Studies in Philosophy Politics and Economics

Paloma Morales (2020) *Testimonial Injustices and Credibility Scores on Social Media: A Response to Regina Rini*, Studies in Philosophy, Politics and Economics, 2(1), 45-49

Testimonial Injustices and Credibility Scores on Social Media: A Response to Regina Rini

Paloma Morales*

* Philosophy, Logic and Scientific Method, London School of Economics, Houghton Street, London, WC2A 2AE, UK. Email: p.morales@lse.ac.uk

Abstract

To stop the proliferation of fake news, Regina Rini proposed the creation of a clear norm of accountability by assigning publicly visible credibility scores to social media users, calculated using the number of fake news the users previously shared (2017). In this paper, I argue that the potential harm these scores could impose onto underprivileged members of society, through the emergence of testimonial injustices and alienation, justifies discarding Rini's proposal. I then devise a modified proposal which consists in assigning credibility scores to pieces of news, thus avoiding testimonial injustices while combatting the spread of fake news.

Keywords: Fake News; Testimonial Justice; Credibility Deficit

In this paper, I question Regina Rini's solution to the problem of *fake news*: assigning a credibility score to social media users (Rini 2017). I argue that the harm to underprivileged members of society that could result from implementing Rini's proposal justifies discarding it as a solution to the spread of fake news. I then introduce an alternative proposal, which could reduce the potential for harm to underprivileged groups. I organise the paper as follows. Firstly, I show that Rini's proposal could lead to a situation of *testimonial injustice* (Fricker 2007) whereby individuals from underprivileged backgrounds would suffer from increased prejudices. Secondly, because the definition of fake news is disputed and because it is virtually impossible to detect all false information, I posit that the unavoidable arbitrariness of the measure could *alienate* the members of society who are the most vulnerable to fake news. Lastly, I propose an alternative solution which averts the emergence of testimonial injustice.

Was COVID-19 "bioengineered" by the USA military (Westcott et al. 2020)? Is the virus transmitted through radio waves using 5G network (Jones 2020)? Has the Russian government freed lions from the zoos to keep its citizens off the streets (Seitz 2020)? The amount of fake news surrounding the recent pandemic is increasing as citizens around the world are trying to make sense of the unprecedented situation they find themselves in. The desire to contain the spread of fake news mostly surges from the observation that citizens of

democracies can only successfully defend their interests if their opinions are correctly informed. In the context of COVID-19, as in that of any crisis, the value of democracy lies in its ability impose checks and balances on political leaders, rewarding or sanctioning their management capacities. However, when fake news interfere in this process, citizens' ability to scrutinise their leaders is hampered. Attempts at reducing the proliferation and spread of fake news are therefore indispensable.

Rini argues that the proliferation of fake news – defined as "false or misleading information presented as news and circulated with the intention to mislead and to be spread" (2017: 45) – can be attributed to the lack of clear *norms of accountability* on social media. Sharing a piece of news on social media – without a clear disclaimer – is not sufficient for guaranteeing assertion and yet is commonly treated as performing exactly this function. This flawed link between sharing and endorsing is, according to Rini, what renders the propagation of fake news possible. As such, pieces of news shared on social media are treated like ordinary testimonies – which, Rini argues, individuals are epistemologically justified in believing – while lacking a central feature of ordinary testimonies: endorsement (2017: 47-48). She therefore claims that solving the problem requires the establishment of a clear norm of accountability and that this should be done by assigning a credibility score to social media users, calculated using the number of disputed stories – i.e. disputed by third party fact-checkers – they have shared.

To introduce what I believe is the central problem to Rini's proposal, it is essential to reflect on the mechanisms enabling disputed stories to appear as flagged on social media. In particular, we must note that the veracity of every piece of news is – due to the immense and constant influx of information on social media platforms – evaluated long after it is published, shared and re-shared. This is indeed why Rini believes that simply flagging disputed stories is insufficient: by the time a piece of news appears as disputed, hundreds more have replaced it. What should instead be done, according to Rini, is holding individuals accountable for sharing the news they decide to share (2017: 55-57). As such, the credibility score would not blame individuals for voluntarily misleading their friends by sharing information they know to be false – since it is likely that when they shared the news it had not yet been fact-checked by the platform – but, rather, it would guarantee that individuals share only what they genuinely believe to be true. In this way, Rini wants to arrive at a clear norm in which *sharing means endorsing*, thereby thwarting the propagation of fake news.

