Technology and Regulation

The Commodification of Attention, Distrust & Resentment: a Threat to (Rawlsian) Justice

Author(s)	Paige Benton
Contact	pbenton@uj.ac.za
Affiliation(s)	Paige Benton is a post-doctoral fellow at University of Johannesburg, African Centre for Epistemology and Philosophy of Science

Keywords	Mutual trust, stability, political liberalism, resentment, reciprocity, distrust, AI, engagement optimisation, John Rawls, justice
Citation	Paige Benton, The Commodification of Attention, Distrust & Resentment: a Threat to (Rawlsian) Justice, Technology and Regulation, 2025, 232-246 • 10.71265/k6r46952 • ISSN: 2666-139X

Abstract

There is a growing body of scholarship on how AI technology can undermine democratic institutions. I present a novel contribution to this literature by accounting for how and why algorithms for engagement optimisation may undermine the necessary conditions for Rawlsian justice. For Rawls's political theory, the ability to form bonds of trust with fellow citizens is an essential condition for citizens to develop their sense of justice, and their sense of justice is necessary for attaining justice. Recommendation algorithms may amplify the space given to hateful, violent, extremist, false, and discriminatory content, I argue. If this is the case, online social media content could undermine the development of mutual trust between citizens, which is necessary for a sense of justice. If citizens can only trust their like-minded members and distrust their fellow citizens, then the possibility of Rawlsian reciprocity in liberal society may not be realised. The lack of reciprocity is a concern as liberal political systems could be inherently unstable, given that affectionate ties needed for mutual cooperation may be undermined.

1. Introduction

One of the challenges facing current liberal democratic societies is the decline of liberal values and the rise of authoritarian, populist, and nationalist sentiments.[1] Several studies have been undertaken to show the link between artificial intelligence (AI) technologies and the rise of populist or extremist groups, demonstrating how AI technology can undermine democratic institutions.[2] Thus, AI as a knowledge-making power can exacerbate this trend. The 'knowledge-making power' of AI refers to AI technologies' impact on how information in societies is generated, organised, and disseminated. AI technologies, including recommendation algorithms, automated bots, and large language models, to name a few, shape the information that individuals encounter online.[3] In this paper, I focus on how one AI technology – the recommendation algorithm (used for engagement optimisation) – influences the digital information landscape, undermining the development of mutual trust and a sense of justice, the necessary conditions for Rawlsian justice.

For John Rawls, a necessary condition for justice is for citizens to have developed an adequate sense of justice. Without a sense of justice, citizens will not form the essential bonds of mutual trust with other citizens who are not like-minded, i.e., who do not share the same conception of the good. It is necessary to form bonds of mutual trust with citizens who do not share the same moral and political views, since it is this bond of trust that is the moral motivation for citizens to acknowledge, respect, and act in accordance with the requirements of justice. Thus, mutual trust is necessary for a stable liberal constitutional democracy.

Engagement optimisation algorithms skew the information landscape in liberal democracies. The business model of Big Tech[4] companies prioritise engagement optimisation to secure financial revenue. The content that is amplified is content that appears to trigger an emotional reaction from people. In this paper, I focus only on how extremist, hateful, violent, discriminatory and false information goes viral.[5] This kind of viral content is 'harmful' precisely because I claim that it is the kind of content that could cause mistrust and resentment between groups that are not like-minded. Suppose citizens can only trust their like-minded members and develop a distrust of citizens who are not like-minded. In that case, citizens may not

---

1. Samuel Scheffler, 'The Rawlsian Diagnosis of Donald Trump' (Boston Review, *2019)* https://www.bostonreview.net/articles/samuel-scheffler-rawlsian-diagnosis-donald-trump/ accessed 3 June 2024

2. Mark Coeckelbergh, *Why AI Undermines Democracy and What To Do About It* (John Wiley & Sons 2024) 1-160; Mark Coeckelbergh, 'Democracy, Epistemic Agency, and AI: Political Epistemology in Times of Artificial Intelligence' (2023) 3(4) *AI and Ethics* 1341; Joe Whittaker, Seán Looney, Alastair Reed and Fabio Votta, 'Recommender Systems and the Amplification of Extremist Content' (2021) 10(2) *Internet Policy Review* 1–29 https://doi.org/10.14763/2021.2.1565 accessed 30 April 2024; Vedran Džihić, 'The Far Right in the Western Balkans: How the Extreme Right Is Threatening Democracy in the Region' (2023) *Austrian Institute for International Affairs* https://www.ssoar.info/ssoar/handle/document/89591 accessed 3 June 2024; Sam Jackson, 'A Schema of Right-Wing Extremism in the United States' (2019) *International Centre for Counter-Terrorism* https://icct.nl/publication/schema-right-wing-extremism-united-states accessed 3 June 2024.

3. Hanna Kiri Gunn, 'Filter Bubbles, Echo Chambers, Online Communities' in Michael Hannon and Jeroen de Ridder (eds), *The Routledge Handbook of Political Epistemology* (Routledge 2021) 259–270; Catherine Smith, 'Automating Intellectual Freedom: Artificial Intelligence, Bias, and the Information Landscape' (2022) 48 *IFLA Journal* 422; Dipto Barman, Ziyi Guo and Owen Conlan, 'The Dark Side of Language Models: Exploring the Potential of Llms in Multimedia Disinformation Generation and Dissemination' (2024) *Machine Learning with Applications* 100545.

4. By 'Big Tech' I refer to the technology companies that have the most influence (often known as the 'Big Five') Alphabet (The parent company of Google), Amazon, Apple, Meta, and Microsoft; *See* Josh Hawley, *The Tyranny of Big Tech* (Regnery Publishing 2021) 1–200.

5. An important caveat to note, is that this paper only focuses on one kind of viral content (i.e., extremist, hateful, violent, discriminatory content). The other kind of viral content is content that promotes emotional connections between users by using humour, feelings of happiness, and emotional connections to activate a dopamine release or provide one with an awe-inspiring feeling. The point of focusing on extremist, hateful, violent, discriminatory content that has gone viral is that this kind of viral content (as I demonstrate in this paper) poses a specific threat to justice that humorous, awe-inspiring and happy viral content does not. I do not deny that viral content could help to foster positive societal outcomes, however it is not within the scope of this paper to assess this. Instead, this paper has a narrow focus on harmful viral content that could threaten the stability of liberal democracies. *See* Karine Nahon and Jeff Hemsley, *Going Viral* (Polity Press 2013) 1-182; Harvey G.O. Igben and Okiemute Endurance Acchugbue, 'Influence of Viral Contents on the Rapid Spread of Information on the Social Media Platforms in Nigeria' (2024) 12(6) *British Journal of Marketing Studies* 24–40.

be willing to act in reciprocal ways with those with whom they hold moral disagreements.[6] This would undermine Rawlsian reciprocity. Without reciprocity, liberal political systems will be inherently unstable as citizens would not have formed the adequate affectionate ties needed for mutual cooperation, which is a precondition for a just society.

