



The Harm of Social Media to Public Reason

Paige Benton¹ · Michael W. Schmidt²

Accepted: 27 August 2024
© The Author(s) 2024

Abstract

It is commonly agreed that so-called echo chambers and epistemic bubbles, associated with social media, are detrimental to liberal democracies. Drawing on John Rawls's political liberalism, we offer a novel explanation of why social media platforms amplifying echo chambers and epistemic bubbles are likely contributing to the violation of the democratic norms connected to the ideal of public reason. These norms are clarified with reference to the method of (full) reflective equilibrium, which we argue should be cultivated as a civic virtue in the context of political justification. The paper demonstrates how epistemic bubbles and echo chambers are likely to be detrimental to the disposition to follow the method of reflective equilibrium in the political context. Based on this diagnosis the paper highlights the need for interventions that help cultivate reflective equilibrium as a civic virtue and the need for corresponding interdisciplinary research.

Keywords Reflective equilibrium · Civic virtue · John Rawls · Echo chambers · Epistemic bubbles

1 Introduction: Social Media and Democracy – A Rawlsian Perspective

Social media has huge potential to offer diverse channels of communication that overcome distances in space, time, and social stratification, and thus might foster mutual understanding and social cohesion between persons. Yet, it is commonly agreed that social media platforms currently have effects that are detrimental to liberal democracies.¹ One reason for this assessment is the claim that social media has amplified the presence of epistemic bubbles and echo chambers, adding to belief polarization, and fuelling moral and political ignorance, hatred, and hostility. Thus, resulting in a breakdown of communication between citizens from

diverse social, economic and political backgrounds (El-Bermawy 2016).

Drawing on John Rawls's political liberalism, we offer an additional explanation of why social media platforms amplifying epistemic bubbles and echo chambers indeed can be seen as contributing to the violation of fundamental norms of liberal democracies. These violated norms are connected to public reason liberalism in general. 'Public reason liberalism' refers to a special kind of liberalism that accepts the ideal of 'public reason'.² The ideal of public reason requires that proposals for the arrangement of the basic structure of society have to be justified to every other reasonable citizen.³ This justification must meet some standard. According to Rawls, for example, the arrangement

¹ For discussion see: (Ball 2021; Pham, Rubel, and Castro 2022; Rini 2019; Introna and Nissenbaum 2000).

✉ Paige Benton
pbenton@uj.ac.za
Michael W. Schmidt
michael.schmidt@kit.edu

¹ University of Johannesburg, Gauteng, South Africa

² ITAS, Karlsruhe Institute of Technology, Karlsruhe, Germany

² There are many forms of public reason liberalism and proponents include Jonathan Quong, Gerald Gaus, John Rawls or Jürgen Habermas (see Rawls 2005; Gaus 2015; Quong 2022; Habermas 2023).

³ Reasonable in the context of this paper is a technical term used in the literature on public reason liberalism. Here are two examples of definitions from that literature to clarify the term. According to Krasnoff (2014, 693) "A reasonable person is a person who sincerely desires to identify, propose, and act on principles that all other such persons could likewise accept". For Quong (2011, 291) reasonable persons adhere to one of the three conditions: (1) They acknowledge reasonable moral pluralism. (2) They regard all citizens as free and equal. (3) A just political system must be a fair and open system for the mutual benefit of all. Thus 'reasonable persons' are those that either do not adhere to one of the three conditions or do not abide by the burdens of judgement and the need to build up a justification on a

of the basic structure of society must be justifiable via the method of reflective equilibrium (RE) (Rawls 2001, 31). Consequently, Rawls requires citizens to employ a version of the method of RE in public political deliberation, where the justification is based on beliefs shared in an overlapping consensus (Rawls 2005, 159–67). It is the corresponding disposition to use this public version of the method of reflective equilibrium for public deliberation that we argue is a civic virtue. Moreover, we argue that it is exactly the abilities and motivations connected with the civic virtue of reflective equilibrium that are negatively affected by epistemic bubbles and echo chambers.⁴ We follow Thi Nguyen's (2020, 142–43) distinction between epistemic bubbles⁵ and echo chambers to describe the social, epistemic, and technological phenomena impacting belief formation.⁶

This paper contributes to the intersection of two core sub-disciplines of philosophy namely, political philosophy and epistemology (more specifically the social epistemology of social media and political epistemology). This paper contributes to the literature at the intersection of these sub-disciplines since we propose: (1) A political framework in which to analyse epistemic phenomena (i.e., epistemic bubbles and echo chambers). (2) Demonstrate how and why these epistemic phenomena could impact the development of political understanding and the possibility of public deliberation and participation necessary for public reason to be achieved. (3) Offer potential solutions to both epistemic and political

basis of beliefs shared in an overlapping consensus for political decisions one is disposed to act upon.

⁴ This paper does not assume that public reason liberalism or the method of reflective equilibrium is accepted generally by all. This paper grounds the claim that the disposition to follow RE should be accepted as civic virtue insofar as one commits to the normative core claim of public reason liberalism, namely that political action should be justifiable to all reasonable citizens.

⁵ In the current debate, there is a wider use of filter bubbles than epistemic bubbles, see: (Introna and Nissenbaum 2000), (Hannák et al. 2017), (Dillahunt et al. (2015). Eli Pariser (2012) defines the term filter bubble to refer to how algorithms on digital platforms limit, and filter the information persons using these platforms experience. In contrast, for Nguyen (2020, 144) and, Miller and Record (2013, 119), epistemic bubbles include filter bubbles as described, but the term extends to account for the mediation of information via 'non-technical' methods. For example, person A chooses to sit with person B instead of person C for lunch in the office because person A and B share the same set of moral beliefs. Epistemic bubbles distinction is attractive precisely because it can account for the epistemic harm of mediated beliefs on both technological platforms and in non-technological settings. For the purpose of this paper, we will only address how epistemic bubbles and echo chambers used on technological platforms undermine the virtue of reflective equilibrium.

