

# MODELLING TESTIMONIAL INJUSTICE

Leon Schöppel (*12334427*)

September 2021

## I. INTRODUCTION

Testimonial injustice, together with hermeneutic injustice, are the two main forms of epistemic injustice identified by Miranda Fricker<sup>1</sup>. Kristie Dotson further differentiates two ways in which testimonial injustice is brought about, two different ways of silencing the testimony of individuals by committing epistemic violence against them. During testimonial quieting the testimony is given less uptake than appropriate, e.g. the speaker is not taken seriously, dismissed. During testimonial smothering, the speaker is coerced/ pressured by the situation to tailor their own testimony to the biases and identities of those listening. It is natural to assume that these two forms of silencing would interplay directly, in that testimonial smothering is a reaction/ adaptation of speakers who experience(d) testimonial quieting. This essay aims to simulate these two concepts and their relation/ interactions using an agent-based model. Particular focus will lie on investigating the effects that the presence of epistemic violence in a collective has on its ability to converge to truth.

In the first section, I will introduce the target phenomena in some detail, paying particular attention to the conditions that have to be met for behaviour to be classified as epistemic violence. Section two will introduce the features and inner workings of my model, and here I will spend some time arguing for why the reader should take it to plausibly represent the target phenomena. Section three is concerned with the results of running the model at specific, argued for settings. Apart from showing that the model for the most part behaves as one would expect it to, the data also points to extremely detrimental effects that the presence of testimonial quieting and smothering can have on the verisimilitude of a societal consensus. In doing so, it provides evidence (of sorts) that in addition to being morally wrong, epistemic violence should also be avoided from the perspective of social epistemology. Finally, after drawing a quick conclusion, I give an outlook as to what else could be investigated using my model, as well as suggest some modifications that might prove fruitful for further research.

---

1. Grasswick 2018

## II. THE TARGET PHENOMENA

In her Book *Epistemic injustice, Power & the ethics of knowing*<sup>2</sup>, Miranda Fricker identifies two major forms of injustice related to knowledge. This relation is in terms of the harm that they create, as they wrong a victim “qua epistemic agent”, “specifically in her capacity as a knower”<sup>3</sup>.

The first such form, **Hermeneutic injustice**, arises from “having some significant area of one’s social experiences obscured from collective understanding, owing to a structural identity prejudice in the collective hermeneutical resource”<sup>4</sup>. The maybe best known example thereof is the lack of collective understanding of the phenomenon of *sexual harassment*, which got resolved with the introduction of the term in the 1970s. Prior to this, the (mostly) women who suffered from sexual harassment additionally suffered the hermeneutic injustice of being unable to properly make sense of their experiences.<sup>5</sup>

**Testimonial injustice** on the other hand is maybe more tangible, as it occurs “when a speaker suffers a credibility deficit due to an identity prejudice on the hearer’s part”<sup>6</sup>. Examples of this form of injustice are so plentiful that I don’t believe giving one to be necessary.

### EPISTEMIC VIOLENCE

Testimonial injustice subjects victims to two different forms of epistemic violence: It either directly harms a speaker by quieting their testimony, or — possibly more indirectly — strong-arms them into tailoring that testimony. The latter is called testimonial smothering.

More generally, Kristie Dotson defines epistemic violence as “a refusal, intentional or unintentional, of an audience to communicatively reciprocate a linguistic exchange owing to pernicious ignorance”, where “pernicious ignorance [...] refer[s] to any reliable ignorance that, in a given context, harms [others].” Finally, “[r]eliable ignorance is ignorance that is consistent or follows from a predictable epistemic gap in cognitive resources.”<sup>7</sup>.

In my model, as I will explain and justify below, reliable ignorance will be a feature of social groups whose members are identifiable via phenotypes. That is, members of such social groups may have a statistical predisposition to be reliably ignorant on certain issues, and that ignorance — in certain contexts — will become harmful. Dotson herself suggests such a dynamic, stating that “the reality that members of oppressed groups can be silenced by virtue of group membership is widely recognized”<sup>8</sup>.

The mentioned refusal is expressed in the form of **testimonial quieting**, in an instance of which a

---

2. Fricker 2007

3. McKinnon 2016, p. 439

4. McKinnon 2016, p. 441

5. Grasswick 2018

6. McKinnon 2016, p. 438

7. Dotson 2011, p. 238

8. Dotson 2011, p. 236, 248

speaker's testimony "is ignored entirely. It's as if she didn't speak at all."<sup>9</sup> But the violence often goes even beyond the "audience fail[ing] to accurately identify the speaker as a knower"<sup>10</sup>, as examples like racial microaggressions<sup>11</sup> and domestic violence<sup>12</sup> show. And even though testimonial quieting is the most direct form of epistemic violence, it is not the only one, as the expected harm of being quietened can stop people from giving risky testimony in the first place.

**Testimonial smothering** is often understood to be a decision not to speak at all<sup>13</sup>. Smothering then refers to the suppression not just of the content of a speaker's testimony, but a suppression of the utterance itself. In this essay, though, I want to treat the phenomenon in a slightly different manner: I will be modelling situations that don't allow participants to simply refuse to be speakers in the first place. Instead, they are forced to give testimony, forced to satisfy the attention drawn to themselves, and therefore have to smother by adjusting the content of their speech. I agree (in the letter, though not in spirit) with McKinnon when she writes about smothered speakers that "[i]t would be wrong to say that they *exclude* themselves from knowledge production"<sup>14</sup>, because I want to say that in many situations they do in fact participate, by giving tailored testimony instead of honest one. This fits with how Dotson describes the phenomenon: "[T]estimonial smothering is the truncating of one's own testimony in order to ensure that [it] contains only content for which one's audience demonstrated testimonial competence"<sup>15</sup>.

As an example, imagine someone called Sabrina working in an office full of people with political opinions wildly differing from her own, but who all agree amongst each other. Imagine further that during a group conversation in which everyone else has already stated their agreement with a — from Sabrina's perspective — ridiculous claim, someone asks her for her opinion. In this situation, now that Sabrina is suddenly in the centre of attention, I want to claim, smothering her testimony requires Sabrina to actually give testimony. Ducking away and finding an excuse would not help her get out of this clutch, but agreeing with the group consensus will. To avoid the harm that accompanies testimonial quieting, Sabrina would be forced to tailor her testimony to the political opinions of her colleagues.

In this essay, I will follow Glenn C. Loury in calling non-tailored testimony *naive*. He makes the case that in the real world, "[n]aive communication — where a speaker states literally all that he thinks [...] is rare and foolish in" "matters of common concern with the intent to shape public opinion or affect policy outcome"<sup>16</sup>. Yet, given that my model will be simplified to the point where it features but a single phenomenon that gives agents a reason to tailor their testimony — namely epistemic quieting — the situation is more straight forward: Every instance of giving testimony is either smothered, or it is naive.

---

9. McKinnon 2016, p. 442

10. Dotson 2011, p. 243

11. Dotson 2011, p. 246

12. Vince 2018

13. McKinnon 2016, p. 443

14. McKinnon 2016, p. 443

15. Dotson 2011, p. 249

16. Loury 1994, p. 431

Three conditions generally need to be met for an act of tailoring one's testimony to count as testimonial smothering<sup>17</sup>:

- (i) The content of the naive testimony would have had to be unsafe and risky.
- (ii) The audience must have demonstrated testimonial incompetence with respect to the content of the naive testimony.
- (iii) This incompetence must (appear) to result from pernicious ignorance.

In my model, the demonstration required by (ii) will be based on observable membership in social groups, meaning that if members of such a group *G* are - on average - perniciously ignorant pertaining to the content of the naive testimony, the (smothered) speaker can identify an audience of majority *G* listeners to be testimonially incompetent. More details on that in the section *The Model*.

#### STANDPOINT THEORY

When ascribing certain features, be it a high likelihood of certain biases, or certain tendencies to quieten others to members of certain social groups identifiable via phenotypes, I am not saying that these properties are essentially tied to displaying those features (and neither am I forced to<sup>18</sup>). Rather, I ascribe to the view that “norms and conventions [...] piggyback on seemingly irrelevant factors like race and gender”<sup>19</sup>. However such roles and conventions came about — O'Connor provides inspiring analysis of how little must be assumed to simulate their evolution — I take it to be the case that contingently, observable identity features do sort humans into social groups that come with certain experiences and learned behavioural patterns.

