedric.tonyedjon@gmail.com

Contents

1	The problem of free speech	3
2	The basics of agent based modeling	5
3	Exploring restrictions on free speech	6
4	Conclusion	8

Abstract: In this short piece I am exploring the epistemic value of limitations on free speech using agent based modeling. I start by reconstruction Mills arguments for absolute free speech. I then introduce epistemic agent based models in the style of Hegselmann-Krause. Finally, I explore an extension of the Hegselmann-Krause model that incorporates different types of censorship.

1 The problem of free speech

The extent to which speech should be free has been philosophically debated for centuries. In his classic *On Liberty* Mill ([1859] 2011) famously argues that everyone should be allowed to say anything. That is, unless they directly harm someone else through saying that thing. He is vehement that the mere falsity, wrongness or harmfulness of expressed beliefs does not warrant their censorship. And although he is talking about political norms, his reasoning is mainly epistemic.

In a nutshell, there are, according to Mill, only three scenarios whenever some expression is censored:

1. the expressed belief is true
2. the expressed belief is partially true and partially false
3. the expressed belief is false

Let's assume that we are the dominant group with the power to censor some belief contrary to ours. If the belief is true then censoring it will move us away from the truth, which is not what we want. Also, human beings don't know anything infallibly. Therefore there is a chance, that any belief contrary to ours is true and we need to debate all of them in order to know for sure. If the belief is partially true and partially false then we are robbing ourselves of an opportunity to learn the truth that is in it. And even if the belief is completely false we should not censor it. Because, so Mill argues, we would miss an opportunity to defend our true belief against a capable and motivated adversary. If we keep doing this, it will cause our belief to

turn into a stale, meaningless and unjustified dogma. Therefore, on Mills account, we collectively loose on the epistemic front whenever we censor. Our assesments become less accurate, less rational, less justified and less truthful.

This is still quite abstract so I will give some concrete examples. According to Mill anyone should be allowed to utter any statement about religion, science, politics or anything else. They should be allowed to distribute political pamphlets even if these are revolutionary, antidemocratic, populist or fearmongering. People should be allowed to predict the end of the world or fantasize about global conspiracies. The only situation where speech should be limited is when the angry mob is literally standing outside some politicians house, torches and pitchforks in hand and this utterance would cause them to storm in and lynch them.(**mill**)

1859 was a long time ago. Since then our world has witnessed the emergence of social media and with it new challenges for political deliberation. Mills ideas were controversial when he first proposed them. Some of them remain just as controversial today. Although free speech is cemented into many state constitutions and a part of the *Universal Declaration of Human Rights*, there are some limitations. For example, while the USA allows it, Germany outlaws Holocaust Denial because the expressed ideas are so deeply wrong that they are dangerous. In addition to that, private social media platforms regulate a large portion of all political deliberation worldwide. They also break with Mill. For example, the social media platform Twitter banned Donald Trump, the former president of the USA while he was still in office. JAN6... Twitter also disabled *Articles of Unity*...**BOTSRESEARCH** In public discourse, some people ask for the social media platforms to finally censor dangerous ideas while others moan their perceived suppression under “cancel culture”. So what are we to make of Mills ideas today? How free should free speech be? Is Mill correct in claiming that any censorship is an epistemic failiure? Or can some forms of it be justified and if so, which ones?

2 The basics of agent based modeling

Agent based models (ABM) are high level representations of interactions between agents. There are boundless options as to what can be considered as an agent in this sense. One option is to model epistemic agents, that is, agents that aim at attaining truth, understanding, knowledge or something of that nature. In an ABM these agents are highly idealized. Their beliefs about the world are represented by a couple of numbers and their actions determined by a small set of rules. This idealization is the source of a big strength and the biggest weakness of ABM. On the one hand, it allows researchers to concentrate on a select few variables. This brings the studied phenomenon clearly into focus while pushing the messy complexities of actual people into a blurry background. On the other hand, it makes the results of the model dependent on the idealizing assumptions being made. Therefore simulations using these models can generate quasi-empirical data. But this data is even more susceptible to scepticism and reinterpretation than empirical data from studies and experiments.

There are two more major strengths when using agent based modeling to explore social epistemology. One strength is that it is easier to perform than human trials. All that is needed is a computer, basic coding skills and some imagination. This is not only cheaper than trying out ideas on humans, it is also safer. The other strength is that in human trials it is excruciatingly hard to operationalize notions like truth, knowledge, understanding or rationality in political discourse. The researchers are a part of the deliberating community — just like their subjects. In order to know what interventions increase truthfulness in some experimental debate they themselves have to know the truth and what is true is exactly what is up for debate. In an ABM this can be solved easily. Truth, rationality, understanding, justification or whatever can be built into a model and different strategies at attaining it can be compared. It's as simple as: The truth is that $\tau = 42\%$.