⁶ For a discussion of the epistemic dimensions of filter bubbles and its impact on belief justification see: 'Justified belief in a digital age: on the epistemic implications of secret internet technologies' by (Miller and Record 2013).

problems facing liberal constitutional democracies today such as the decline of epistemic and political trust.

It is important to note, that within these sub-disciplines there is a debate on the causal connection between epistemic bubbles, echo chambers and belief polarization. There is no consensus if belief polarization exists and is prior to epistemic bubbles and echo chambers, or if belief polarization is a result of epistemic bubbles and echo chambers.⁷ This paper does aim to stay agnostic with respect to this debate and only presupposes that current social media platforms at least significantly amplify epistemic bubbles and echo chambers. Moreover, our thesis is not concerned with belief polarization which, indeed, might be reasonable in specific circumstances. Belief polarization does not necessarily entail social instability or infringement of political autonomy in a society based on the principles of public reason liberalism as long as the essential elements of the overlapping consensus are not impaired by the polarization. Whether or not epistemic bubbles and echo chambers lead to belief polarization is not our concern here. Rather our concern is that they are likely to contribute to the violation of basic norms of liberal democracies with respect to the ideal of public reason. Insofar as epistemic bubbles and echo chambers prevent reasonable debates between a diversity of persons and their epistemic beliefs, we claim that epistemic bubbles and echo chambers do pose a threat to liberal democracy, regardless if they are responsible or not for belief polarization, since they pose a threat to reasonable agreement. This argumentative positioning adds to the novelty of our approach to show that the harm of social media can be framed with respect to concepts other than belief polarization, such as with respect to civic epistemic virtues.

To demonstrate how epistemic bubbles and echo chambers undermine the conditions of Rawlsian public reason, the paper is laid out as follows. In Sect. 2, we first provide an exposition of the method of RE and the norms of public reason. We then, demonstrate how and why exhibiting the disposition to follow RE in the political context should be seen as a civic virtue. In Sect. 3 we address how epistemic bubbles and echo chambers impact public reasoning in terms of undermining the civic virtue of RE. In Sect. 4 we highlight the need for interventions for fostering the civic virtue of RE by exploring different possibilities. Section 5 explores whether a prevailing strong civic virtue of RE can significantly contribute to the effective prevention of filter bubbles and echo chambers. We summarize our argument in Sect. 6.

⁷ For discussion on this literature broadly construed see: (Sunstein 2018, Lewandowsky et al. 2012; Zuiderveen Borgesius et al. 2016; Bail et al. 2018; Dubois and Blank 2018)

2 Employing the Method of (full) Reflective Equilibrium as a Civic Virtue

2.1 An Analysis of Reflective Equilibrium with Reference to Four Core Rules of the Method

RE is commonly understood as a method of justification. It was popularized by John Rawls in *A Theory of Justice* and his subsequent works (Rawls 1974, 1999, 2001, 2005). RE aims to achieve beliefs (or other doxastic states) that are justified for an epistemic agent. However, there is a danger of misunderstanding since it is not entirely clear what kind of justification RE aims at. It is not the kind of justification that figures prominently in epistemology.⁸ What RE aims at is a kind of justification which is mostly interpreted as not necessarily securing a reliable connection to truth – it is not (in conjunction with true beliefs) sufficient for knowledge. RE aims at a kind of justification that is seen as a form of rational entitlement to believe only in light of one's own beliefs or with regard to beliefs that are shared within a given debate, group or society (Schmidt 2024; see also Kauppinen and Hirvelä forthcoming). A prominent interpretation is that this is directed towards the epistemic achievement of understanding (Elgin 1996, 2017; Baumberger and Brun 2020; Beisbart and Brun 2024)

The basic idea is the following: Beliefs are justified in this internal or dialectical sense *if and only if* they are part of a RE state. Epistemic agents (whether natural persons or collective agents, such as courts) achieve a RE state if they succeed in forming their beliefs into a sufficiently wide and harmonious system of beliefs, they reflectively judge as the most plausible. In order to arrive at a RE, one should scrutinize one's beliefs in an integral way and mutually adjust theoretical principles and judgements of all levels of generality. So, one can make a distinction between the state of RE and the method of RE that aims at achieving this state. However, state and method are closely related since the method is the only – although fallible – way of reaching a RE state (Schmidt 2024). A coherent state achieved by some kind of brainwashing process presumably would not count as a RE.

We build upon the position that following the method of RE consists in following four interconnected rules (Schmidt 2024, 170):

1. *Minimalistic Foundationalism:*

Justification via MRE is tied to the epistemic agents' own beliefs and evaluations. Epistemic agents may (only) include beliefs in the RE process that they themselves happen to hold. They are also entitled to include

hypotheses or theories they deem to be relevant or worth considering.

2. *Minimalistic Fallibilism:*

The epistemic agent should consistently treat all beliefs that enter the RE process or result from it as defeasible and revisable.

3. *Moderate Holism:*

a) In the RE process epistemic agents must consider all beliefs, theories and hypotheses—including their inferential relations—that they would deem to be relevant for the current inquiry after due reflection. They are relevant if and only if their inclusion would foreseeably alter the core result of the process. Considerations regarding work and time constraints can justify a deviation from the ideal of identifying all relevant beliefs, theories and hypotheses.

b) Conflicting beliefs, theories and hypotheses are not balanced against each other in isolation, but as part of possible systems of belief, which result from alternative plausible adjustments. These systems of belief—which are candidates for a RE state—are evaluated as a whole.

4. *Minimalistic Rationality:*

Epistemic agents should choose the candidate system of beliefs which, as a whole, exhibits the highest plausibility—in light of all inferential relations and the strength of the agent's beliefs. If the epistemic agents succeed in adjusting their beliefs accordingly, they have achieved a RE."