Going from there, one should expect questions of social relations and interactions to depend on social location of agents<sup>20</sup>. With this essay, as well as the literature on epistemic violence putting a special focus on majority-minority-relations, I want to “ascrib[e] contingent epistemic privilege to (some) subdominant standpoints”<sup>21</sup>. In this, I am building on the established notion that “some standpoints [...] have the especially salient advantage that they put the critically conscious knower in a position to grasp the effects of power relations on their own understanding and that of others.”<sup>22</sup>, claiming that “the perspectives of subordinated social groups have an epistemic advantage regarding politically contested topics related to their subordination, relative to the perspectives of the groups that dominate them.”<sup>23</sup>. Take as an example of these two points the idea that “feminist arguments that gender is an epistemically

---

17. Dotson 2011, p. 244

18. Wylie, Figueroa, and Harding 2003, p. 4

19. O'Connor 2019, p. 4

20. Grasswick 2018

21. Wylie, Figueroa, and Harding 2003, page 21

22. Wylie, Figueroa, and Harding 2003, p. 13 and 14

23. Anderson 2020

relevant category of social location apply (only) as long as the society under consideration is structured along the lines of gender.”<sup>24</sup>. To what extent situatedness applies to any specific choice of social groups and issue is something “that cannot be settled in the abstract, in advance” but instead “require[s] the second order application of our best research tools to the business of knowledge production”<sup>25</sup>.

Still, for group identity to serve as the basis of simulating epistemic quieting and smothering as defined above, it requires more than just any identity-dependent bias and behaviour pattern. It requires instances of reliable, harmful ignorance, and identity based reliable, harmful ignorance at that. Yet, given how important “their location in a hierarchically structured system of power relations”<sup>26</sup> is for the epistemic standpoint of human beings, and given how these systems are often organized and structured around phenotypes of humans, it is not a far reach to say that such ignorance would prove both reliable and identity based. If social struggles have shown anything, then that systems like the patriarchy, classism and institutional racism are as hard to weed out as they are harmful, exactly because they reliably protect themselves, and in doing so reproduce their respectively relevant social identities.

As an example from outside feminist philosophy, Du Bois’ concept of *double consciousness* introduced in his 1903 book *The Souls of Black Folk*<sup>27</sup> ascribes to black people living under racial oppression in the US a certain dual understanding of society, in that their oppression requires and allows them to be actively aware — in addition to their own perspective — of the perspective of the white society. Clearly, here the situatedness of a marginalized group affords them a special epistemic standpoint, a special ability to know and understand social relations. And just as clearly, this epistemic privilege is born out of the (at the time) extremely established and reliable — as well as extremely harmful — practice of racial segregation. Another example of applications of the concepts of testimonial smothering and quieting along the lines of phenotype-based social groups: Rosa Vince<sup>28</sup> argues — based on various empirical studies — that (mostly) women regularly suffer specifically epistemic violence in the form of testimonial quieting and smothering when trying to refuse sex to their (mostly male) partners.

But even when members of certain identity groups tend to be clearly and reliably ignorant, that might not yet be harmful (over and above the inherent harm caused by ignorance<sup>29</sup>). And indeed, my model will not presuppose this. Instead, it makes the harm caused by reliable ignorance context-dependent: Only if the power dynamics in a specific so-called communication game are such that the victim appears as a misfit or easy target so to speak, will agents feel emboldened enough to become epistemically violent by quieting the victim. Consequently, only in games in which agents must assume to be misfits will they (rightfully) expect to face violence and attempt to avoid it by tailoring their testimony.

---

24. Grasswick 2018, my brackets

25. Wylie, Figueroa, and Harding 2003, pages 22, 23

26. Wylie, Figueroa, and Harding 2003, p. 9

27. Du Bois 2015

28. Vince 2018

29. Dotson 2011

### III. THE MODEL

I've implemented my model<sup>30</sup> using *NetLogo 6.1.1*<sup>31</sup> and its *arrays-extension* (arrays are mutable and therefore allow for faster computation than NetLogo's native lists). It is agent-based, and so within it "social systems are modelled as a set of agents located within a suitable space, a set of behavioural rules for the agents and, lastly, a mechanism that guides the agents' interactions, including their actions' feedback on the structure"<sup>32</sup>. In this case, with the target phenomenon being the interplay between testimonial quieting and smothering, I will simulate the communication and epistemic violence between agents of a minority and a majority group. As is common in modelling, I've made multiple simplifying assumptions about the phenomenon, as well as about the properties of agents, but I will spend some time in this section providing reasons both for these simplifications and for the details of my implementations. One implicit, yet particularly central, assumption is that I take the cited literature to adequately pick out real phenomena. In this light, I hope that my model can provide a plausible account of how these two forms of epistemic violence might interplay, and thereby provide what one could call how-plausibly understanding of those phenomena.

#### THE CENTRAL PROPOSITION

Central to every simulation in this model is a single proposition  $\phi$  with an objective probability value  $p(\phi)$  in the interval  $[0, 1]$ <sup>33</sup>. The value of  $\phi$  will either be randomly assigned each run, or — based on a toggle in the user interface (UI) — be set to  $1/2$  for easier comparison between runs.  $\phi$  is meant to be in an area within which epistemic (dis-)privilege based on group identity is to be expected. By assigning an objective value to  $\phi$ , I am avoiding general relativism about truth, and as explained above, I am not committing to a general standpoint-epistemic position. Rather, I only need to defend the somewhat trivial point that there are indeed specific issues or questions, for knowledge about which identity matters. Then, my model simulates how effects generated by many individual instances of testimonial quieting first lead to tailored testimony, and as a result to problematic effects on the verisimilitude of the collectives mean credences.

#### THE WORLD

The model features a world of adjustable size, split up into patches coloured in a checkerboard-pattern. These are used to sort agents into communication games, see subsection *The Game*. The world does not wrap at its edges, meaning agents can freely wander, e.g., from the topmost row of patches directly to

---

30. Schöpl 2021

31. Wilensky 1999

32. Klein, Marx, and Fischbach 2018, p. 11

33. In this regard it is similar to Douven 2010 and Christensen 2007

the bottom of the world.

## AGENTS

Agents are split up into at least two, possibly three, groups. These are meant to represent identity groups that both come with epistemic (dis-)advantages in certain areas, and come with certain markers/phenotypes, allowing agents to discern the group identity of other agents.

Agents have individual credences  $cr(\varphi)$  in the central proposition. Given that I take  $\varphi$  to be harder to epistemically access for people of a certain identity group, agents have individual biases that are controlled via sliders adjusting the average bias of their group. Nevertheless, if toggled on in the UI, it is possible for individual agents to be outliers and defy their group identity. An agent's prior  $cr(\varphi)$  is calculated by adding their bias  $b$  to  $p(\varphi)$ . In addition, some random noise  $\in [-0.2, 0.2]$  is added.  $b$  is inferred from the mean bias of the agent's group  $b_{group}$  either by  $b = random([0, 2])b_{group}$  or, in case of allowing for outliers,  $b = random([0, 2.2])b_{group} - \frac{b_{group}}{10}$ . In both cases, the mean bias of any group will fluctuate around the setting, though if overlap is toggled on, the spread is minimally wider.

Agents wander around aimlessly in the world, in order to repeatedly find themselves on patches with a ever new random collection of other agents. This will become relevant in the next subsection, as the position of agents decides about their participation in communication games.

The two forms of epistemic violence central to the aims of this paper come with utility values for those being subjected to them. In this model, the utility of being smothered depends on the victim, in an effort to reflect how differences in character influence how harshly people take having to tailor their testimony. The utility of being quietened depends on the perpetrator / aggressor, to reflect how different people use methods of varying severity when behaving in epistemically violent ways. In addition, based on group identity, agents will have a tendency to be epistemically violent, which will be explained in the subsection *Testimonial Quieting*.