In section 1, I provide an exposition of Rawlsian theory to demonstrate the conditions of reasonable moral and political knowledge claims, the role of mutual trust as a moral motivation for a sense of justice, and the stages of moral development necessary for an adequate sense of justice. In section 2, I briefly sketch the moral views that have historically existed in the United States of America. I demonstrate how feelings of resentment and distrust as a result of individuals' conflicting ideas of the good and the threat of instability have existed prior to the development of AI technology. In section 3, I address how epistemic bubbles, echo chambers, group polarisation, false information and conspiracism have been a part of the historical knowledge landscape of democracies before AI. Yet their effects are amplified by the business model of AI. Then, in section 4, given that AI appears to amplify moral and political disagreement, hateful and discriminatory content, I demonstrate how this leads to moral views that undermine liberal values as being given a place to flourish in the public sphere. It is the flourishing of these values that undermines the development of trust and a sense of justice.

2. Rawlsian Exposition

Let us first discuss what epistemically justified political and moral knowledge is for John Rawls. Political and moral knowledge have contrasting constraints for Rawls. Moral knowledge is knowledge that is associated with a prescription of the good. The good is a category term to refer to a moral value, goal, or final end that is meaningful for a person based on a full or partial comprehensive doctrine that they follow.[7] Persons follow certain conceptions of the good and comprehensive doctrines over others due to what moral truth claims they find persuasive. For instance, a citizen may choose to subscribe to Buddhism as opposed to Christianity if they find the Four Noble Truths necessary for a meaningful life. Although moral truth is important for the construction of citizens' rational life plans it should not be the foundation for political knowledge. This is why, in constructing political liberalism, he replaces the search for moral truth with the search for moral reasonability as the foundation of justice.

Political constructivism is one of the core features of Rawlsian justice. It requires that the condition of reasonability and not truth be extended to persons, principles, political judgements, and institutions.[8] Truth should not be the foundation of political knowledge since it cannot form the basis of political agreement needed for an overlapping consensus, given liberty of conscience. Liberty of conscience is a presupposed feature of liberal society, it entails that persons have the freedom to construct, revise and follow any idea of the good.[9] A just society for Rawls can accommodate a multitude of conflicting ideas of the good. Thus, reasonable moral pluralism[10] is a precondition for any liberal democratic constitutional theory of justice,

6. One may wonder, do people need to trust others that stand in moral opposition to them? For instance, should trans people trust transphobes? There are legitimate epistemic reasons for marginalised persons to have distrust in groups or people that fundamentally threaten their existence, thus there are good reasons to justify the epistemic safety that echo chambers, and epistemic bunkers provide to marginalised persons. However political stability in liberal democracies does require that those who stand in moral opposition to one another (even when their political relationship may be characterised by an unequal power dynamic) need to have a minimal level of trust in one another. Insofar as both groups of people can trust one another to uphold the demands of justice. For further discussion see: Jennifer Lackey, 'Echo Chambers, Fake News, and Social Epistemology' in Sven Bernecker, Amy K Flowerree and Thomas Grundmann (eds), *The Epistemology of Fake News* (1st edn, Oxford University Press 2021) 206; Katherine Furman, 'Epistemic Bunkers' (2023) 37 *Social Epistemology* 197; Paige Benton and Michael W. Schmidt, 'The Harm of Social Media to Public Reason' (2024) 43 *Topoi* 1433.

7. For a distinction between partial and fully comprehensive see: John Rawls, *Political Liberalism: Expanded Edition* (Columbia Classics in Philosophy edition, Columbia University Press 2005) 13–14.

8. Rawls (n7) 93–96.

9. Rawls (n7) 94, 150-151.

10. "The fact of reasonable pluralism implies that there is no such doctrine, whether fully or partially comprehensive, on which all citizens do or can agree to settle the fundamental questions of political justice". John Rawls, *Justice as Fairness: A Restatement* (Harvard University Press 2001) 24.

such as his. Due to liberty of conscience and reasonable moral pluralism, if political knowledge were to be grounded on moral truth claims, then persons would be relying on their comprehensive moral doctrines in forming their knowledge of what a good political principle, political judgement or institution ought to be.[11] This is problematic since it could lead to the political system promoting certain conceptions of the good over others, which not all persons could agree to.[12]

For Rawls, it is necessary to secure the right kind of power dynamics between persons and their ideas of the good for social and political stability. The 'right kind' of power balance would be one in which citizens do not attempt to use state institutions to promote their conception of the good. This could lead to political instability, when persons in power change then state institutions would be held to the contingency of person(s) vying for their ideas of the good to be promoted which in theory may contribute to deep civil disagreement on the questions of the good that cannot be reasonably resolved given the burdens of judgement.[13] Thus, the alternative for Rawls is for political knowledge to meet the criterion of reasonability as opposed to truth.

For political knowledge to be reasonable it requires persons to have the "willingness to propose and abide by fair terms of social cooperation among equals and their recognition of and willingness to accept the consequences of the burdens of judgment".[14] In contrast to the reasonable conditions for agreement previously described, if "prejudice and bias, self- and group interest, blindness and wilfulness" are the sources of justification for persons' political beliefs, then these beliefs are founded on an unreasonable basis for disagreement.[15]

For Rawls, the ability of persons to adhere to the reasonable conditions of belief formation or justification rests on the conception of the person he develops. Briefly, he defines persons as having two capacities. The capacity for a sense of good and a sense of justice.[16] One exercises their sense of good when deliberating about their rational life plan and engaging in associational life. In comparison exercising one's sense of justice requires citizens to propose arguments that are accessible for all fellow citizens to accept when deliberating on issues of justice.[17] What qualifies an argument as 'accessible' is to ensure that the justifications citizens provide appeal to political values, such as liberty, equality, and equal opportunity.[18] In other words, citizens must provide public reasons for matters of justice such as political policies. Public reason is public insofar as the reasons provided are justified by an appeal to the values embedded in the public political culture of the liberal political tradition.[19]

The criterion of reasonability encompassed in public reason ensures that an overlapping consensus can be reached as the foundation for public consensus is that which persons can consider acceptable to agree to in light of the circumstances of justice. The motivation for persons to adhere to public reason, reasonable disagreement, and the exercise of political values such as "toleration and mutual respect, and a sense of fairness and civility" is the development of the moral feeling of trust.[20]

Trust is an important element in the stability of Rawlsian justice. The social basis of mutual trust is grounded on the equal liberties of persons as a constitutional essential.[21] People have equal rights and as such in theory, they should have trust that the system is built for the equal benefit of all citizens.[22] Moreover, trust is