If Rini's proposal were efficient, that is, if social media users only shared information they believed to be correct, the credibility score would measure nothing else than the accuracy of their initial judgement. Hence, someone would receive a high score if the news they believed to be true actually turned out to be true. Conversely, a low score would indicate a greater propensity to believe in fake news. In a study, Allcott et al. found that an individual's level of education and the probability of believing in fake news were negatively correlated (Allcott et al. 2017: 17). Thus, based on Allcott et al. study, we can conjecture that individuals with lower levels of education would receive a lower score. Furthermore, it is known that level of education and socio-economic background are strongly correlated (Hansen 1997; Galindo-Rueda et al. 2004; Dos Santos et al. 2011) so that underprivileged groups receive systematically lower levels of education. Rini's measure would consequently result in assigning lower credibility scores to members of those groups. In the following paragraph, I argue that the systematic assignment of lower score to members of underprivileged groups could lead to the emergence of *testimonial injustice.*

Testimonial injustice is a situation in which, due to prejudices against the socio-economic group to which a speaker belongs, "she receives less credibility than she otherwise would have", thereby suffering from a *credibility deficit* (Fricker 2007: 20). As argued above, under

Rini's proposal, members of underprivileged groups would systematically receive lower credibility scores. This could allow for the emergence of prejudices against these groups if, for instance, statistics reflecting the relationship between socio-economic group and credibility score were disseminated to the wider population. These negative prejudices would, in turn, generate a credibility deficit for members of underprivileged groups, such that their mere belonging to the group would reduce the amount of credibility they receive in democratic debates, regardless of their individual credibility score. This is hugely problematic in a democratic system which asserts that all individuals should have an equal say in politically relevant matters, despite not currently being in a state of perfect equality regarding education and information.

One could respond to this worry by saying that a society of ill-informed citizens is as dangerous for the health of a democracy as the credibility deficit from which the underprivileged could suffer, were Rini's proposal implemented. Underprivileged groups should of course have the right to be listened to as much as they should have the right to vote but both rights are merely instrumental to the representation of their interests. Accordingly, it is essential that their opinion regarding the effectiveness of means to protect their interests is accurately informed. So it could be argued that Rini's proposal would indeed benefit the underprivileged members of society by enabling them to better defend their interests. In the next paragraph I argue that the proposal's ability to fulfil this educational purpose depends on individuals' perception of the score. It seems that the contested nature of fake news and the methodological obstacles to creating an infallible measure could prevent the score from being accepted as a legitimate measure, especially in the eyes of the underprivileged groups who would systematically receive lower scores.

As low credibility scores would most likely be attributed to underprivileged groups of society, which could result in them suffering from a credibility deficit, the probability of these groups contesting the scores in non-negligible. This is particularly problematic as the notion of fake news is disputed. Even if we were to set aside the debate and adopt Rini's definition, we would still need to make decisions when building a measure, as her definition is silent regarding important nuances. These include whether a fake news "is a completely false story or a partially true story" or whether the phrase can apply to "an individual posting on social media without doing so on the behalf of a news outlet" (Habgood-Coote 2019: 1039). Indeed, even if academics were to agree on an exhaustive definition, any measure of fake news would still, by nature of its complexity, be subject to errors. It could for example mistakenly include satires or miss out on scientifically fraudulent articles. These arbitrary mistakes, even if they turned out to be negligible, would undoubtedly undermine the legitimacy of the score. If individuals with low scores distrust – perhaps justifiably – the neutrality of the measure, it would fail to fulfil its educational purpose. In the rest of this paper, I present a modified version of Rini's proposal which, despite being unable to eliminate the risk of alienation, can prevent the emergence of testimonial injustice.

Rini acknowledges the possibility of alienation but asserts that this problem is "unavoidable for any serious institutional response to fake news" (2017: 58). I find this claim highly plausible. Any attempt at solving the problem of fake news will be ineffective when interacting with individuals unwilling to update their beliefs when presented with new evidence. Instances of such epistemic unresponsiveness are anti-vaccine communities who, despite the publication of numerous debunking research, refuse to vaccinate children by fear of it generating autism (Battistella et al. 2013; Paynter et al. 2019). Still, an alternative solution should work out what is, all things considered, better for marginalised groups. As argued above, any attempt at tackling the problem of fake news will, to a greater or lesser extent, alienate some members of society. Testimonial injustice, however, can be avoided if we introduce some modifications to Rini's proposal. I do so in the following paragraph.