Based on whether the simulation employs the learning function, agents either have direct access to certain mean values in the model, i.e., average credences and quieting tendencies of groups and the mean utility of being quietened, or they will have to slowly collect experience and data to calculate (at first rough, later more accurate) estimates of these values. In the later case, they will remember various details about previous games.

## THE GAME

Once per *tick*, NetLogo's discrete units of measuring time progression, agents are sorted into communication games based on their position in the world. These games consist of three consecutive rounds, and are played on every patch containing two or more agents. I take these games — the participation

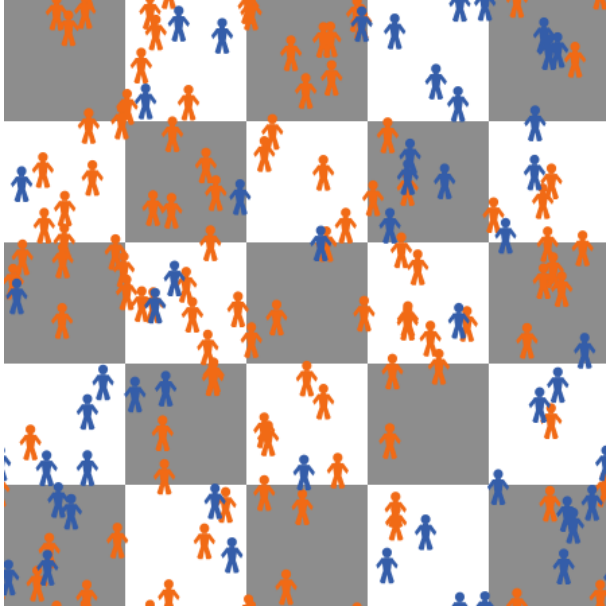


Figure 1: A typical setup with a 5 by 5 patches world containing two groups of 150 total agents,  $\frac{2}{3}$  of which are members of group A.

in which agents do not get to choose — to represent real situations that too are often unavoidable in practice, e.g., work meetings, court hearings. If one switches off epistemic violence in the model, then agents will skip the game, merely honestly stating their credences and updating fairly on the testimony given by others.

In **round one**, every agent  $A$  will identify the group identities of all the other participants in their respective game. Based on either the objective mean credences of the different groups, or — if the simulation employs the learning function —  $A$ 's experience of said credences,  $A$  will calculate the expected patch consensus. I have implemented two different versions thereof, one where agents include their own credence in the calculation, and one where they omit it. I take the latter to be more realistic, as here agents make a prediction of what the other participants' consensus would be, without  $A$  present.

Afterwards, during **round two**, every agent  $A$  goes on to calculate the expected utility of giving *naive*, honest testimony. To do so, they try to assess whether they will be quietened, for every other participant present. This assessment is based on  $A$ 's information about each group's quieting thresholds and mean credences.  $A$  then multiplies the utility value of being quietened by the number of suspected offenders, and compares that to  $A$ 's own utility of having their testimony smothered. The values used in these calculations are yet again either the objective means, or collected over  $A$ 's experiences from past games. Based on the resulting expected utility,  $A$  will either *be smothered*, or give honest testimony, that is, state their actual credence.

Finally, in **round three**,  $A$  collects their input from the testimonies given by the other participants. For each testimony given,  $A$  decides, based on conditions explained below, whether they should *quieten*



the speaker or give them appropriate uptake. After collecting the input,  $A$  will weigh it according to the settings chosen in the UI and update their own credence accordingly. If the learning function is employed, agents use this round to update their beliefs about the credences of other agents (based on group identity), as well as about tendencies to and utility of quieting.

The consecutive nature of these rounds results in the impossibility for agents to wait with certain choices or actions until they know exactly how everyone else behaves. I take this to be an upshot of the model, given that people are constantly forced to act on uncertain information. Still, this means that agents *signify ignorance* merely by virtue of their belonging to certain groups, as e.g., they will not be able to give immediate feedback to testimony given by the first speaker.

### TESTIMONIAL SMOTHERING

If an agent  $A$  comes to the conclusion that being smothered is — given the specific situation and context — less bad than giving honest testimony and being quietened in return, they can instead tailor their testimony. I have implemented three versions of this:

1.  $A$  seeks to withdraw from the game entirely, giving  $N/A$  testimony, or rather no testimony at all. At first glance, this might strike the reader as the most appropriate implementation, given what I've cited in the previous section. Given that the communication game is meant to represent situations which agents are forced to participate in, though, I find a simple refusal to participate to not be satisfactory.
2. Instead, I will mainly focus on a version of smothering in which  $A$  will *utter the expected patch consensus*, to fly under the radar of the other participants. Here,  $A$  tailors their testimony to exactly what they believe their audience wishes to hear, as calculated in round one of the game.
3. As a last option,  $A$  could also choose to split the difference to the expected patch consensus, merely softening their testimony. I have included this version because it seems realistic that in many cases, victims of epistemic violence might see a chance to get some of their points across, even if they are forced to adjust what they actually want to state. Still, as this type of smothering doesn't entirely smother the victim, I will disregard it in the *Results* section.

### TESTIMONIAL QUIETING

When receiving testimony from a speaker  $B$ , an agent  $A$  will assess whether or not to quieten  $B$ . I want to stay agnostic about the degree to which this assessment constitutes a conscious deliberation in actual humans. In this model, though, the agent's deliberation includes the following conditions:

- $A$  is not a member of the identity group that  $B$  belongs to.

- *A* is backed up by the other present participants, in that their own credence is closer to the actual patch consensus than the testimony given by *B*.
- *B* gave testimony that is further from *A*'s credence than *A*'s quieting threshold allows for.
- Optional 4th condition: *B* gave testimony that is further from the actual patch consensus than *A*'s quieting threshold allows for.

If *A* chooses to quieten *B*, they will not offer appropriate uptake to their testimony, in one of the following ways:

1. *A* slots in their own credence for *B*'s testimony when calculating the input from the current game. This somewhat strong and possibly unrealistic version of quieting might represent cases in which dissenting opinions are received as so out-there as to be taken as evidence for one's own opinion.
2. In this softer version of the same general idea, *A* might split the difference with *B*'s testimony and slot the result into the input.
3. The final version, which I will use for the results shown in this paper, *A* will entirely ignore *B*'s testimony, calculating the input as if it was never given, as if *B* wasn't even present. This — in my opinion — most closely resembles the phenomenon of quieting as discussed in the literature.

If the learning function is employed, based on their experience with *A*, *B* will try to assess the average quieting threshold of members of *A*'s identity group. In case of quieting, *B* may also update their belief about the average utility of being quietened, and use that for future calculations of expected utility of giving honest testimony.

## EXPERIMENTS

If toggled on in the UI, agents will take time off between every game to perform experiments concerning the central proposition  $\varphi$ . These experiments are meant to represent epistemic progress going beyond the results of mere communication, but may not entirely surpass the identity based biases that agents bring with them. Agents are not meant to have direct access to  $p(\varphi)$ , but instead are merely (indirectly) attracted to the truth.<sup>34</sup> In the base version, an agent's credence in  $\varphi$  at tick  $t + 1$  is calculated based on a weight  $w$  as such:  $cr_{t+1}(\varphi) = cr_t(\varphi)w + p(\varphi)(1 - w)$ .<sup>35</sup>

Yet as this updating rule doesn't reflect an agent's individual bias, I have added two versions of biased experiments, in which agents obtain slightly misleading results from their experiments, and are pulled towards a direction indicated by their personal bias. The first of these options adjusts  $p(\varphi)$  in the above

---

34. Hegselmann and Krause 2006, p. 3

35. Douven 2010, p. 149

formula by adding random noise  $r \in [0, 1]$  multiplied by the agent’s bias  $b$ , thereby creating a persistent negative epistemic impact:  $cr_{t+1}(\varphi) = cr_t(\varphi)w + (p(\varphi) + rb)(1 - w)$ . The second version multiplies the agent’s bias by the difference  $d = |cr(\varphi) - p(\varphi)|$  and adjusts  $p(\varphi)$  by the resulting amount, so the experiments become less biased as the agent’s credence approaches the truth:  $cr_t(\varphi)w + (p(\varphi) + db)(1 - w)$ .