11. 'Moral truth claim' implies that the kinds of claims proposed are claims founded on a comprehensive notion of the good. A 'comprehensive notion of the good' refers to a doctrine, theory or way of life that prescribes what is morally valuable and desirable to a person. *See* W.B Mahan, 'The Right and The Good in Theory and Practice' (1924) 34(1) *The Monist* 112-127.
12. Rawls (n10) 192–194.
13. Rawls (n7) 54–56.
14. Rawls (n7) 94.
15. Rawls (n7) 58.
16. For a full discussion see: Rawls (n 7) 47–86.
17. John Rawls, 'The Sense of Justice' (1963) 72 *The Philosophical Review* 281, 282.
18. Rawls (n7) 194.
19. John Rawls, *Justice as Fairness: A Restatement* (Harvard University Press 2001) 26–28.
20. For a detailed discussion of trust and guilt in relation to his three stages of moral development see: Rawls (n19) Chapter VIII and Rawls (n17).
21. Rawls (n7) 181.
22. Rawls (n10) 126, 138, 218.

further developed among citizens firstly when they see the basic structure of society satisfying the principles of justice. Secondly, when citizens see their fellow citizens developing and acting in line with the political values, they too are more likely to develop and act from the same virtues gradually and over time as their confidence in the political system and one another grows. Rawls refers to this aspect of trust as citizens' mutual public recognition for one another.[23]

Trust plays a central role in the condition of reciprocity[24] between citizens. Reciprocity is achieved when citizens are willing to respect all citizens' entitlement to basic liberties and adhere to conditions of reasonability. Trust is necessary for stability as it is a core determining feature as to whether citizens can uphold or violate the criterion of reciprocity. Rawls developed his theory of moral development[25] and reasonable moral psychology[26] to account for how a minimal level of trust and confidence can develop in citizens to help generate a just society.

During all three stages of development[27], i.e., morality of authority, morality of association, and morality of principles, children and citizens learn from one another to reciprocate ties of affection and general moral rules. Briefly, during the first stage, children develop feelings of trust towards their parents. It is these feelings that motivate children to follow the rules of the household. Following rules starts the child's development of a sense of justice. During the second stage, citizens learn how to be just in the connections they form in their associational life. Institutions such as families, schools, and religious organisations require their members to be able to uphold shared rules of engagement that control the internal running of their organisations. Members choose to act in line with these rules out of the trust that their fellow members will do the same. At this stage, citizens further develop their sense of justice as they choose to uphold rules for the collective benefit of membership with like-minded individuals. The final stage is the morality of principles. Just as citizens understand the benefits of mutual cooperation within their associations, at this stage, they learn how these benefits of cooperation can be extended to broader political society. An adequate sense of justice is achieved when citizens acknowledge and trust that their fellow citizens are committed to a shared common political aim. Rawls states that civic friendship is a possibility that results when citizens acknowledge their shared commitment to the principles of justice and maintain adherence to political values for their shared benefits.[28]

Rawls in *Political Liberalism* states: "If other persons with evident intention strive to do their part in just or fair arrangements, citizens tend to develop trust and confidence in them; iv) this trust and confidence becomes stronger and more complete as the success of cooperative arrangements is sustained over a longer time".[29] Trust is essential to the development and reproduction of a just society, as it is a precondition needed for just citizens. Without the adequate development of one's sense of justice, citizens will not develop the moral feeling of guilt when they transgress the rules of justice. The feeling of associational guilt is necessary for citizens as it is the moral feeling that enables citizens to want to rectify their behaviour and "make reparations" when trust is broken.[30]

The point of ending this brief exposition with a discussion of Rawlsian trust and guilt is to show that a precondition for reasonable political beliefs for Rawls depends on citizens internalising the demands of justice. Thus, citizens must adhere to the requirements of reasonability, exercise political virtues, and develop their sense of justice, as this is essential for fair cooperation. According to Rawls: "When fair terms are not honoured, those mistreated will feel resentment or humiliation, and those who benefit must either recognise their fault and be troubled by it, or else regard those mistreated as deserving their loss. On

23. Rawls (n19) 405–409.
24. Reciprocity is defined as "all who do their part as the recognized rules require are to benefit as specified by a public and agreedupon standard"; Rawls (n10) 24.
25. Rawls (n19) Sections 70,71,72 in Chapter VIII.
26. Rawls (n7) Lecture II, Section 8.
27. Rawls (n19) Chapter VIII.
28. Rawls (n19) 415.
29. Rawls (n7) 86.
30. Rawls (n17) 105.

both sides, the conditions of mutual respect are undermined".[31] When 'the conditions of mutual respect are undermined', the adherence to and the putting forth of reasonable political beliefs becomes unlikely as conditions for reasonable political beliefs and mutual trust are eroded, as I demonstrate in section 4 of this paper.

Now that I have outlined the ideal theoretical conditions for achieving justice in a liberal democratic society, it is important to examine the non-ideal conditions of the liberal democratic society of the United States of America (USA) to assess the challenges facing the development of trust.

3. The knowledge landscape and instability of non-ideal liberal constitutional democracies:

In Section 2 I provided an exposition of the conditions for reasonable political beliefs in Rawlsian theory. Let us now turn to the knowledge landscape of non-ideal liberal constitutional democracy of the USA to discuss the conditions that could undermine reasonable political belief formation and break down trust before the impact of AI algorithms. Fundamentalist, illiberal and prejudiced values and groups impacted the social, political and knowledge landscape of the US (and other liberal democracies) prior to AI. Below, I demonstrate how these phenomena challenge the reasonable knowledge landscape and have undermined the normative foundations of a liberal democratic society.

Let us look at fundamentalists. Historically, the cases of Wisconsin vs Yonder[32] and Mozert vs Hawkins[33] highlight the tension between liberal education and illiberal doctrines. In the first case, Amish[34] parents fought to remove their children from state high schools due to their fear of their children being exposed to liberal education and associational ways of life. In contrast, in the second case, the plaintiffs (Protestant fundamentalists) challenged the removal of textbooks that exposed their children to the liberal values of tolerance and moral pluralism. In both cases, plaintiffs claimed that exposing children to liberal values and pluralist culture undermines the respective fundamentalist values and lifestyles that these parents teach and desire for their children. Thus, fundamentalist view cultural pluralism as a threat to the reproduction of their way of life. Fundamentalists will have a hard time supporting the liberal political system since they cannot buy into the core liberal assumptions, such as moral pluralism and tolerance. Fundamentalists in a liberal society may find it especially difficult as their notion of the good is marginalised. It is essential to look at historical examples in the knowledge landscape prior to AI, in order to assess how moral values or values of comprehensive doctrines could be in tension with, or undermine, political values before the disruption of engagement optimisation algorithms.

Similar to fundamentalists, citizens who hold prejudiced or illiberal views may find themselves at odds with the liberal system, struggling to accept its core tenets and feeling marginalised within a political system that prioritises equality and diversity. The liberal commitment to equal rights and opportunities for all individuals may directly contradict the discriminatory attitudes and actions of sexist and racist individuals, making it difficult for them to reconcile their beliefs with the principles of liberalism. Persons who adhere to rigid, exclusionary and discriminatory doctrines will be required to suppress or restrict their views to the private sphere in favour of adherence to political values of inclusivity and tolerance in the public sphere.