2.2 Justification in the Political Context: Reasonable Pluralism

The method of RE must be complemented in a specific way – following Rawls – if one aims at a justification that is adequate in the political context of liberal democracies. To see why this is the case, one can ask the question: Whose beliefs should be taken into account while employing RE in the political sphere? This question becomes urgent if one accepts the normative claim that principles of justice must be justifiable to all reasonable citizens.⁹

⁹ We will not defend this core claim of public reason liberalism in the present paper and do not have a more compelling defence than the ones which were already offered. However, we think that the arguments in favour of this claim with regard to political autonomy and stability are sound. In case one does not agree, our findings might be only interesting from a theoretical perspective, otherwise, they have also practical implications (see also Sect. 1).

⁸ Justification is mostly associated with the analysis of knowledge as "justified true belief" (see Gettier 1963; Ichikawa and Steup 2018).

Prima facie, it is not clear that every reasonable citizen can easily be included in a specific public RE process. There are what Rawls calls the “burdens of judgement” (Rawls 2005, 54–58). In some domains, especially in the moral domain, these burdens lead to deep disagreement, where the disagreement is not based on irrationality but on adequate thought processes and adjustments. When freedom of thought and speech is granted, especially in liberal democracies, we should thus expect not only pluralism in matters of morality but reasonable pluralism of different comprehensive moral doctrines, i.e., philosophies, worldviews, and religions that provide answers to basic moral questions or promote a conception of the good life. Rawls calls this the fact of reasonable pluralism (Rawls 2005, 144). With respect to the normative claim that principles of justice or basic political regulation must be justifiable to all reasonable citizens, we might also call it the problem of reasonable pluralism since a RE process that is successful for one person might not yield a RE for another person with different moral beliefs. Since these disagreements on moral beliefs can be reasonable, if they are caused by the burdens of judgement and not some irrational adjustments,¹⁰ and one accepts fallibilism one has a duty of civility or tolerance towards those opposing views.¹¹ This includes a kind of epistemic humility since one acknowledges that one’s position might need revision and there is no certain moral superiority regarding one’s position. Given that no person can legitimately claim definitive superiority for their moral beliefs there is a need for a method of public justification that is not relying on a specific comprehensive moral doctrine.

Rawls provides a solution for the problem of reasonable pluralism: If we base our justification concerning the

principles of justice and basic political regulation on an overlapping consensus, on which all reasonable comprehensive moral doctrines agree despite their incompatible parts, a public justification is appealing to every reasonable citizen. This consensus should include the modes of public reasoning and justification itself. However, this leads to a modification of RE in the political context due to a necessary reformulation of the rule of minimalistic foundationalism (insofar as one accepts the proposed rule-based analysis of RE presented in subsection 2.1).

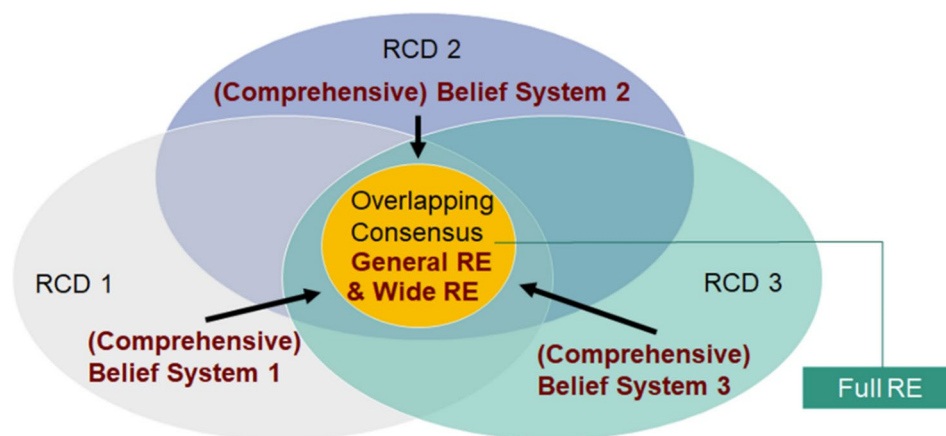
Minimalistic Foundationalism in Political Contexts

Justification in liberal democracies is tied to the beliefs and evaluations which are shared by all reasonable citizens in an overlapping consensus. Epistemic agents may (only) include beliefs in a public RE process that they in fact do share with other reasonable citizens. They are also entitled to include hypotheses or theories that all reasonable citizens deem to be relevant or worth considering.

If one achieves a RE that is based on such an overlapping consensus in the political context and this RE is sufficiently wide and accepted generally as such by all reasonable citizens, one achieves what Rawls calls “full RE” (Rawls 1995, p. 141; 2001, 31 f.; Daniels 1996, 144–75; 2020) (see Fig. 1).

Important to note is that marginalized groups, cannot simply be excluded or considered unreasonable just because one does not share their epistemic standpoint. However, some beliefs from marginalized epistemic standpoints might not be initially shared beliefs, so with respect to the proposed

Fig. 1 Scheme of a wide, general and (thus) full RE in a liberal democracy with three reasonable comprehensive moral doctrines (RCD)



¹⁰ ‘Irrational adjustments’ for Rawls would be adjustments to one’s belief system as a result of judgements made under conditions of unreasonable disagreement. “Prejudice and bias, self- and group interest, blindness and wilfulness” are the six sources of unreasonable disagreement (Rawls 2005, p. 58).

¹¹ The duty of civility has, besides the justificatory dimension, also other dimensions see: (Bardon et al. 2023, 2024).

interpretation of the method of full RE, there is a complication to this statement. This complication, or so we argue, can be solved if there is agreement that different epistemic standpoints imply domain knowledge that is important for political decisions and must be included in the quest for full RE by testimony. One can interpret Rawls's burdens of judgement to include this kind of epistemic humility and inclusivity (see especially consideration d) in Blake 2014, 75). Thus, a diversity of epistemic standpoints should be considered when pursuing full RE.

2.3 The Disposition to Follow the Method of (full) Reflective Equilibrium as a Civic Virtue

(Full) RE itself as an adequate method of public justification must be part of the overlapping consensus of a well-ordered (liberal democratic) society, if it is an obligation to justify principles of justice or proposals of basic political regulation via RE to all reasonable citizens. Accordingly, citizens have to develop the disposition to use RE as a method to justify their political claims to their fellow citizens with the aim of a full RE. Political decisions should be based on a shared justification wherever possible, but especially on basic matters. When the overlapping consensus is too insubstantial to allow for a shared justification, at least the mode of procedural solution (e.g., a majority vote in parliament or court) should be based on a shared justification (see Brändle and Schmidt 2021, 1491–93).