#### LEARNING FUNCTION

As mentioned throughout this section, the model features a toggle for *employ learning function*. If turned off, agents can directly access data about the mean credences and thresholds of groups, as well as know the mean utility value of being quietened. This is somewhat unrealistic, especially in scenarios where one group constantly tailors their testimony when interacting with agents from another group. Therefore, I will pay particular attention to results obtained from runs in which agents must slowly gather data about the credences of each group, about the utility of being quietened, as well as try to assess how likely it is to be quietened in the first place.

Most of these values are calculated in a trivial fashion: E.g., an agent  $A$  updates their belief  $c$  about the mean credence of group  $B$  after receiving testimony  $t$  from a particular member of that group, by the following formula:  $c_{new} = \frac{cx+t}{x+1}$ , where  $x$  is the total number of previous encounters between  $A$  and agents of group  $B$ . As an optional setting in the UI, one can choose to — in cases where  $A$  quietens  $B$  — let  $A$  update not on the testimony given by the victim, but instead slot  $A$ ’s own credence in for  $t$ .

I have implemented two versions of calculating the average quieting thresholds of groups, but only *adjust expectations* will be used in the results of this paper. Here, during round 3, based on the *difference to the testimony* given by a (potential) aggressor  $A$  in the previous round, a (potential) victim  $B$  will try to predict, whether they would be quietened. Then, if they predicted wrongly,  $B$  will adjust their belief about the average quieting thresholds of  $A$ ’s group at the start of the next game. If, for example,  $B$  predicted the threshold of  $A$ ’s group to be too low in 3 cases (i.e., unexpected non-quieting), correctly in 2 and too high in 1 case (unexpected quieting),  $B$  will adjust the threshold upwards twice (to expect to be quietened more easily in the future).

#### IV. RESULTS

In addition to manually exploring my model, I have run a total of 1100 simulations using *BehaviorSpace*. For each of the settings discussed below, I collected the results from 100 runs for 1000 ticks each. As the results from these simulations support me in thinking that my model behaves quite deterministically, they should allow me to make more general claims about what my model shows, than individual runs would. Yet for the sake of clarity, I still chose to include figures from individual, representative runs.

Certain features, like total number of agents, absolute quieting thresholds and world dimension I identified to be less significant. While for reasons of robustness they remain as options in the model, they will remain fixed in all the runs considered in this section. As described above, I considered additional interpretations of certain phenomena and implemented them for the reader to play around with, but they will not be featured in the results section.

In every run that I will be discussing in this section, the following settings were selected:

- The collective was made up of 150 agents, 100 of which belong to group *A*, 50 to group *B*. Group *C* was not included in the simulations, for the sake of clarity.
- The world consisted of 5 by 5 patches.
- $p(\varphi)$  was set to  $1/2$ .
- Input from the communication games was weighted 0.1, while input from the experiments was weighted 0.01.
- Both groups had their mean quieting thresholds set to 0.1, so agents were — on average — quite willing to quieten others.
- Agents of group *A* were set to a mean bias of  $-0.2$ , while agents of group *B* were not biased themselves. *Allow-outliers* was toggled on, meaning there will be some rare agents defying their group's biases.
- If injustice was toggled on and for the reasons stated above, I chose as patch-consensus-type *omit own credence*, as smothering-type *utter expected patch consensus* and as quieting type *ignore fully*. I went without the 4th condition for quieting, though not much appeared to hinge on that during my test runs. As mean utility values for smothering and quieting I chose 3 and 2 respectively, so that — on average — it is better to be quietened by one aggressor than to tailor one's testimony, but better to succumb to smothering than to be quietened by two (or more) aggressors.
- If the learning function is employed, I chose the most effective and realistic settings:

Agents update their belief about others' credences based on the actual testimony given. Whenever an agent *A* quietens another *B*, I expect *A* to notice what it is *B* said, despite the fact *A* doesn't offer proper uptake to the testimony when it comes to updating *A*'s own  $cr(\varphi)$ . Otherwise, agents like *A* would be delusional as to the openly uttered testimony of others like *B*.

And, as agents aren't mind-readers, they have to update their beliefs about other's credences based on the testimony of those they encounter, instead of the actual credence of said agents.

For the learning-function specifically, I had to decide on prior values to initialize the agents with, as choosing the wrong values here could have detrimental effects on the results. If, for example, every agent of group *B* started out believing the mean credence of *A* to be a value that's much higher than it actually is, this will — at the start of any run — impact their estimations of the patch consensus. I have therefore chosen *custom* initialization values as follows: For the expected utility value of being quietened, agents start out naive, meaning that they don't expect any harm to come from this form of epistemic violence until they experience it for the first time. When it comes to agents' prior beliefs about the mean quieting thresholds of other groups, I was unsure how to progress: Would the naive assumption be to believe the threshold to be so large as to never be met? I decided to initialize this value with a probability value  $\in [0, 1]$  that is randomly chosen for each individual agent. For the initial credences about other group's mean credences, the *custom* settings include randomization as well, though for the sake of comparability between the runs featured in this section, I have temporarily replaced it by the fixed value of  $1/2$ .

## MEASUREMENTS

The model includes plots, which allow the user to observe, over time, how values central to the simulation develop. The lines are coloured the same way that the agents of the respective groups are, that is *orange* for members of group *A*, *blue* for members of group *B*, and if included, *brown* for members of group *C*. The following data is plotted:

- For each group, the mean credences held and testimonies given by its members. They are contrasted (in black) by the objective chance of  $\varphi$ , which in the results shown in this section is always  $1/2$ .
- The overall mean distance from the truth (mean error), that is how far on average every agent's credence in  $\varphi$  is from the objective chance of  $\varphi$ . For a simulation containing  $n$  agents, this amounts to  $\frac{1}{n} \sum_{i=1}^n |cr(\varphi) - p(\varphi)|$ .
- For each group (and a total in black), how often its members have so far been quietened and smothered. Note that these are absolute values, not adjusted for group proportions.
- For the learning function specifically, I've added a handful of plots allowing the user to observe how well agents of each group — on average — estimate the utility value of being quietened, as well as the mean credences and quieting thresholds of each group.

For each setting, I have included results from runs that were just long enough to make the relevant effects most clearly visible. Hence, keep in mind that the number of ticks differs wildly from figure to figure.

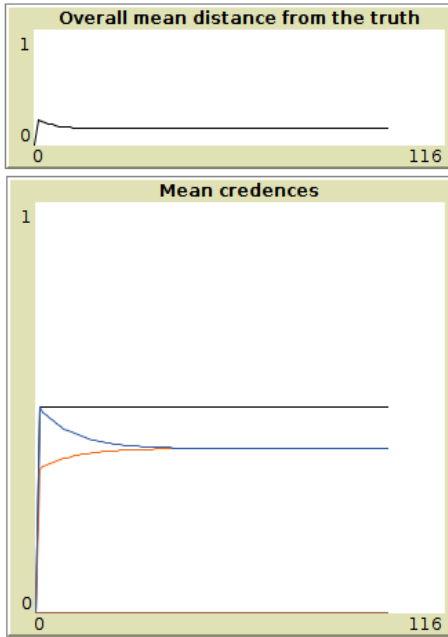


Figure 2: Neither injustice, nor experiments, 100 ticks;

#### NO EPISTEMIC VIOLENCE

Just as one would expect, if epistemic violence and experiments are toggled off and agents *skip the games* instead of playing them, a collective consensus will ensue relatively quickly (see figure 2). Where this consensus lands depends on the bias-settings and sizes of the different groups, as all that occurs is that the agents influence each other to arrive at a certain credence shared by every agent in the model. In these runs, not only does the mean error not converge to 0, it barely changes at all, as the community lacks epistemic access to  $p(\varphi)$  beyond what was encoded in their initial priors.

If in such simulations, bias-free experiments are toggled on (see figure 3), the mean group credences quickly coalesce and converge to  $p(\varphi)$ .