Historically, this tension between associational groups rejecting liberal values can be seen in instances such as the Ku Klux Klan in 1963 with the Baptist Church Bombing, the creation of political organisations such as the Moral Majority, the Christian Coalition and the Family Research. All of these groups demonstrate a rejection of liberal values insofar as they reject extending equal rights and liberties to persons based on either their race, gender or sexuality, or they aim to expand their conception of the good into political institutions.

31. Rawls (n7) 302–303.
32. *Wisconsin v. Yoder*, 406 U.S. 205 (1972).
33. *Mozert v. Hawkins County Board of Education*, 827 F.2d 1058 (6th Cir. 1987).
34. Stephen Macedo, 'Liberal Civic Education and Religious Fundamentalism: The Case of God v. John Rawls?' (1995) 105 *Ethics* 468.

Associational groups such as these undermine the conditions for reasonable political belief formation, impacting the formation of trust and the development of a sense of justice. As they reject at least one of the three conditions of reasonable participation, namely: (1) reasonable moral pluralism, (2) all citizens must regard others as free and equal, and (3) the political system is an open and fair system for the mutual benefit of all citizens.[35] Citizens who reject these conditions, Rawls regards as unreasonable citizens.[36]

If these citizens reject the conditions of reasonableness for managing public life, they will be unable to propose political claims that other citizens would consider reasonable. Additionally, they may not recognise the legitimacy of reasonable political claims made by their fellow citizens. As Rawls states, "unreasonable doctrines are a threat to democratic institutions since it is impossible for them to abide by a constitutional regime except as a modus vivendi".[37] In other words, unreasonable persons adhere to the rules of society for instrumental purposes. They are biding their time until the power dynamics shift, giving them the opportunity to insert their conception of the good into the political institutions. Thus, unreasonable persons pose a threat to the stability of a liberal democracy precisely because they cannot adhere to the liberal requirements of justifying political knowledge via the conditions of public reason.

Samuel Scheffler[38], diagnoses the decline of liberal values in America and the rise of populism. He states that Trump's presidential win in 2017 is a sign of the "vindication of liberal theory" as opposed to the demise of it. As mentioned, Rawlsian liberalism posits the ideal of reciprocity as a necessary condition for the stability of a liberal society. Scheffler claims that the USA fails to meet the Rawlsian reciprocity criteria.[39] The failure of social reciprocity can be seen in protests such as the Black Lives Matter (BLM) movement. Whereby black citizens protesting systemic racism is suggestive of the fact that there is a breakdown of reciprocity, insofar as political and social institutions fail to protect the interests of black and white citizens equally. This leads marginalised citizens (i.e., citizens facing structural discrimination) to feel resentment towards the institutions as they feel that they are unfairly disadvantaged. Scheffler argues that if reciprocity is not adequately developed then American liberal democratic society may be inherently unstable and this could lead to an increase in resentment among citizens.

In contrast to Scheffler, who demonstrates how civil disobedience is a sign of resentment and lack of reciprocity, Alessandra Tanesini[40] develops an alternative account to explain 'the politics of resentment'. Persons who engage in this are persons who "experience a loss of some entitlements previously conferred to them in virtue of their dominant ethnic or gender identity". Thus, these individuals view themselves as suffering an injustice when political power dynamics change.

Arlie Hochschild in *Strangers in Their Own Land Anger*[41] provides sociological investigations into Tea Party voters in Louisiana, highlighting the grievances of persons engaged in the politics of resentment. Hochschild presents a metaphorical story to expose the sentiments of her interviewees. Below are extracts taken from the deep story:

35. Jonathan Quong, *Liberalism without Perfection* (Oxford University Press 2011) 1-352.

36. Important to note is not all fundamentalists, prejudiced or illiberal groups are considered unreasonable on the basis of having conflicting beliefs. They are unreasonable if they impose their moral beliefs on the political system. Persons who upload their illiberal beliefs in the private sphere but are willing to adhere to fair terms of cooperation in the public sphere would be considered reasonable.

37. Rawls (n7) 489.

38. Scheffler (n1). In contrast, Weithman focuses on the economic inequalities and sociological evidence from Hochschild in *Strangers in Their Own Land: Anger and Mourning on the American Right* (2016) to argue that the ability for Rawlsian reciprocity to be achieved could be challenged by the downward-directed resentment citizens have for those that are benefiting from redistributive economic policies. *See* Paul Weithman, 'Comment: Reciprocity and the Rise of Populism' (2020) 26 *Res Publica* 423.

39. The second failure of Rawlsian reciprocity for Scheffler (n1) is evidence of the excessive increase in economic inequalities in the US, which I do not address in this paper.

40. Alessandra Tanesini, 'The Politics of Resentment: Hope, Mistrust, and Polarization' in Hana Samaržija and Quassim Cassam (eds), *The Epistemology of Democracy* (Routledge 2023) 115–134.

41. Arlie Hochschild, *Strangers in Their Own Land: Anger And Mourning on the American Right* (The New Press 2016) 135-137.

> *"You are patiently standing in a long line leading up a hill, as in a pilgrimage. You are situated in the middle of this line, along with others who are also white, older, Christian, and predominantly male, some with college degrees, some not... Many in the back of the line are people of colour – poor, young, and old, mainly without college degrees...You think of things to feel proud of – your Christian morality, for one. You've always stood up for clean-living, monogamous, heterosexual marriage... Liberals are saying your ideas are outmoded, sexist, and homophobic, but it's not clear what their values are ... Blacks, women, immigrants, refugees, brown pelicans – all have cut ahead of you in line... You resent them, and you feel it's right that you do ... People complain: Racism. Discrimination. Sexism. You've heard stories of oppressed blacks, dominated women, weary immigrants, closeted gays, and desperate refugees but, at some point, you say to yourself, you have to close the borders to human sympathy – especially if there are some among them who might bring you harm... If you can no longer feel pride in the United States through its president, you'll have to feel American in some new way – by banding with others who feel like strangers in their own land".*

These condensed partial extracts highlight an alternative kind of resentment than that which Scheffler argues for. In Scheffler's case, members of the BLM movement are not members of a historically dominant group or protesting an entitlement they have lost. Historically (and currently) black persons in America have been oppressed, and their protests against the policing system highlight the systemic injustices that are persistent in these political institutions. Scheffler's example of resentment demonstrates a failure of the USA to create institutions that are fair and equal for all.

In contrast, Hochschild's investigations support Tanesini's claim by demonstrating how a dominant group (i.e., white Christian men) feel that they have lost previous entitlements (i.e., moral dominance and political privilege) and as a result, feel powerless.[42] These voters feel resentful that Christianity is no longer the moral unifying feature of America. With the rise of moral pluralism and liberty of conscience, these voters are confronted by views that challenge theirs and view their moral values as "outmoded, sexist, homophobic".[43] The sentiments of these voters seem to suggest that they long for a time when their religion was prioritised by the state. They also feel that they are losing access to resources they are entitled to. The resentment these voters feel towards 'line cutters' is directed at historically disadvantaged persons, as there is a sense that the latter are undeserving of the advantages. Thus, the dominant group views marginalised groups' advantages as unjust.