To safeguard the political autonomy of citizens and the stability of liberal democracies, we argue that RE needs to be cultivated as a civic virtue. Rawls claims that there are necessary political virtues that citizens need to exercise in the public sphere, and in our case in the digital sphere when exercising their sense of justice. The political virtues for Rawls are the “[...] virtues of fair social cooperation such as the virtues of civility and tolerance, of reasonableness and the sense of fairness.” (Rawls 2005, 194; see also Boettcher 2014) Citizens have to actively exercise these political virtues when participating in public life in order for the advancement of a stable constitutional liberal democratic society. These political virtues are the vital capacities of citizenship, without these capacities, citizens would not be able or willing to engage in cooperation on free and equal terms. We argue that the four rules of RE – adjusted minimalistic foundationalism, minimalistic fallibilism, moderate holism, and minimalistic rationality – provide a rule-based analysis of these capacities insofar as they are concerned with public justification.

The disposition to follow the method of (full) RE needs to be recognised as a civic virtue insofar as this is exercising key justificatory capacities necessary for citizenship. Developing the disposition to follow RE including the necessary

abilities or competencies alongside the motivation to use them with the aim of a full RE, of course, involves some training, implicit or explicit. One can succeed to a higher or lower degree in acquiring the disposition to follow the method of (full) RE. If one agrees that it is a public good to acquire this disposition to a sufficient degree, then one might reasonably regard this disposition as a civic virtue. Thus, we speak of RE as civic virtue or the civic virtue of RE, in short, from now on.

2.4 The Civic Virtue of RE in the Wider Context of Rawlsian Public Reason Liberalism

At this point, we would like to highlight the connections of core concepts of Rawls's theory with our interpretation of RE as a civic virtue. An essential aspect that ensures political autonomy and the stability of an overlapping consensus – and a liberal society itself – is for citizens to be able to develop and assert a reasonable justification for any position they wish to advance in the public sphere. As stated previously, this ability for reasonable justification is connected with the citizens' duty of civility. Citizens exercise their 'duty of civility' when appealing to public reason to justify policies, legislation, and distribution of goods. Rawls states clearly that a requirement for any democratic citizen is learning how to reason in such a manner that they can explain and justify their choice or action on public terms that other citizens could find reasonable to accept (Rawls 2005, 460). These public terms are an appeal to liberty, equality, and equal opportunity as primary political principles. When citizens exercise this duty, they are acknowledging to one another that there are limits to their comprehensive doctrines and a specific space they can occupy. Specifically, they acknowledge that it is unjust for them to try and control the mechanisms of the state to assert their comprehensive doctrine (Rawls 2005, 460). When citizens realise this and uphold their duty of civility, they help to safeguard the democratic institutions themselves. For example, the voting process would then be protected from citizens advancing their own preferences. Citizens would also protect themselves from an overthrow of majority rule, and the threat of a comprehensive doctrine influencing the foundation of state law (Rawls 2005, 219). Thus, the duty of civility is closely connected with the adjusted rule of minimalistic foundationalism as part of RE as a political virtue. When exercised, citizens who hold diverse moral views are able to disagree about these views, while at the same time arriving at reasonable discussion in the public sphere.

Without the prevalence of the civic virtue of RE, citizens would also not be able to be politically or epistemically autonomous. We interpret Rawls's idea that justice demands citizens to be free and equal, including this epistemic

dimension. As we demonstrate in the following section, epistemic bubbles and echo chambers undermine the political and epistemic autonomy of citizens in digital spaces. Agents in these epistemic systems are either not exposed to external sources (i.e., epistemic bubbles) and as such have unequal access to epistemic sources, or the external sources they are exposed to are discredited via epistemic manipulation (i.e., echo chambers). Both of these phenomena help to disrupt a steady stream of contrary evidence and counterarguments. This contributes to undermining the agency of citizens insofar as they reshape digital spaces that undermine the free and equal conditions necessary for liberal democracies' legitimacy.

The legitimacy of liberal democracies depends on public justification that helps realise and express the autonomy of citizens. Reasonable disagreement between groups is expected considering liberty of conscience and reasonable moral pluralism. As mentioned earlier in this section, the burdens of judgment are the foundation of this reasonable disagreement. When deliberating political matters, persons must take into account that reasonable disagreement can be a result of analysing evidence, the interpretation of moral and political concepts and values, a person's lived experience, the appeals to various frameworks for assessment, and, lastly, the difficulty in prioritising some values at the expense of others (Rawls 2005, 56–57). These are the six sources of reasonable disagreement. In opposition to this, "prejudice and bias, self- and group interest, blindness and wilfulness" are the six sources of unreasonable disagreement (Rawls 2005, 58). Epistemic bubbles and echo chambers help to develop an unhealthy epistemic system in the digital sphere that fosters the six sources of unreasonable disagreement. Unreasonable disagreement has the potential to undermine the overlapping consensus given that unreasonable disagreement arrives from citizens prioritising their own moral convictions and ideas of the good over their commitment to principles of justice that can be accepted in light of a shared justification.

3 How Does Social Media Affect Reflective Equilibrium as a Civic Virtue?

The formation of the culture in the public sphere as well as the culture in the private sphere is influenced by what Rawls describes as the 'nonpublic political culture'. This encompasses various mediums such as newspapers, reviews, magazines, television, and radio (Rawls 1997, 768n13). Current liberal democracies' political culture includes *X* (formerly known as *Twitter*), *Facebook*, *Instagram*, *YouTube*, *TikTok*, *Reddit* and *Twitch* to name a few. Given that these digital technologies offer the communicative acts that are

necessary to reproduce the political and private culture in liberal democracies. These digital platforms serve as avenues for citizens to express both their sense of justice and their sense of the good.¹²

Citizens exercise their sense of the good on these same platforms when promoting their comprehensive moral doctrines, private moral values and their belief systems. Insofar as citizens exercise this capacity, they are exercising their non-public reason. Citizens exercise their sense of justice on social media platforms when they engage in public debate on legislative issues such as climate change policies. Exercising one's sense of justice in the digital sphere may take the form of one of the following actions: A person may promote their political position on political issues by signing online petitions, creating a political blog, live-streaming rallies, creating online campaigns, sharing, liking, commenting and reposts media and ideas of fellow users.