If persistent bias is introduced to these experiments (see figure 4), convergence to  $p(\varphi)$  is hindered by the biased results of the majority group, which negatively impacts the epistemic endeavour of the collective. While I do find this result to realistically reflect certain epistemic features within social systems, as Dotson writes: “the state of reliable ignorance insures that an epistemic agent will consistently fail to track certain truths”<sup>36</sup>, it is also extremely pessimistic. As in this model, agents cannot get rid of their biases, no matter how often they experiment and communicate, they will never arrive at the truth. Thus, the no-injustice, perpetual-bias case sets an uninteresting baseline for the further results in this paper, and I will — for now — skip over further considerations of experiments of this type.

Softening the bias type to *resolving* merely increases the amount of ticks it takes the collective to converge to  $p(\varphi)$  (see figure 5). This result is to be expected, as the negative influence that agents’ bi-

36. Dotson 2011

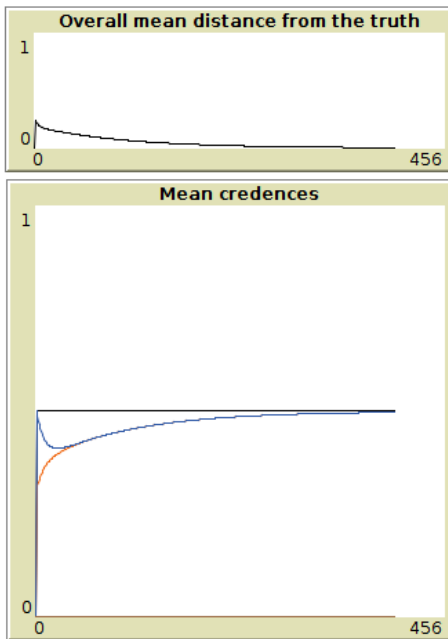


Figure 3: No injustice, bias-free experiments, 400 ticks;

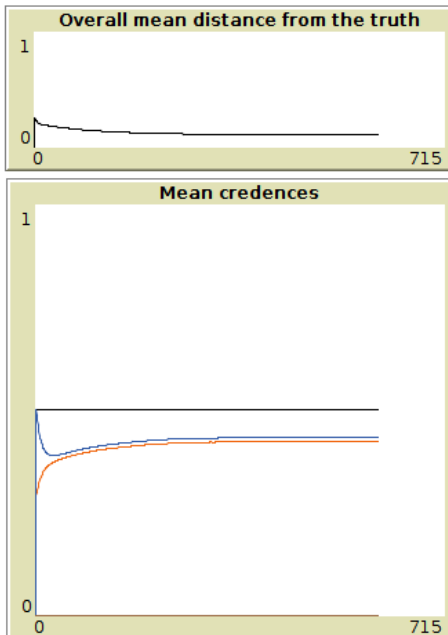


Figure 4: No injustice, perpetual-bias-experiments, 600 ticks;

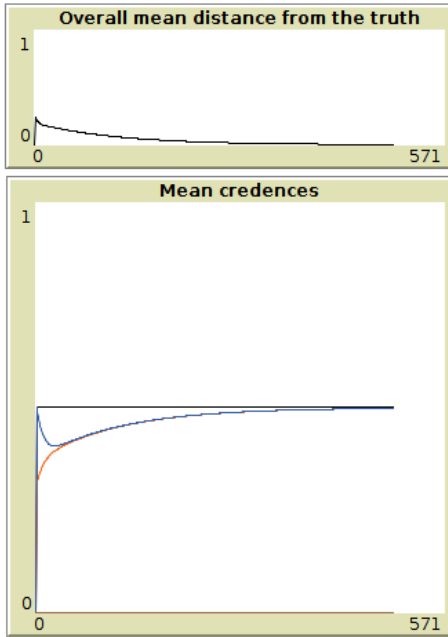


Figure 5: No injustice, resolving-bias-experiments, 500 ticks;

ases have on the outcome of their experiments (or rather, their interpretations of their results) decreases as they communicate amongst each other. Igor Douven<sup>37</sup> takes the amount of ticks required for convergence to be a meaningful measurement of a collective’s adeptness at communicating and investigating the truth. And while I have some worries about how exactly we should interpret *a tick* with respect to real social situations, I tend to agree that in this situation, less is more: The faster the collective arrives at  $p(\varphi)$  (or within a small enough margin of, let’s say 0.01), the better.

These simulations in the absence of epistemic violence will serve as base cases for later results to be compared against. Given that neither quieting nor smothering occurs, the results of these runs do not in any way depend on the utility values of these practices. Additionally, as agents don’t require any insight about the other agents over and above their stated testimony, whether or not the learning function is toggled on doesn’t matter either.

#### EPISTEMIC VIOLENCE, NO LEARNING

Let me now present what happens when epistemic violence is featured in runs, while the agent’s behaviour is guided not by their previous experiences, but by objective values within the simulation.

The first setup to consider is one without experiments (see figure 6): Here, both quieting and smothering happen quite a lot, especially during the first 30 or so ticks, when the two groups’ mean credences have not yet coalesced. On average, it appears that any individual agent of group *B* is just under 2 times as likely to be smothered than is any agent of group *A*, whereas for quieting, it is just over 2 times as

37. Douven 2010



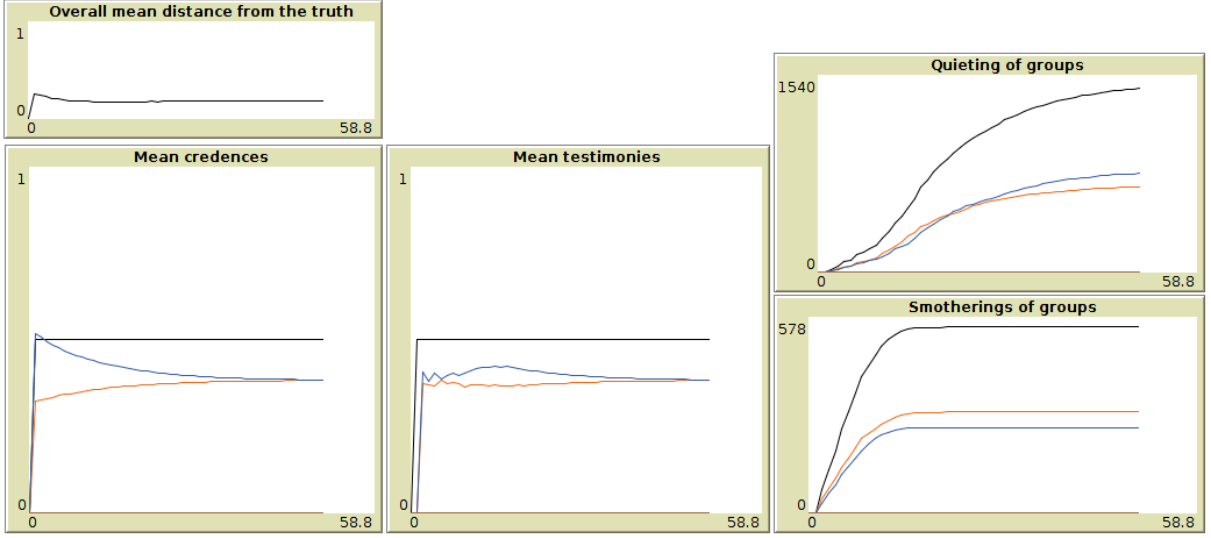


Figure 6: No experiments, 50 ticks;