I have claimed that assigning a publicly visible credibility score to users might result in underprivileged groups suffering testimonial injustices. Nevertheless, Rini's criticism that independent fact-checkers evaluate the veracity of news long after they are shared and re-shared and therefore play a negligible role in the combat against fake news, is reasonable. As such, an alternative solution could be to use Rini's *social media users'* credibility scores – calculated exactly as she proposes – to assign a preliminary credibility score to *pieces of news,* before independent fact-checkers can establish with accuracy the veracity of the information. This preliminary credibility score would be determined relative to the credibility scores of the social media users who have shared the news. However, the credibility scores of users themselves would not be made public. Of course, the news score would not be entirely accurate since it is perfectly conceivable that some users share both true and false news, but so would Rini's users score, as it would never be immediately updated. In so doing, this proposal intends to prevent the creation of testimonial injustice by avoiding the emergence of prejudices against the marginalised groups who have a higher propensity to believe in fake news, while providing an indicator of the accuracy of online information that overcomes the delay problem of independent fact-checking. As noted earlier, this proposal would nonetheless unavoidably alienate some social media users who would distrust the classification of the news. However, this seems an unescapable downside of any institutional attempt at combatting the spread of fake news. Whether decision-makers wish to implement an institutional solution to the spread of fake news will depend on how they weight its potential educational benefits against the potential costs of alienation.

To summarise, I have argued against the implementation of Rini's proposal, using two justifications. I have firstly claimed that giving a credibility score to individuals based on whether the news story they shared turned out to be disputed or undisputed by social media platforms' standards, could lead to increased prejudices, mostly against the least socio-economically privileged members of society. Secondly, I have argued that because of the flawed definition of the concept of *fake news* and the unavoidable limitations any measure, Rini's proposed solution could be rejected by the individuals it aims to protect. This is because they could feel alienated from a system which gives them less credibility due to their background – they would receive lower scores *because* they are less educated – and does so by using a non-neutral measure. I have then suggested a modified version of Rini's proposal which, by not making social media users' scores publicly available, could prevent the emergence of testimonial injustice. I have however highlighted that, as any proposal to combat the spread of fake news, my suggestion is unable to entirely eliminate the risk of alienation.

Acknowledgements

I thank Dr Liam Kofi Bright, Professor Alex Voorhoeve and the two student editors for their insightful comments and suggestions on previous drafts.

References

Allcott, H., & Gentzkow, M. 2017. "Social media and fake news in the 2016 election". *Journal of economic perspectives* 31(2): 211-36.

Battistella, M., Carlino, C., Dugo, V., Ponzo, P., and Franco, E. 2013. "Vaccines and autism: a myth to debunk?". *Igiene e sanità pubblica* 69(5): 585-596.

Dos Santos, M. D., & Wolff, F. C. 2011. “Human capital background and the educational attainment of second-generation immigrants in France”. *Economics of Education Review* 30(5): 1085-1096.

Fricker, M. 2007. *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Galindo-Rueda, F., Marcenaro-Gutierrez, O. and Vignoles, A. 2004. "The widening socio-economic gap in UK higher education." *National Institute Economic Review* 190(1): 75-88.

Habgood-Coote, J. 2019. “Stop talking about fake news!”. *Inquiry* 62(9-10): 1033-1065.

Hansen, M. N. 1997. “Social and economic inequality in the educational career: Do the effects of social background characteristics decline?”. *European Sociological Review* 13(3): 305-321.

Jones, R.C. 2020. “Coronavirus: Fake news is spreading fast”*. BBC News*. February 26 2020. URL: https://www.bbc.com/news/technology-51646309

Paynter, J., Luskin-Saxby, S., Keen, D., Fordyce, K., Frost, G., Imms, C. and Ecker, U. 2019. “Evaluation of a template for countering misinformation—Real-world Autism treatment myth debunking”. *PloS one* 14(1).

Rini, R. 2017. “Fake news and partisan epistemology”. *Kennedy Institute of Ethics Journal* 27(2): E-43.

Seitz, A. 2020. “Lion not released on Russian streets to keep people home”*. apnews.com*. March 24 2020. URL: https://apnews.com/afs:Content:8679310226

Westcott, B. and Jiang, S. 2020. “Chinese diplomat promotes conspiracy theory that US military brought coronavirus to Wuhan”. *CNN.com*. March 14 2020. URL: https://edition.cnn.com/2020/03/13/asia/china-coronavirus-us-lijian-zhao-intl-hnk/index.html