Digital platforms become the domain where public reason and non-public reason¹³ collide with one another as it is mostly here that citizens share their moral, economic, social, and political viewpoints. Therefore, the content shared on social media can contribute to the reinforcement of democratic political values, thus strengthening citizens' commitment to public reason. Yet, it can promote perspectives and activities that undermine these political values, and thus amplify existing social and historical biases present in democratic societies (Benton 2023). Digital platforms can facilitate large-scale discussions on social and political issues yet, at the same time, facilitate the polarisation of views that aid in fragmenting political society as opposed to enabling the social consensus necessary for democracy to flourish (Bozdog and van den Hoven 2015, 251). With the rise of fake news, there is a call for the regulation of information on media platforms to encourage fair epistemic participation in digital technologies (Smith and Niker 2021, 1–2). Due to the fact that social media platforms enable the promotion and sharing of both political and private moral values the digital sphere interconnects the public and so-called non-public spheres.

In both the public and non-public sphere, citizens should have epistemic participation as free and equal individuals.

¹² The moral power of the sense of justice accounts for the capacities citizens have to uphold and deliberate about principles of justice and rules for mutual cooperation (Rawls 2005, p. 19). The moral power of the sense of the good is defined by Rawls as the ability one has to reason from, revise, edit and follow their own chosen conception of the good (Rawls 2005, p. 19).

¹³ We use 'non-public reason' here to refer to the kind of Rawlsian reasons that citizens can provide for justification in their private lives. These reasons generally rely on arguing in favour of a moral truth claim that underlies the comprehensive conception of the good that citizen subscribes to. This reason is a non-public reason as it does not need to meet the conditions of public reason, namely the criterion of reciprocity and the burdens of judgment (Rawls 2005, 53–56).

Firstly, insofar as they participate in the digital sphere for engagement with exercising their moral power of the sense of the good individuals should have the freedom and equality to express their reasonable notion of the good as they see fit. Secondly, in terms of exercising their sense of justice, individuals are bound by the duty of civility. They should respect the norms of public reasoning and thus exhibit and cultivate the civic virtue of RE.

One of the greatest challenges to epistemic participation is the role of echo chambers and epistemic bubbles. Echo chambers and epistemic bubbles are two areas of concern in the digital sphere that we argue are likely to encourage behaviour that conflicts with and undermines the disposition to follow the rules of full RE as a method of public justification. Thus, these two features of current social media - inter alia - are likely to negatively affect the corresponding civic virtue of RE.¹⁴ Even if initially plausible, however, in the end, this is, of course, a claim that has to be investigated empirically. So, our argumentation in the next section only aims at making plausible that such an empirical inquiry is a worthy enterprise. Let us first look at epistemic bubbles.

3.1 Epistemic Bubbles

An epistemic agent is situated in an epistemic bubble - at a specific time and concerning a specific topic - *if and only if* the epistemic agent has been exposed for an extended period only to a one-sided coverage of information and arguments by omission of relevant information and arguments that provide an alternative perspective or additional evidence. We follow Nguyen in his characterisation of epistemic bubbles. To reiterate epistemic bubbles are prevalent in both the technological and non-technological spaces. They are characterised by a uniformity of information or beliefs that end up restricting the epistemic agency of the person (Nguyen 2020, 141, 144). Thus, in an epistemic bubble that exists in the digital sphere, relevant alternative voices to a conversation in digital public spaces are omitted (Nguyen 2020, 141). Consequently, individuals in epistemic bubbles exist in a flawed epistemic system since the system is narrow. By 'narrow' we mean that the kinds of information and justification for beliefs are restricted. Relevant testimony, evidence and facts could be missing due to the narrow 'bubble' of information the individual finds themselves in (Goldberg 2011, 93-94).

Epistemic bubbles can be a result of an individual's own doing, whereby a person actively engages only with a limited set of persons, those that share their respective

moral, political, philosophical, metaphysical, or religious beliefs and values (Nelson and Webster 2017). Epistemic bubbles, however, can also result from a person's rather narrow engagement with alternative views which are then externally reinforced through social media recommendations such as algorithms with opaque filtering rules (Miller and Record 2013). The filtering rules mostly aim to suggest content that, with respect to the available data, is statistically likely to keep the user engaged in the platform (Pariser 2012). In effect, these filtering algorithms often suggest in the political context content in which similar or like-minded views are expressed. This is an overlap between epistemic bubbles and filter bubbles concerning social media.

Once inside an epistemic bubble, individuals experience what Nguyen refers to as "a self-reinforcing epistemic filter". By being exposed to ideas that only reinforce their beliefs individuals develop a false sense of epistemic confidence which results in these individuals either refusing to or not seeing the need to challenge, reflect on, or revise their initial beliefs. This can help to reinforce cognitive bias and further entrench a flawed epistemic system by the perpetuation of confirmation bias (Nguyen 2020, 143-44). While algorithmic filtering reduces the probability of encountering another viewpoint, the false sense of epistemic confidence could rule out the desire to hear from the other viewpoint even if that viewpoint was available. Epistemic confidence may not always be a side effect of being in an epistemic bubble, for example, a climate change denialist may doubt her beliefs on climate change, while not interrogating her beliefs, or searching for alternative sources of information. Thus, theoretically, she may choose to stay in her "informational cocoon" (Sunstein 2007, 10-19) even if she is a fallibilist about her own beliefs. One situated within an epistemic bubble receives fewer challenging views that indicate that their opinions need elaboration or even revision. This does, of course, not entail that one's epistemic confidence will be inflated (that can be the case even if the overall confidence level is still comparably small). However not being exposed to available informational debunkers does, or so we claim, make it at least likely that one's epistemic confidence becomes inflated.