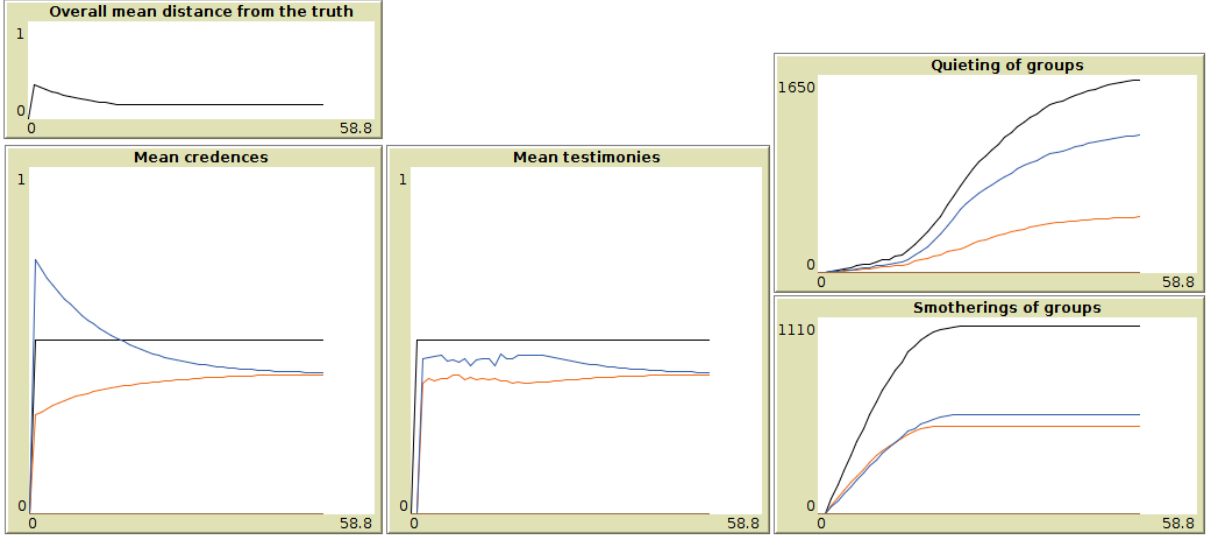


Figure 7: No experiments, biased minority group, 50 ticks;

likely. This for the most part adequately reflects the target phenomenon, with one caveat: That the majority agents are impacted by injustice to even this degree might be due to the fact that their credences are less monolithic, given the spread resulting from multiplication with an agent's group-bias is non-existent for group  $B$ . In a simulation where the mean bias of group  $B$  is 0.2, the effect disappears (see figure 7).

In any case, without experiments it takes the collective slightly longer to come to an agreement, and the biases of the majority are slightly more present in the consensus once it is established. This is because the testimonies of the minority are — relative to group sizes — both tailored and not offered proper uptake more frequently.

If experiments with resolving bias are re-introduced (see figure 8), the mean error again starts to converge to  $p(\phi)$  and does so fairly quickly. This is partly due to the fact that once the mean credences

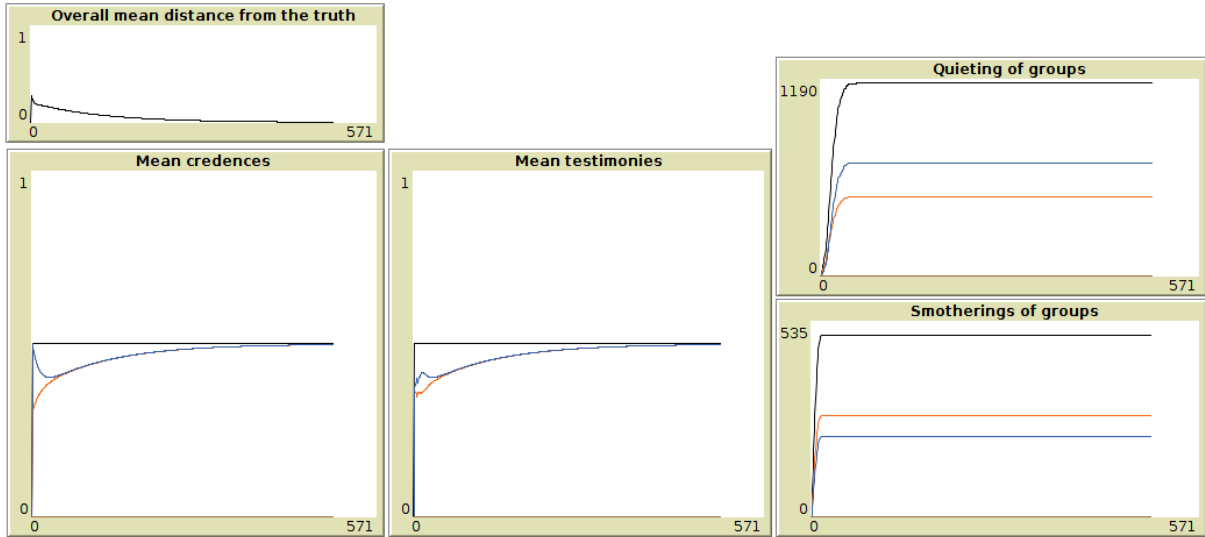


Figure 8: Resolving-bias-experiments, 500 ticks;

of the two groups approach each other, the basis for epistemic violence disappears, as no agent will even have credences that might be quietened by others. This could be regarded as a weakness of my model, especially past the first few dozen ticks. I will mention some possible additions in the *Outlook* sections, which could help break up the perhaps unnatural dichotomy between heavily biased and differing priors on the one hand, and the unified credences at the end of every simulation on the other.

Overall, the simulations with epistemic violence that is not based on the experience of individual agents, have not provided surprising results. As expected, minority agents were affected to a higher degree, and the collective benefitted less from their unbiased mean credence.

#### LEARNING FUNCTION

Before I will turn to general results obtained with the learning function turned on, let me quickly discuss the results obtained specifically about the data collection done by agents (see figure 9):

- (i) Agents that are members of minority identity groups tend to be much quicker at realizing the stakes of being quietened, due to the fact that they experience it much more frequently on average.
- (ii) While agents tend to develop fairly accurate understandings of each group's mean credences quite quickly, this is negatively affected by widespread smothering. Trivially, if agents of a group always, e.g., give testimony much lower than their actual credences, their mean credence will generally be underestimated, even by the members of this very group.
- (iii) When it comes to assessing the mean quieting thresholds of other groups, problems arise in runs in which all opinions quickly coalesce into a unified one. As this puts a stop to most instances

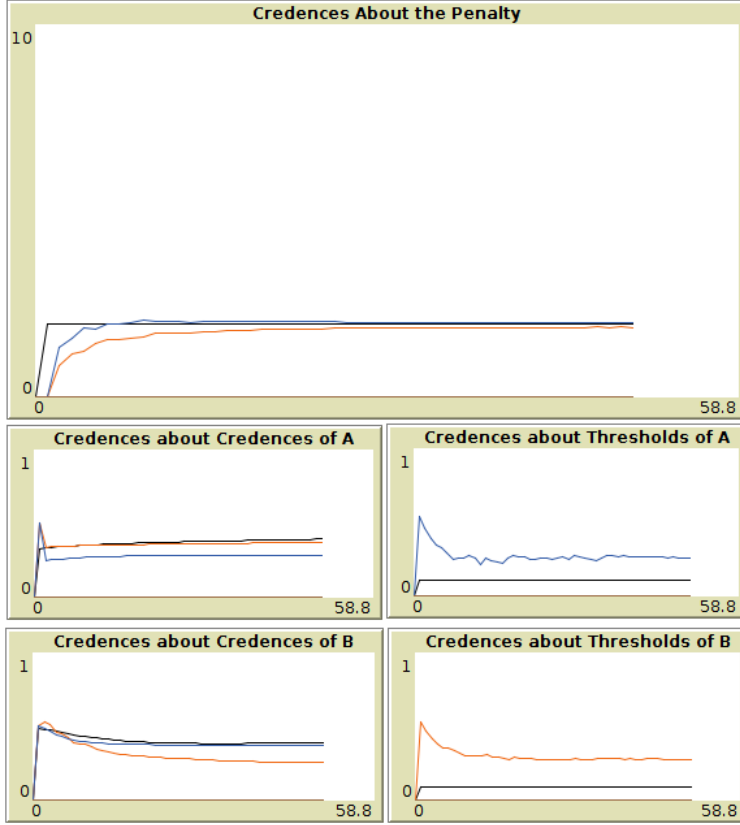


Figure 9: Learning results, resolving-bias-experiments, 50 ticks;

of quieting, agents will not have the time required to properly adjust their beliefs about the mean quieting thresholds.

I take these three observations to support my model, as they reflect what I would expect of the target phenomena. Still, to resolve any potential problems resulting from observation (iii), I have added a modified version of the learning function in which agents have direct access to the quieting tendencies of each group.