One approach to ABM of social epistemology was developed by Hegselmann and Krause 2009. Their idea was to represent an agent as an integer $1 \leq i \leq n$. To each agent there corresponds a number $x_i(u) \in (0, 1]$ which represents its beliefs about

or assesment of some state of affairs. For simplicity, this assesment only has one dimension. The truth that each agent aims at through their assesments is $\tau \in (0, 1]$. Each agent also has a group of peers $X_i(u) = \{j : |x_i - x_j| < \epsilon\}$. These are the agents whose assesments are sufficently close to their own. With each time step every agent updates their assesment based on the truth and the mean assesment of their peers at the former time step. $\alpha \in [0, 1]$ represents how strongly they trust their peers. Riegler and Douven (2010) later extended this model by limiting communication between agents to those that are close in a two dimensional space and by introducing random noise into agents access to the truth. To keep things manageable, I am introducing noise but otherwise staying with Hegselmann and Krauses original model. Therefore agents update their beliefs like:

$$x_i(u + 1) = \left(\frac{\alpha}{|X_i(u)|} \sum_{j \in X_i(u)} x_j(u) \right) + ((1 - \alpha)(\tau \cdot \text{noise}))$$

3 Exploring restrictions on free speech

I extended this model in order to explore the epistemic impact of restrictions on free speech. I considered three scenarios. In each scenario there is a silenced group S and all other agents are in the dominant mainstream group M . These groups determine who agents can listen to, meaning who they can have in their peer group. For each scenario there are two settings. The silenced group can keep their trust in the mainstream. In this case the only beliefs any agent takes into account are those of the agents in the mainstream group: $X_i(u) = \{j : |x_i - x_j| < \epsilon \wedge j \in M\}$. But censorship can lead the people that are being silenced to loose trust in the media in which they are censored. This may cause them to seek out alternative means of communication. I represent this phenomenon in the model with a setting that allows agents to loose trust in the mainstream. In this case silenced agents only take into account the beliefs of other silenced agents: $X_i(u) = \{j : |x_i - x_j| < \epsilon \wedge j \in S\}$.

The three types of censorship are by range, arbitrary group and popularity. Censorship by range represents one powerful entity like a state or a social media plat-

form censoring a range of beliefs that they deem to be wrong to be expressed. In this case, the silenced group at a time step is the set of all agents with assessments within the range: $S(u) = \{i : \phi \leq x_i(u) \leq \psi\}$. Arbitrary group silencing represents censorship based on arbitrary non-epistemic criteria like skin color, gender, heritage, etc. Here, a fixed ratio of agents $\kappa \in [0, 1]$ is permanently assigned to the silenced group: $\{i : i < \text{round}(\kappa \cdot n)\}$. Lastly, people can face a social kind of silencing. They can be shunned, insulted and bullied for their beliefs by individual people, leading them to withdraw from public discourse without any institution formally censoring them. I represented this by letting the silenced group at time u be a proportion ρ of the agents with the largest mean distance from their assessment at u to that of any other.

With Python, TkInter and matplotlib I created a graphical desktop application that can be used to easily play around with these parameters. Through experimentation I noticed some tendencies. All of these are in scenarios where the silenced group keeps trust in the mainstream.

If a range of beliefs is censored that is far outside the truth, for example with $(\phi = 0, \psi = 0.2, \tau = 0.4)$ then deliberation barely changes. This can be observed quite clearly by comparing a simulation with censorship to one without where all the random values stay the same:

If, however, a range around the truth is censored, it can lead agents to form polarized groups around the truth, and these groups won't always converge.

It can also push all agents out of the range into one direction, away from the truth.

This model doesn't show an epistemic disadvantage of silencing wrong beliefs, but agents also don't justify their beliefs here. They just evaluate some state of affairs and then access their evaluations directly. Since the benefits that Mill lists for allowing the expression of false beliefs relate to their justification, it is therefore not surprising that the model does not show any of them.

Arbitrary silencing didn't show a noteworthy effect under most conditions. It slightly decreased the accuracy of the group when almost all agents were part of the silenced group and trusting in the mainstream ($\kappa = 0.99$). This is likely because the few mainstream agents remaining had fewer resources to explore and therefore

could not beat the noise as well as the larger compared group.

This seems implausible in reality. It could be an artifact of the idealizing assumption that all agents are equally likely make any observation. In reality, an epistemic disadvantage in silencing people based on the colour of their skin, their gender and so on might be that they have important things to say that the dominant group overlooks (**harding1992**;...). If this is the case, a more accurate model of arbitrary silencing should assign different noise functions to agents belonging to the different groups.

I didn't find any cases where silencing beliefs because they were unpopular had a significant impact. This also seems implausible in reality. The reason for this happening could be similar to the one above. The model assumes that at each time step agents go out into the world and are able to make any observation any other agent could make. In reality, as Mill points out, one of the largest epistemic dangers that we face in silencing unpopular beliefs is missing out on the true and partially true observations that are contrary to our own and hidden within.

Another significant limitation of this model is that I am evaluating freedom of speech from a purely epistemic standpoint. In reality, other things matter too. Objectifying and dehumanizing pornography, for example, could be censored in order to avoid indirect harm to women. (... do this, but it is doubtful how effective their attempt is.) Maybe, an agent based model that aims to contribute to the debate around free speech should take some kind of utility or ethical normativity into account.

4 Conclusion

Let's summarize this. The question about the limits of free speech has been important and controversial for centuries and it remains hotly debated today. There are good arguments for allowing vast liberties but the debate is not settled. Agent based models are a way to explore limitations on free speech cheaply and safely, in a setting where it is clear how effective they turn out to be and in a way that allows us to focus on the variables that matter. The Hegselmann-Krause inspired model that

I explored here suggests that a deliberating group can become polarized and inaccurate when the truth is censored. Here, Mill is correct. But it also suggests that censoring falsehoods is neither harmful (contra Mill) nor helpful to our epistemic ends. Furthermore, according to the model, arbitrarily silencing a group of agents barely detracts from our epistemic ends as long as this group is not huge. Even if the group is huge the effect is small. According to this model, silencing unpopular beliefs does not make a group less accurate.

The results of these explorations should be taken with a grain of salt. They are based on simplifying assumptions that might be a bit too simplifying. Especially the assumption that all agents are equally capable observers in every respect, the pure focus on epistemic value to the exclusion of individual utility or ethical value and the abstraction from justifications appear to be doubtful.