In both Scheffler's and Tanesini's accounts of resentment, both groups (i.e. those that are marginalised by the system (black women and men), and those that view themselves as marginalised by the system (white Christian males) have lost trust in political institutions. It appears as if each group believes political institutions (such as the police system) serve the interests of the opposing group rather than their own. It is outside the scope of this paper to assess the legitimacy of each group's distrust and resentment here. It is pertinent to note that both groups (i.e. marginalised and dominant groups) at the very least experience minimal distrust with one another regarding how their interests are represented in political institutions. Let us now look at how this mutual distrust undermines reciprocity.

For Rawls (as stated in section 2), trust is an essential moral feeling underlying a person's moral motivation for fair social cooperation. Mutual cooperation requires citizens to be able to trust fellow citizens who are not like-minded. In other words, citizens who share no similarity in terms of their moral beliefs must be able to trust other citizens' commitment to upholding the requirements of justice and be committed to cooperation on the foundation of freedom and equality for all. When trust is eroded, it can lead to resentment, undermining reciprocity and hindering cooperative efforts.

The point of highlighting the tension between liberal and illiberal values and groups is to demonstrate the moral conflict underlying the development of reasonable political beliefs. When citizens fail to propose

42. Tanesini (n40) 116; Margaret Urban Walker, *Moral Repair: Reconstructing Moral Relations after Wrongdoing* (Cambridge University Press 2006) 108.

43. Hochschild (n41) 135-137.

reasonable political beliefs, in theory, this can cause instability as it reflects the lack of commitment to exercising one's sense of justice, which I examine in section 4. It is essential to emphasise that instability created by resentment and distrust between citizens existed before AI's interference, complicating efforts to build a cohesive and cooperative society. Highlighting this tension emphasises that the challenges in achieving mutual trust and cooperation are deeply rooted and multifaceted, extending beyond technological disruptions to fundamental disagreements on core moral values. In the following section, I examine how the AI industry amplifies this moral disagreement, resentment and distrust.

4. AI recommendation algorithm amplifies distrust and hate in non-ideal liberal constitutional democracies:

AI algorithms mediate individuals' existence in the world and to those around them. AI algorithms work in the background of our daily lives, shaping our experiences and decisions of what we should buy, how we should vote, and what we should believe in.[44] Algorithmic recommendations direct the kinds of information individuals receive from governments, media outlets, companies, organisations and other citizens. People engaging with the information received, like and share some information, while scrolling past and discarding other information. This continual recommendation, interaction and selection process leads to personalised information networks and the viral explosion of some information over others.

In current debates, analysis of personalised information networks focuses on the impact of epistemic bubbles, echo chambers, fake news, and group polarisation on individuals' belief formation.[45] Some argue that these phenomena in the digital environment undermine democracy.[46] Cohen and Fung[47] in their comparative analysis of information production and flow in America during the mass media versus the digital public spheres, demonstrate that neither is conducive to democratic objectives. Similarly, it is crucial to recognise that epistemic bubbles, echo chambers, fake news, and group polarisation affected the informational landscape of liberal democracies such as the USA even before the AI revolution. All these phenomena, whether experienced because of AI algorithmic influence or not, are detrimental to democracy since they heighten moral disagreement.

Epistemic bubbles[48] form as a result of individuals being selective about the information they are exposed to. Whether they are self-imposed (i.e., choosing to only listen to the same radio show) or a result of the environment (i.e., communities relying on their one local newspaper), they limit persons' exposure to a broader range of views which can lead to reinforcing existing beliefs and having an inflated self-confidence of one's beliefs. In contrast to epistemic bubbles, echo chambers form and reinforce themselves when they actively discredit outside sources. C. Thi Nguyen refers to this as the 'disagreement-reinforcement' mechanism.[49] It is this mechanism that leads to distrust, as echo chamber members learn to develop an inflated distrust of non-members and a heightened trust and confidence in their group's beliefs. Cults are associational groups in liberal society currently (and

44. Sinan Aral, The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health—and How We Must Adapt (Crown Publishing Group 2020) 1–416.

45. Engin Bozdag and Jeroen van den Hoven. 'Breaking the Filter Bubble: Democracy and Design' (2015) 17(4) *Ethics and Information Technology* 249.

46. Lucas Introna and Helen Nissenbaum, 'Shaping the Web: Why the Politics of Search Engines Matters' (2000) 16 *The Information Society* 169; Eli Pariser, The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think (Reprint edn, Penguin Publishing Group 2012) 1–304; Boaz Miller and Isaac Record, 'Justified Belief In a Digital Age: On The Epistemic Implications of Secret Internet Technologies' (2013) 10 *Episteme* 117; Mostafa M. El-Bermawy, 'Your Filter Bubble Is Destroying Democracy' (Wired, 18 November 2016) https://www.wired.com/2016/11/filter-bubble-destroying-democracy/ accessed 24 August 2023; Dave Kinkead and David M. Douglas, 'The Network Unlikely Destroys Big Data & the Epistemic Justifications of Democracy' in Kevin Macnish and Jai Galliott (eds), *Big Data and Democracy* (Edinburgh University Press 2020) 114–131; Mark Coeckelbergh (n2) 1341.

47. Joshua Cohen and Archon Fung 'Democracy and the Digital Public Sphere' in Lucy Bernholz, Helene Landemore, and Rob Reich (eds), *Digital Technology and Democratic Theory* (University of Chicago Press 2021) 23-62.

48. C. Thi Nguyen, 'Echo Chambers and Epistemic Bubbles' (2020) 17 *Episteme* 141, 144.

49. Nguyen (n48) 141, 147.

before the digital age) that exercise(d) this 'disagreement-reinforcement' mechanism.[50]

Both epistemic bubbles and echo chambers help exacerbate group polarisation[51] since individuals within epistemic bubbles and echo chambers become more entrenched in their views or groups' beliefs. In contrast to this argument, other theorists argue that individuals exposed to alternative beliefs can lead to them holding a more extreme position than their initial view, further contributing to group polarisation.[52] Group polarisation is not new in liberal democracies. Historically, wherever there are moral and political conflicts, there is polarisation to some degree. For instance, one can see group polarisation historically in America such as during the Civil War between anti-slavery and pro-slavery movements.[53] However, polarisation currently is the highest it has been in 140 years, suggesting there is something unique about the current social and political climate that increases polarisation.[54]

Fake news and conspiracism add to the intensity and proliferation of group polarisation. Fake news is misleading or incorrect information (i.e., fabricated stories, exaggerated claims and distorted facts) presented as factual news. Fake news aims to mis/disinform [55] persons and manipulate public opinion. Although the term fake news gained popularity in 2016, the phenomenon of false information in the public sphere to sway public opinion is not new.[56] There have always been instances of misleading, false or exaggerated claims in the public sphere, such as 'yellow journalism' in the early 1950s.[57]