#### FINAL RESULTS

Now, let me turn to the results obtained by employing the learning function in the above-described setup. Here, we notice a new phenomenon (see figure 10): *Overcompensating minority agents*. Despite starting out having a mean credence that is closer to  $p(\varphi)$  than that of the majority group, the testimony given by the average agent of group  $B$  is consistently further from the truth than that of  $A$ . Due to having to tailor their testimony to what they believe are the credences of members of  $A$ ,  $B$ -agents end up pulling the overall consensus away from what they themselves actually believe.

As a result, some runs cease to converge to  $p(\varphi)$  after a while, but instead plateau at a mean error of up to roughly  $1/2$  of where the simulation started. This is a deeply worrying result, as now the difference

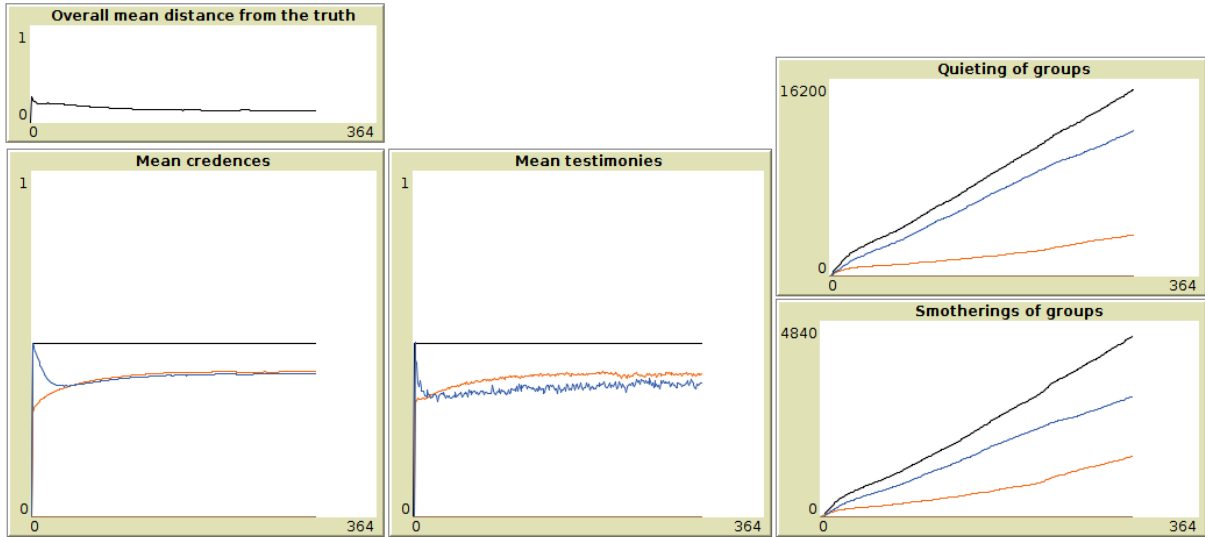


Figure 10: Injustice, normal learning function, 300 ticks: Overcompensation by minority group may stop convergence to  $p(\phi)$ .

that the addition of epistemic violence makes for the overall convergence to truth is no longer merely one of delayed success. It doesn't just take the collective longer to figure out  $p(\phi)$ , they fail to do so entirely. Keep in mind that this obtains despite the fact that agents perform experiments with resolving bias every round, which in no-injustice runs had them converge to  $p(\phi)$  quickly and consistently.

Interestingly enough, if the learning function is employed with the new addition of giving agents direct access to the quieting tendencies of each group, these problematic effects are further strengthened: At first, minority agents give testimony not only further from the truth than their actual credences, but worse than even those of the majority. Now, the mean error even increases over time, as the minority agents overcompensate so heavily as to pull the majority deeper into biased territory. Not only does the collective no longer converge to  $p(\phi)$ , but they slowly worsen over time (see figure 11).

After a while though, once the two group's mean credences start to coalesce, it appears that minority agents may become convinced of the erroneous position they have been forced to take, and the original setup is swapped — though heavily pulled towards the prior biases of the majority. This result makes me unsure as to how significant the results of my model are beyond the first hundred ticks, especially given how quickly the two groups converge to each other. Is it really realistic that a minority group would end up not only over-compensating in their tailored testimony, but actively become convinced by the majority? Wondering about this issue, I have chosen to take another look at letting agents be perpetually biased. Here, while the overall convergence to truth is just as hindered, the questionable effect can be avoided, as seen in figures 12 (modified learning function) and 13 (normal learning function). I will mention some possible adjustments to the model that would make this bias type more interesting in the *Outlook* section.

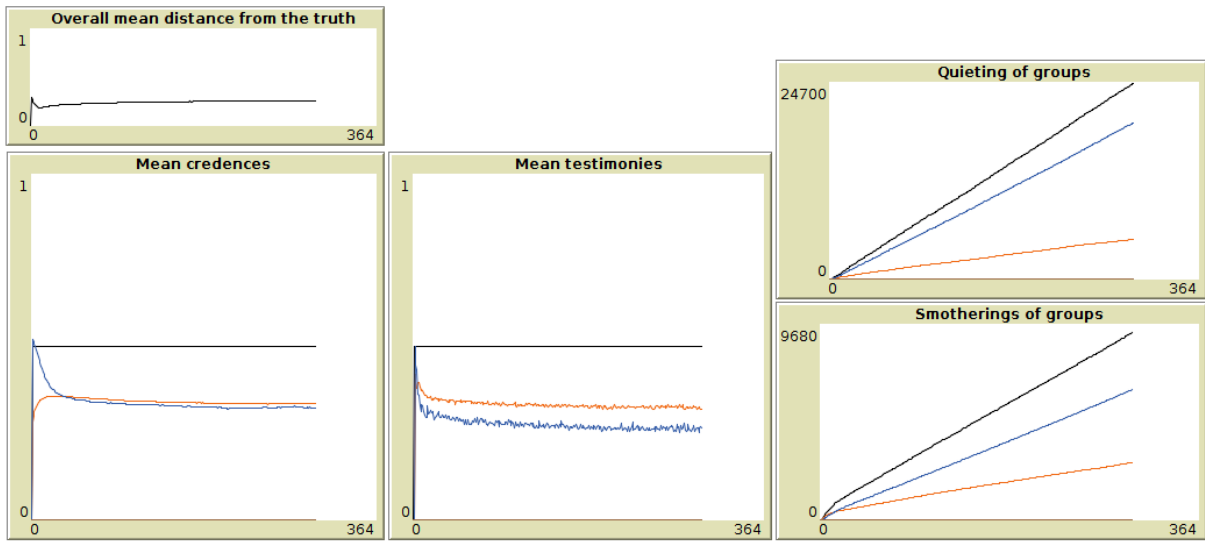


Figure 11: Injustice, modified learning function, 300 ticks: Overcompensation by minority group increases mean error.

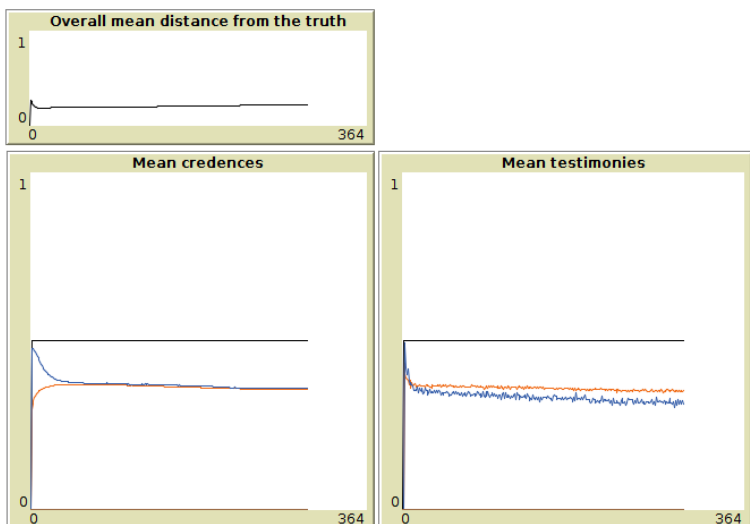


Figure 12: Injustice, modified learning function, perpetual-bias-experiments, 300 ticks;