With respect to the civic virtue of RE epistemic bubbles thus are likely to undermine the motivation and ability to follow at least three of the corresponding rules, namely 'Holism', 'Fallibilism', and 'Rationality'.

- 1) 'Holism' requires that the epistemic agents take into account all arguments concerning a topic, they would deem relevant after due reflection. An epistemic agent in an epistemic bubble, however, is only exposed to a rather narrow range of information and arguments. Whether the epistemic bubble arises from the individuals'

¹⁴ At least insofar as it is plausible that continued behaviour that neglects or conflicts with abilities and motivational sets that are necessary for the disposition to follow the method of (full) RE likely affects this disposition negatively.

self-imposition or as a result of algorithmic outcomes, the ‘bubble’ is characterised by an omission of opposing epistemic channels of information. Thus, preventing the agent from receiving a diverse set of epistemic sources to test their beliefs against. An epistemic agent in such a bubble is accustomed to taking into account a rather narrow range of information and arguments and thus is likely to become disposed to stop searching for further counter-evidence rather prematurely. The disposition to act in line with the rule of holism requires agents to actively search out alternative informational sources. Persons who actively remain in ‘informational cocoons’ may have various reasons for not wanting to search out alternative informational sources, such as for their own cognitive comfort of being surrounded by informational sources and persons that reaffirm their own beliefs. Thus, agents in self-imposed epistemic bubbles most likely will not be willing to uphold this rule since it requires them to be willing to dismantle their epistemic bubble.

- 2) ‘Fallibilism’ requires that epistemic agents consider all beliefs within the RE process as defeasible and revisable. However, an epistemic agent in an epistemic bubble is not accustomed to assessing arguments contrary to their initial convictions in a critical and open-ended inquiry. As stated above, since agents in epistemic bubbles are exposed to views that reinforce their epistemic position as opposed to challenging it, they are also likely to develop a false sense of epistemic self-confidence. Thus, their ability and motivation to revise their beliefs in light of arguments are likely compromised. Being committed to the rule of fallibilism isn’t merely a matter of an abstract intellectual commitment. It involves a corresponding disposition to act that might become impaired when one’s epistemic confidence gets inflated. At least, having a disposition to follow the rule of fallibilism involves the practice of revising one’s beliefs from time to time which involves handling the stress and uneasiness associated with this task. Both agents with low confidence or inflated confidence in epistemic bubbles, in theory, may not be willing to undertake this epistemic responsibility and hence choose to remain in their ‘informational cocoon’. Thus, adhering to the rule of fallibilism involves the agent being willing to ‘pop’ their epistemic bubbles when they become aware of them. Since this becomes less easy when one is situated in an epistemic bubble for a longer period, the disposition to follow the rule properly is likely to become negatively affected.
- 3) ‘Rationality’ is likely to be negatively affected since ‘Fallibilism’ and ‘Holism’ are closely intertwined with it. If the disposition to revise one’s beliefs and to

consider serious alternative accounts is affected, the disposition to choose the belief system which is on reflection the most plausible with respect to all relevant alternative belief systems is likely affected as well. In all these cases we assume that a disposition grows and shrinks with corresponding habituation or lack thereof.

Epistemic bubbles do not appear to undermine the motivation and ability to follow the other rule of full RE namely ‘refined minimalistic foundationalism’. An epistemic agent in a flawed epistemic network such as an epistemic bubble can still adhere to the condition that public justification is a result of political convictions, which they think are shared by reasonable citizens and thus respect the burdens of judgement with respect to comprehensive moral doctrines. However, even if the diversity of moral worldviews is still apparent in a liberal society, the disposition to acknowledge reasonable pluralism with regard to some political beliefs might be affected as well in the end. Although epistemic bubbles are a danger to public reason, there are in principle easy ways to dissolve them, e.g., by exposure to omitted sources over an extended period (see also Sect. 5). Agents in echo chambers, however, have less chance of expanding their epistemic circle due to the epistemic distrust inherent to the system.

3.2 Echo Chambers

Epistemic agents are situated in echo chambers, *if and only if* they are members of a group, whereby the group itself is enclosed in the following way: The views of the group (especially the opinions of its leaders) are reinforced whereas all perspectives and arguments contrary to the beliefs propagated within the group are deliberately discredited in an inadequate way (Nguyen 2020, 147; Jamieson and Cappella 2008, 75–90). With this definition in mind, and in light of RE as a civic virtue we claim that in echo chambers perspectives and arguments are discredited inadequately if they cannot be justified in light of the epistemic standard given by the method of RE. Being situated in an echo chamber results in agents being overly reliant on only the epistemic sources the group deems ‘trustworthy’. The propagation and maintenance of distrust towards outsiders in echo chambers mirrors the epistemic systems created by cults, whereby the authority of outside epistemic views are undermined (Nguyen 2020, 147). Agents in epistemic bubbles typically do not actively resist or distrust outside epistemic sources, whereas echo chambers are epistemic systems that can account for how some citizens in liberal democracies have clear resistance to accepting strong arguments for otherwise plausible (and rather consensual) positions, such as in the case of climate change denialists (Nguyen 2020, 142).

Agents in echo chambers are taught to distrust other epistemic sources, which is achieved mainly by undermining the credentials of epistemic outsiders and overinflating their own credentials. The result of this is the creation of a false dichotomy of epistemic beliefs based on an ‘us vs. them’ narrative. This epistemic system, based on an inadequate justification of distrust, is self-reinforcing as long as epistemic outsiders are distrusted, and the agents in the echo chambers reaffirm the superiority of their beliefs. In echo chambers, persons’ epistemic freedom is limited if not virtually non-existent for the agent, given the limited resources for the critical assessment of the dogmas within the echo chamber (Coeckelbergh 2023, 1347). This active discrediting becomes the ‘epistemic wall’ surrounding the echo chamber metaphorically isolating members and non-members. Thus, widening the epistemic circle by exposure to alternative viewpoints can help dissolve an epistemic bubble but does not help collapse the echo chamber. If agents in an echo chamber are provided with epistemic predictions of how non-members will react to them, and non-members react as predicted, then the non-members end up helping reinforce the epistemic trust of the echo chamber by verifying the predictions within the echo chamber. Nguyen refers to this verification as the “disagreement-reinforcement mechanism” (Nguyen 2020, 147).