Fake news thrives in what Cohen and Fung call a many-to-many communication network.[58] Print media and media broadcasters are examples of one-to-many communication technologies; they are characterised by a narrow concentration of voices that portray views and information to large audiences, while at the same time, these audiences do not have the space to respond. In contrast, in a many-to-many communication network (indicative of a digital information network) the flow of information is not top-down. Here, there is a broad base of voices that can create and disseminate information while simultaneously being able to respond to other views in turn. According to Cohen and Fung[59], content moderation takes place before news is aired or printed in a one-to-many information system, thus reducing the scale of false information. Meanwhile, in a many-to-many communication environment, fake news thrives as content moderation only occurs after posting and relies on user feedback. Although Cohen and Fung are correct that some content moderation happens after the fact, there is 'ex ante' moderation in many-to-many communication networks. 'Ex ante' moderation is when companies like Meta use algorithms or filters to remove content before users see it.[60] The use of ex ante may weaken Cohen and Fung's reasoning for the increased spread of fake news on

50. For a further discussion of cults see: Margaret Thaler Singer and Janja Lalich, Cults in Our Midst: The Hidden Menace in Our Everyday Lives (Jossey-Bass 1995) 1–381.

51. Christopher A. Bail and others, 'Exposure to Opposing Views on Social Media Can Increase Political Polarization' (2018) 115 (37) *Proceedings of the National Academy of Sciences* 9216 https://doi.org/10.1073/pnas.1804840115 .

52. Robert B. Talisse, 'Problems of Polarization' in Jeroen de Ridder, Michael Hannon and Miranda Fricker (eds), *Political Epistemology* (Routledge 2021) 209–226.

53. Jack M. Balkin, 'The Cycle of Polarization' in *The Cycles of Constitutional Time* (Oxford University Press 2020) 187–211.

54. Cohen and Fung (n47) 37; Nolan McCarty,and Keith T. Poole, 'An Empirical Spatial Model of Congressional Campaigns' (1998) 7 *Political Analysis* 1; Nolan McCarty, 'In Defense of DW-NOMINATE' (2016) 30(2) *Studies in American Political Development* 172; Nolan McCarty, *Polarization: What Everyone Needs to Know* (Oxford University Press 2019) 30.

55. Sander van der Linden, *Foolproof: Why Misinformation Infects Our Minds and How to Build Immunity* (WW Norton & Company 2023) Chapter 4.

56. Julien Gorbach, 'Not Your Grandpa's Hoax: A Comparative History of Fake News' (2018) 35 *American Journalism* 236; Julie Posetti and Alice Matthews, 'A Short Guide to the History of "Fake News" and Disinformation' (International Center for Journalists, July 2018) accessed 24 August 2023 https://www.icfj.org/news/short-guide-history-fake-news-and-disinformation-new-icfj-learning-module

57. David R. Spencer, The Yellow Journalism: The Press and America's Emergence as a World Power (Northwestern University Press 2007) 1–272.

58. Cohen and Fung (n47) 36.

59. Cohen and Fung (n47) 37.

60. Vincent Chiao and Alon Harel, 'Content Moderation Online: Regulation Ex Ante versus Ex Post' (2023) *University of Illinois Law Review* 1587; Paddy Leerssen, 'An End to Shadow Banning? Transparency Rights in the Digital Services Act Between Content Moderation and Curation' (2023) 48 *Computer Law & Security Review* 105790.

such platforms; however, on the other hand, one needs to question the effectiveness of ex ante moderation, given that context and nuance are essential to determining the boundary of acceptability.[61] For the relevance of this paper, it is adequate to note that many-to-many communication networks are a breeding ground for fake news, whether it be because of ex post moderation or the failure of effective ex ante moderation.

Fake news and the new conspiracism[62] help to reinforce one another. Rosenblum and Muirhead characterise classical conspiracism as a movement that some persons engage in to make sense of the world; in doing so, they attribute a powerful group of people as being solely responsible for certain social and political occurrences. Classical conspiracists, in justifying the connections they make, rely on drawing connections between who, where, what, why and how, thus demonstrating an alternative trail of causal connections. Examples of events that classical conspiracists have developed alternative explanations for are the moon landing and 9/11. In contrast, Muirhead and Rosenblum state: "The new conspiracism satisfies itself with a free-floating allegation disconnected from anything observable in the world".[63] The new conspiracism does not rely on evidence, argumentation or explanation instead, they call for collective action around forms of "bare associations" linked to fake news. As Muirhead and Rosenblum point out: "With every use of the term fake, conspiracists insist on the reality of a plot to make up news stories, concoct fictitious intelligence reports, and manufacture data—deliberately, not want only. And the conspiracist response is not correction or setting things straight; "fake" is the entire response. There is nothing more".[64]

Overall, the occurrence of selective exposure, confirmation bias, active discrediting, heightened in-group trust, heightened out-group distrust, the entrenchment of moral and political views, the creation of misinformation, and the reliance on bare associations all form as a result of these phenomena. All of these elements could contribute to fostering a general atmosphere of political distrust among citizens and thus may contribute to a fragmented informational landscape.[65] A fragmented informational landscape impairs democratic discourse, as individuals become less willing to engage with or understand opposing perspectives, leading to a more divided and contentious society. Now I turn to an exposition of how and why AI recommendation algorithms may amplify this already fragmented informational landscape of liberal democracies.

A core feature of the business models of Google, Meta, X, Instagram, YouTube, etc. is optimising user engagement. These companies are concerned with optimising user engagement as increased engagement translates to increased financial revenue. Algorithms help to optimise user engagement by attempting to predict the likelihood of user engagement with certain kinds of content by analysing users' behaviour, demographic and geolocation data. Narayana points out that recommendation algorithms are most effective at identifying content that the majority will like and niche content that specific subgroups will be interested in.[66] Due to the former, recommendations can create digital information cascades, leading to viral information trends. Information cascades can have good or bad societal implications depending on the content.