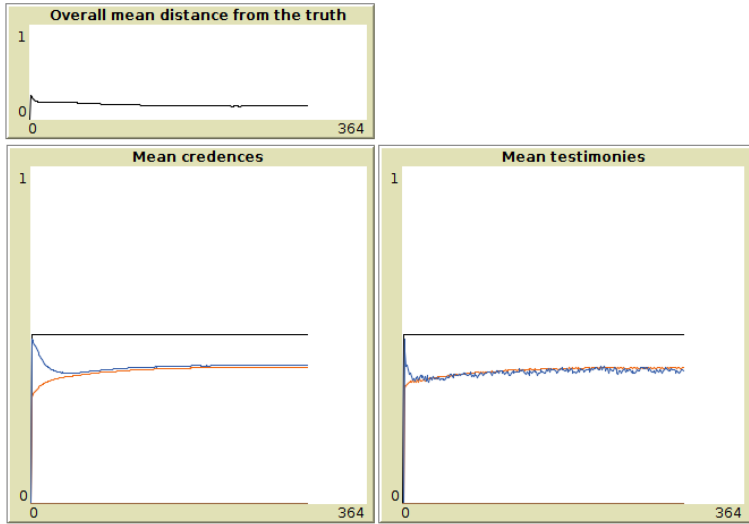


Figure 13: Injustice, normal learning function, perpetual-bias-experiments, 300 ticks;

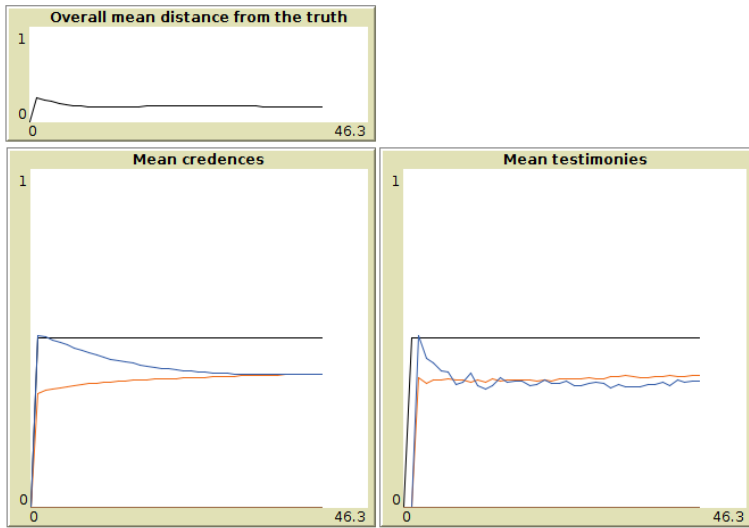


Figure 14: Injustice, normal learning function at 50 ticks: Overcompensation happens way before the minority becomes more biased than the majority.

In the face of worries about the meaningfulness of high tick-count results though, I have included another figure (14), which shows how the overcompensation starts well before any problematic effects appear. Here it is very clear how long before the mean credences of groups *A* and *B* swap places, members of *B* heavily overcompensate and wrongly tailor their testimony.

No matter the type of learning function employed, what remains consistent is that both forms of epistemic violence are by far more frequently suffered by the minority group. Only about 50% of all cases of smothering were those in which members of group *A* tailored their testimony, and only about 30% of quietings were committed against agents of group *A*. This is despite the fact that group *A* is twice as big as group *B*, meaning that per capita the difference is even more striking.

## V. CONCLUSION AND OUTLOOK

Overall, while this essay is just as much an exercise in agent based modelling as it is an investigation of the target phenomena, I hope that the results point to some worrying effects epistemic violence can have on a collective's convergence to truth. Minority agents may be pressured to overcompensate by tailoring their given testimony to the more extreme opinions of the majority group, resulting in the collective's inability to resolve the biases introduced by the standpoint of majority agents — despite otherwise useful experiments being conducted.

Insofar as the reader finds my model plausible enough to take these effects seriously, this essay should point them to even more reasons to combat epistemic violence: Not only ought we to avoid it on ethical grounds, but there are avoidable social-epistemological effects that make epistemic violence detrimental for everyone involved, even for the members of the majority group.

### POSSIBLE MODIFICATIONS OF MY MODEL

As — for reasons of brevity — I was forced to restrict my analysis to a quite narrow set of settings, much remains to be explored about this model. Things to consider include playing around in more depth with different biases, group sizes, the inclusion of a third group, different utility values, as well as different learning-function-specific prior beliefs of agents. And of course, there is the possibility that in building this model more or less from scratch, I have overlooked some features of the target phenomena that the reader finds central enough to include it in the simulations.

But even after this has been done, other avenues remain, of which I will briefly suggest two:

Firstly, the problematic epistemic effects described above may be addressable via modification of how minority agents communicate with each other: When minority agents never manage to receive honest testimony from each other, they remain ignorant of their own group's mean credences. The addition of *safe spaces* in which agents get to communicate free from fear of epistemic violence comes to mind. It strikes me as a plausible addition to the model to include communication rounds for specific groups only. Given that the assumptions I encoded in the model would make these rounds injustice-free by default, it would be interesting to see whether the following effects ensue: Do agents get more accurate ideas about their own social group? Do they also realize that they're not alone in constantly smothering their testimony? And furthermore, should agents, once (and if) they realize the discrepancy between their group's mean credence and the testimony given by the other group(s), become more hesitant to update on inputs that they suspect to be the results of epistemic violence?

Secondly, in the *Results* section I have mostly disregarded perpetual-bias-experiments for the present paper. Still, in further research it may be very interesting to consider models in which agents can update their bias over time, and/ or new agents are added between rounds, while old ones are phased out over

time. In the resulting simulations, perpetual biases types may well provide valuable insights into the target phenomena, for the reasons outlined previously.

## REFERENCES

- Anderson, Elizabeth. 2020. "Feminist Epistemology and Philosophy of Science." In *The Stanford Encyclopedia of Philosophy*, Spring 2020, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.
- Christensen, David. 2007. "Epistemology of disagreement: The good news." *The Philosophical Review* 116 (2): 187–217.
- Dotson, Kristie. 2011. "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia* 26 (2): 236–257. <https://doi.org/10.1111/j.1527-2001.2011.01177.x>.
- Douven, Igor. 2010. "Simulating peer disagreements." *Studies in history and philosophy of science part A* 41 (2): 148–157.
- Du Bois, William Edward Burghardt. 2015. *The souls of black folk*. Yale University Press.
- Fricker, Miranda. 2007. *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
- Grasswick, Heidi. 2018. "Feminist Social Epistemology." In *The Stanford Encyclopedia of Philosophy*, Fall 2018, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2018/entries/feminist-social-epistemology/>.
- Hegselmann, Rainer, and Ulrich Krause. 2006. "Truth and cognitive division of labor: First steps towards a computer aided social epistemology." *Journal of Artificial Societies and Social Simulation* 9 (3): 10.
- Klein, Dominik, Johannes Marx, and Kai Fischbach. 2018. "Agent-based modeling in social science, history, and philosophy. An introduction." *Historical Social Research/Historische Sozialforschung* 43 (1 (163): 7–27.
- Loury, Glenn C. 1994. "Self-censorship in public discourse: A theory of "political correctness" and related phenomena." *Rationality and Society* 6 (4): 428–461.
- McKinnon, Rachel. 2016. "Epistemic injustice." *Philosophy Compass* 11 (8): 437–446.
- O'Connor, Cailin. 2019. *The origins of unfairness: Social categories and cultural evolution*. Oxford University Press, USA.



- Schöppel, Leon. 2021. "ABM-Testimonial-Injustice v1.0," <https://github.com/leon-schoeppl/abm-testimonial-injustice/releases/tag/v1.0>.
- Vince, Rosa. 2018. "Testimonial Smothering and Pornography." *Feminist Philosophy Quarterly* 4 (3).
- Wilensky, U. 1999. "NetLogo," <https://ccl.northwestern.edu/netlogo/>.
- Wylie, Alison, Robert Figueroa, and Sandra Harding. 2003. "Why standpoint matters." *Science and other cultures: Issues in philosophies of science and technology* 26:48.