The unhealthy epistemic system of echo chambers is not particular to social media, this system has been used in various social groups that wish to isolate its members from non-members, such as cults (ibid.). However, the danger of this is amplified via algorithmic technologies due to the increase in scale and availability of invitations to join such echo chambers. Given this unhealthy epistemic participation, the disposition to follow all of the rules of full RE that should be cultivated as a civic virtue are undermined by echo chambers. Moreover, they are undermined in a rather different way than in the case of filter bubbles:

- 1) ‘Holism’: The disposition and ability to follow the rule of holism is likely to be affected since here, just like with epistemic bubbles, the epistemic agents in echo chambers are getting used to a narrow consideration of information and arguments. However, it is not the case that alternative and conflicting views are simply omitted – they are taken into account (in a certain sense) insofar as they are discredited. The ‘narrowness’ here is restricted to the reasoning concerning the discrediting of the beliefs and arguments of epistemic agents outside of the echo chamber and is thus subtler but poses more of a threat to the rule of holism, due to the danger of the “disagreement-reinforcement mechanism”.
- 2) ‘Fallibilism’: Similarly, the disposition and ability to follow the rule of fallibilism is likely to be affected

by echo chambers. Although epistemic agents in echo chambers are accustomed to revising their beliefs, these revisions, however, are superficial and one-sided, due to their falsely inflated epistemic confidence. Agents inside echo chambers do not consider the views of experts or epistemic authorities duly, since they distrust these from the beginning without assessing their track record or epistemic situatedness. As a result, they do not regard these alternative outside opinions as having any credibility. Agents trapped in echo chambers (e.g., with a conspiracy theory that discredits all outside perspectives) virtually always privilege the beliefs shared in the echo chamber over other beliefs that turn out to conflict with them. Thus, members of echo chambers are disposed to behaviour contrary to ‘fallibilism’ where every type of belief is treated as defeasible and revisable.

- 3) ‘Rationality’ again is affected, insofar as ‘Holism’ and ‘Fallibilism’ are affected. A further issue might be that the desire to be accepted by the in-group of the echo chamber and corresponding power dynamics leads to the custom of accepting belief systems one does not, in the end, find the most plausible on reflection. From a purely epistemic perspective, this might be conceived as the wrong kind of reason. Moreover, this might be conceived to be irrational, all things considered, insofar as one cannot internally justify adequately that being part of the ingroup is something desirable all things considered.
- 4) ‘Adjusted minimalistic foundationalism’: Where epistemic bubbles did not impact the disposition to follow the adjusted rule of minimalistic foundationalism, echo chambers likely affect members’ commitment to this rule. To recap, the rule in a political context requires epistemic agents to base their justification on beliefs that are shared with fellow citizens in an overlapping consensus. However, epistemic agents in echo chambers are systematically discouraged from looking for shared beliefs with fellow citizens outside the echo chamber, since fellow citizens are discredited a priori based on distrust of outsiders. Agents in echo chambers develop epistemic vices such as close-mindedness, due to the false sense of trust they experience for the beliefs within their echo chamber (Battaly 2018). For instance, climate change denialists are unlikely to widen their beliefs beyond their echo chamber since the unhealthy epistemic system they are in also erodes the trust these agents have in institutional structures of society, such as ‘mainstream media’ or ‘the scientific community’ (Nguyen 2020, 149).

In a healthy epistemic system, agents must rely on institutional trust structures (i.e., appeal to responsible journalists,

scientific journals, peer review journals etc.) whereby an agent can trust evidence even though the agent did not gather the information, the agent trusts the sources from where the information derives its credibility from (Hardwig 1985). Echo chambers, help to break down the institutional trust agents have and this poses an epistemic threat to public reason since public spheres of discussion become “digital islands of isolation that are drifting further apart each day” (El-Bermawy 2016). Echo chambers undermine public reason in that their existence breaks down sympathetic engagement with citizens who hold alternative moral, economic, and political beliefs while actively searching for common ground (Kinkead and Douglas 2020, 127). Thus, people outside echo chambers who lose trust in those occupying echo chambers in fact undermine the aims of public reason. The basic normative ideas of public reason liberalism require citizens in the public sphere to develop a shared public justification of political actions, the epistemic achievement. Then agents losing trust in others impacts the ability for this achievement of reasonable public deliberation to occur. If truth were the epistemic goal for public reason liberalism, then it might be a good thing to simply lose trust in people in echo chambers, but this is not an adequate strategy if one tries to achieve a public justification.

Moreover, echo chambers also pose a threat to epistemic agents outside echo chambers – they also likely affect the disposition to follow the adjusted rule of minimalistic fallibilism of reasonable citizens negatively: Epistemic reasonable citizens justifiably distrust some or even many beliefs of epistemic agents who are unfortunately situated in echo chambers. However, in doing so they might also implicitly accept the ‘us vs. them’ narrative that typically discourages them from arguing with these unfortunate citizens at all and to look for beliefs that are still shared with agents situated in echo chambers, thus also compromising their civic virtue of RE and duty to civility to a certain degree.

In contrast to the above discussion highlighting the negative impacts of echo chambers, can there not be some subset of enclosed epistemic spaces to be perceived as epistemically beneficial, such as Lackey’s argument for echo chambers (2021) and Furman’s argument for epistemic bunkers (2023)? There are two reasons for denying that enclosed epistemic spaces are epistemically beneficial for RE in the context we focus on.