According to whistleblower Francis Haugen, Facebook's viral posts contain misinformation and harmful content and encourage division and anger.[67,68] The fact that hateful, violent, extremist, and prejudicial content becomes viral is not suggestive that the majority want to see it or find it good, but rather that it

61. Johana Bhuiyan, 'Google Refuses to Reinstate Man's Account After He Took Medical Images of Son's Groin' *The Guardian* (22 August 2022) https://www.theguardian.com/technology/2022/aug/22/google-csam-account-blocked accessed July 2024.
62. Russell Muirhead and Nancy L. Rosenblum, A Lot of People Are Saying: The New Conspiracism and the Assault on Democracy (Princeton University Press 2019) 1–232.
63. Muirhead and Rosenblum (n62) 38.
64. Muirhead and Rosenblum (n62) 26.
65. Empirical research would need to be undertaken to substantiate this point. It is beyond the scope of this paper to address this empirical gap. It would be essential for empirical research to be conducted by other social sciences to demonstrate the strength or weakness of this claim.
66. Arvind Narayanan, 'Understanding Social Media Recommendation Algorithms' (Knight First Amendment Institute, 9 March 2023) https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms accessed 24 July 2024.
67. Narayana (n66) 60-61.
68. Amnesty International, 'Surveillance Giants: How the Business Model of Google and Facebook Threatens Human Rights' (2019) 5–17 https://www.amnesty.org/en/documents/pol30/1404/2019/en / accessed 24 July 2024.

gains traction due to its shock value.[69] Recommendation algorithms that learn from behavioural data are learning to amplify content that persons like, share or comment on based on "unconscious, automatic and emotional reactions".[70] When people see a post and experience anger, fear, or outrage, they share it with others, not necessarily because they agree with the content, but because the content has captured their attention.[71] Algorithms interpret these interactions as signals of interest and preference.

When it comes to viral content, there appears to be an asymmetrical power basis created in the digital information landscape between emotion-inducing content compared to content that is not. Extremist, hateful, violent, prejudicial and intolerant views may receive greater amplification due to this content potentially being anxiety and anger-evoking. This is not to suggest that positive content that is tolerant, inclusive, compassionate, and respectful does not go viral. Positive, awe-inducing content does go viral; in fact, some studies suggest that its virality is higher than negative (i.e., anxiety and anger-evoking) content.[72]

Berger and Milkman acknowledge that although positive content may go more viral. Virality is influenced by more than content being positive or negative. Their study suggests that "content that evokes more anxiety or anger is actually more viral".[73] Their research reaffirms their hypothesis: feelings of arousal inform the social spreading of information. According to their study:

> "... a one-standard-deviation increase in the amount of anger an article evokes increases the odds that it will make the most e-mailed list by 34% (Table 4, Model 4). This increase is equivalent to spending an additional 2.9 hours as the lead story on the New York Times website, which is nearly four times the average number of hours articles spend in that position. Similarly, a one-standard-deviation increase in awe increases the odds of making the most e-mailed list by 30%".[74]

Given this empirical research, it seems likely that sensational, controversial, polarising, extremist, hateful, violent, prejudicial, and intolerant content would have a higher probability of going viral due to evoking feelings of arousal. If this were the case, a potential feedback loop could form, where people are continuously exposed to more content that is emotionally charged, violent, and misleading, as opposed to content that is awe-evoking.

Considering that 98% of Meta's revenue comes from advertisements, the company's business model relies on grabbing and keeping the attention of the platform's users. If social media platforms cannot keep the user's attention, advertisements as the primary source of revenue may not be a workable business model since they commodify attention. It seems possible that, given (1) the business models of Big Tech and (2) that anger-evoking content has a 34% higher engagement rate than other content, Big Tech have a vested financial interest in amplifying extremist, false and discriminatory content, as that is what brings them revenue.[75] The point of this statement is not to validate the amplification of this content but to highlight the inherent tension between maximising user engagement to increase revenue vs the need to develop strong institutions that can help support the stability of liberal societies and encourage liberal civic values such as tolerance, reasonableness, etc.

69. Berger and Milkman argue that viral content is content that provides persons with feelings of awe, anger or anxiety. *See* Jonah Berger and Katherine L. Milkman, 'What Makes Online Content Viral?' (2012) 49 *Journal of Marketing Research* 192.
70. Narayana (n66) 36.
71. Tim Wu, The Attention Merchants: The Epic Scramble to Get Inside Our Heads (Atlantic Books 2017) 1–403.
72. Berger and Milkman (n69).
73. Berger and Milkman (n69) 8.
74. Berger and Milkman (n69) 8.
75. Amnesty International (n68) 10.

5. AI as a threat to Rawlsian justice

Extremist, hateful, violent, intolerant and discriminatory values existed in liberal democracies prior to AI (as demonstrated in section 3), so why then are AI technologies such as optimisation algorithms a threat to Rawlsian justice specifically? It is precisely because extremist, hateful, violent, intolerant and discriminatory values that should be minimised in liberal democracies are in fact amplified due to the feelings of arousal that they evoke, which Big Tech is capitalising on in their ever-expanding goal to commodify attention.

Extremist, hateful, intolerant and discriminatory values are values that are sometimes associated with unreasonable doctrines. To recap, unreasonable doctrines are those that reject reasonable pluralism, the freedom and equality of all citizens, and reject the idea that society should be structured for the equal benefit of all citizens. Persons holding extremist and intolerant views may be less accepting of reasonable moral pluralism, since, at times, they often aim to impose their comprehensive doctrine on those around them and on the political system itself. This, in theory, contradicts the Rawlsian requirements of justice since comprehensive ideas of the good cannot form the foundation of political institutions. Persons promoting hateful and discriminatory values target specific groups for exclusion, oppression and violence, thus undermining the values of freedom and equality in general. Moreover, persons who promote policies that favour one group or reject distributive policies for previously oppressed groups undermine the claim that political institutions should benefit all.

Recommendation algorithms that optimise user engagement may increase citizens' exposure to persons who appear to reject reasonable pluralism, freedom, and equality of fellow citizens and oppose political systems for the equal benefit of all. This exposure, in theory, can result in people feeling a sense of fear as opposed to the feeling of trust being amplified. Both fear and trust are moral motivations. For Rawls, trust is the moral motivation that requires exercising one's sense of justice. In contrast, fear is a moral motivation that prevents citizens from exercising mutual trust and their sense of justice. If you fear a person or group, you see them as threatening your way of life. In theory, increased distrust and fear could lead to a breakdown of reciprocity. There may be no reciprocity between citizens without mutual trust. Without reciprocity and mutual trust, citizens may not be willing to develop civic friendship. In *The Monarchy of Fear* [76], Martha Nussbaum highlights how fear has infiltrated the relationship between American citizens in the political sphere. She claims that informational cascades heighten generalised fear and become more dangerous due to their proliferation on social media and the internet.[77]

Fear was the moral motivation driving distrust in the examples given in section 3. Wisconsin vs Yonder and Mozart vs Hawkins feared their traditional values would be eroded by exposing their children to liberal values. Scheffler's example of the BLM movement demonstrates the legitimate fear and anger black American citizens feel as they may see their interests (more fundamentally, their dignity) being undermined by political institutions, such as law enforcement. In contrast, Hochschild's deep story highlights the current fear of white Christian men, who may fear that their cultural values may no longer be shaping public discourse. Thus, they lament a time when their moral values and doctrine were prioritised in the political system.

The point of reiterating that these diverse groups experience different reasons for fear is to demonstrate why there may be a breakdown or an underdeveloped civic friendship. These groups may view their way of life or their freedom and equality as under threat. When persons develop fear and distrust, they could develop feelings of resentment towards their fellow citizens and the political system for not awarding them what they (1) are entitled[78] to or (2) what they think they are entitled to.[79] The viral content promoted on social media by engagement optimisation algorithms appears to contribute to a negative arousal feedback loop. By 'negative arousal feedback loop' I mean, when citizens see anger-evoking content online, they are

76. Martha C. Nussbaum, The Monarchy of Fear: A Philosopher Looks at Our Political Crisis (Simon & Schuster 2018) 1-272.
77. Nussbaum (176) 50.
78. For example, in the case of black Americans being legitimately entitled to the same fair and equal treatment by law enforcement as their white American citizens.
79. In the case of Hochschild's deep story, the potential perceived entitlement that some white Christian men and women have for their cultural values to shape public discourse.

more likely to (1) share this content and (2) see this content longer, as anger-evoking content has a higher potential to consume the top of one's feed. The exposure to it and the desire to share anger-evoking content may potentially fuel distrust, fear, anger and resentment both on and offline. This negative arousal feedback loop could help to contribute to moments such as the Charlottesville march in 2017.[80]

According to Rawls, "prejudice, bias, self and group interest, blindness and wilfulness" are the sources for unreasonable disagreement.[81] As Big Tech companies use recommendation algorithms to optimise engagement, due to their financial incentive, they promote content (arousal-inducing content) that, in theory, appears to help foster the sources of unreasonable disagreement. These six sources of unreasonable disagreement contribute to the breakdown of a reasonable knowledge landscape since they are not publicly justifiable forms of disagreement. For example, a citizen supporting an anti-immigration policy, basing this justification for this policy on xenophobic prejudice, is not publicly justifiable to other citizens. This is the case since prejudice undermines the core moral values of liberty and equality of persons. Suppose such citizens further entrench themselves in their initial belief instead of weighing alternative evidence. In that case, those same citizens would not exercise the political virtue of reasonableness.

Rawls claimed unreasonable comprehensive doctrines did not threaten the stability of society if kept to the fringe groups. Likewise, unreasonable disagreement poses less of a threat to stability if it is limited. Without this limitation, unreasonable doctrines and unreasonable forms of disagreement may pose a danger to the reasonable knowledge landscape that democracies need to sustain trust in society. When optimising engagement, Big Tech could be using algorithms to do precisely that, to make views that are unreasonable in the Rawlsian sense go viral, for the goal of commodifying the attention of users.

This potential virality of unreasonable disagreement that promotes hateful and intolerant views threatens the normative foundation for liberal democratic societies. If persons are constantly exposed to content that demonstrates how their fellow citizens uphold values that threaten their freedom or equality, then this may threaten even those who once bought into the system of reasonable justification and political values. If persons who uphold the political values of tolerance, civility, reasonableness, fairness, and willingness to propose and adhere to the fair terms of social cooperation, see their fellow citizens not upholding the same commitment to a democratic society this could lead to these very citizens to reject the political values they once supported, for fear that they need to protect their interests.

In theory, if citizens start to reject the liberal values and conditions of a reasonable informational landscape, this could threaten the possibility of achieving an overlapping consensus in the United States. Empirical research needs to be undertaken to determine the impact of engagement optimisation algorithms on citizens' willingness to disagree with the constitutional essentials. If there is a high level of disagreement, this could indicate an unravelling of the overlapping consensus, as there is no longer a minimal buy-in to the core liberal values. Extensive sociological research would need to be done to substantiate this claim, but theoretically, if the suppositions above are correct, and empirical research does substantiate the fact that arousal-evoking content does go viral, then the stability of liberal democracy in the USA may be deeply threatened. It may also turn out that there was never an overlapping consensus in the first place, but only a modus vivendi.

6. Conclusion

In conclusion, engagement optimisation threatens Rawlsian justice because optimisation algorithms seem to skew the information landscape in liberal democracies insofar as content that goes viral is content that evokes awe, anger, or anxiety. This skewed informational landscape can potentially encourage more moral

80. Empirical research would need to be undertaken to substantiate the casual relationship between arousal inducing content and its effects to undermine moral motivations essential for one's sense of justice.

81. Rawls (n7) 58. Note, I am not suggesting that BLM is based on unreasonable forms of disagreement, quite the opposite, BLM protestors appeal to liberal values of liberty and equality to highlight how the American law enforcement system operates unequally, reducing their freedom.

disagreements between citizens, escalating into perceived existential battles, given the negative arousal feedback loop that could occur. Citizens may view one another then as moral enemies to be defeated rather than fellow citizens to be engaged with. This would undermine the possibility of constructive dialogue and compromise. The possible implication of a disrupted informational landscape could be the erosion of the foundations of a democratic society where conflicting moral views struggle to coexist and persons fail to reconcile their conflicting views through reasoned debate and mutual respect. Thus, it seems plausible that a sense of justice may not be adequately achieved in this information environment.

As I demonstrated in the historical examples of section 3, extremist, intolerant, and discriminatory views are not a result of AI technology but rather amplify their danger for liberal constitutional democratic systems. Engagement optimisation algorithms may amplify distrust, fear, anger, and resentment precisely because the content that elicits this emotional reaction will go viral. The implication of this argument is that it raises questions on the nature of Big Tech's business model and its alignment (or lack of) with democratic aims. Engagement optimisation algorithms amplify arousal-evoking content in order to keep the user's engagement on their platforms in the name of financial benefit for the elite few. Thus, these companies prioritise user engagement over developing companies that are firmly committed to fostering a stable liberal democratic society for the benefit of all citizens.

In *Political Liberalism* (2005), Rawls states: "that there are doctrines that reject one or more democratic freedoms is itself a permanent fact of life or seems so. This gives us the practical task of containing them—like war and disease—so they do not overturn political justice".[82] Just as Rawls suggests, unreasonable doctrines should be contained, so should extremist, hateful, intolerant, and discriminatory viral content.

The pertinent question is, what recommendations can be implemented to reduce the risk of anger-evoking viral content from building distrust and resentment in liberal democracies? This is an essential question for future interdisciplinary empirical inquiry, undertaken by AI industry experts, anthropologists, sociologists, psychologists, computer and data scientists, legal scholars, policy makers, as well as philosophers. This interdisciplinary collaboration would help to achieve rigorous, verifiable, empirical knowledge, which may support normative implications, such as those made in this paper, thereby helping to ensure that AI serves democracy as opposed to having the potential to undermine it.

7. Acknowledgements

I want to thank Veli Mitova and Karen Frost-Arnold for their comments on an earlier version of the paper. Thanks to the audience's comments and suggestions from the TILTing Perspectives 2024 Conference. I gratefully acknowledge the financial support provided by the University of Johannesburg and Tilburg University for conference attendance, which contributed to the development of this article. This article is part of research conducted under the GES 4.0 SI-SDG Project, 'Safeguarding Democracy in the Age of AI', funded by the University of Johannesburg under grant number 105591.

82. Rawls (n7) 64-65.