Firstly, Lackey (2021, 207) defines echo chambers as enclosed spaces that reinforce the group’s view, the view is ‘echoed’ within the space as all other views are “either absent or drowned out”. Her definition does not include the active discrediting of outside viewpoints. Thus, Lackey and Nguyen’s definitions of echo chambers differ. Insofar as we accept Nguyen’s definition of echo chambers and its notable component the ‘disagreement-reinforcement mechanism’.

We argue that the definitional distinction impacts the comparison, as the disagreement-reinforcement mechanism is a harmful epistemic feature that is not present in Lackey’s echo chambers or Furman’s (2023) epistemic bunkers. Both Lackey and Furman account for exclusionary epistemic spaces as productive for safeguarding truth or protecting marginalized persons. However, we argue that enclosed epistemic spaces, be it Lackey’s echo chambers or Furman’s epistemic bunkers, are still harmful from a political perspective.

This brings us to the second reason. Lackey (2021, 216) states: “When I’m reading an article about the impact of climate change on wildlife, what is the benefit of clicking on a button from the perspective of a climate change denier? When I’m reading an article about the Sandy Hook school shooting, what epistemic advantage is there to also learn about the perspective of a Sandy Hook truther? Sure, I will be exposed to different views, but at the expense of something even more fundamental to democracy: truth”. When we focus on the realm of public deliberation, achieving or preserving truth is not the only epistemic goal – while we agree with the normative claim of public reason liberalism, that we should act on beliefs that are justified for ideally all reasonable agents, we cannot simply ignore other opinions for the sake of preserving or gaining true beliefs for some group of individuals. In other words: We focus on a different epistemic achievement than truth, namely a shared public justification for political action. When truth, well-being, or other values, are reasons for justifying enclosed epistemic communities or epistemic bunkers, they are still harmful to the public sphere as they prevent engagement between people that are socially, culturally, epistemically and politically dissimilar.¹⁵ Let us now turn to how we can foster RE as a civic virtue.

4 How to Foster Reflective Equilibrium as a Civic Virtue?

In the preceding sections, we highlighted normative implications of public reason liberalism and suggested that the disposition to follow the norms of public reasoning is endangered by epistemic bubbles and echo chambers which are fuelled by the current design of social media. This last claim is, of course, a claim and must be investigated empirically. What we thus provide are initially plausible hypotheses for empirical inquiry, which is urgent if one accepts the normative claims of public reason liberalism. The aim

¹⁵ It is a further research question of how to balance the individual or public benefits of Lackey’s echo chambers or Furman’s epistemic bunkers with the harm or benefit they may pose to individual or public deliberation.

of the present section is to provide such initially plausible hypotheses for the question of how to foster RE as a civic virtue, namely the corresponding disposition. There might be very different ways to foster RE as a civic virtue and thus we try to provide an overview of some of the possible measures, we deem most relevant, namely civic education in general by the state and civil society, and changes in the design of social media and its regulation. We'll elaborate on each of these in turn.

4.1 Civic Education

One essential measure for fostering the disposition to follow the four rules of RE is civic education. In *Justice as Fairness: A Restatement* Rawls acknowledges the role of education:

“Their education should also prepare them to be fully cooperating members of society and enable them to be self-supporting; it should also encourage the political virtues so that they want to honour the fair terms of social cooperation in their relations with the rest of society” (Rawls 2001, 156).

Therefore, educational institutions are required to teach children their basic liberties, and political virtues such as equality, tolerance, solidarity etc. Due to this, educational institutions run by the state or civil society would be key to teaching RE as a civic virtue. Critical thinking programs are a good example of such cultivation of RE as a civic virtue since there is a close connection between the rules of RE and guidelines of critical thinking (Schmidt 2022, 377–80; for critical thinking guidelines see Hitchcock 2022). The cultivation is not thought of as something that must be made in an explicit and theoretical way, it can be achieved also by practising the rules without consciously following them and thus RE might be cultivated as a civic virtue in any social institution through socialisation. This includes the encouragement of public debates that exhibit an adequate plurality and assessment of alternative arguments and positions.

A specific measure of civic education that is directed against the dangers of social media concerning the civic virtue of RE is to promote AI and digital literacy. This includes knowledge about the application and mechanisms of filter algorithms in social media and the available options people have to deal with them. AI literacy is an emerging field that arose from the necessity to determine the essential capacities individuals need in order to responsibly and critically engage with AI technology. There are a variety of frameworks of AI literacy however many share similar competencies such as 1: Developing a basic technological understanding of AI technologies, 2: Critically assessing the outputs of this technology and 3: Learning responsible methods for engaging in the use of this technology such as by exercising scepticism (Long and Magerko 2020; ‘Grade

Band Progression Charts’ 2021). If individuals are educated on the potential harms of these technologies, they could be better equipped to understand how they could be influenced by these harms. The development of such competencies would help raise awareness of potential harms such as epistemic bubbles and echo chambers, that one might encounter.

As discussed in Sect. 3 epistemic bubbles and echo chambers undermine the disposition to follow the rules of RE. To recap, ‘Holism’, for example, requires that epistemic agents take into account all arguments concerning a topic, they would deem relevant after due reflection. Individuals who have exposure to AI literacy and are introduced to the technological knowledge of AI would be aware of the existence of recommendation algorithms that filter their content on social media. Learning that there is a filter while being trained on the harms of filters and the methods on how to dissolve the bubble could provide the theoretical knowledge individuals need to form a basic motivation to widen their epistemic network. AI literacy frameworks developed by Long and Magerko argue for 16 competencies. Two of these are, the know-how to recognise the strengths and weaknesses of AI technology (competency number five) and the capacity to know how AI technology arrives at decisions (competency number eight) (Long and Magerko 2020). If individuals are equipped with these competencies to a sufficient degree, persons may be more open to looking at arguments that are critical of their initial position. In terms of ‘Fallibilism’, this rule requires that epistemic agents consider all beliefs within the RE process as defeasible and revisable. AI literacy has the potential to get individuals to adhere to this rule as AI literacy could help reduce the overinflated epistemic self-confidence individuals in epistemic bubbles have for their positions and help them get accustomed to the dynamics of reasonable belief changes. AI literacy could help agents in echo chambers to assess the validity of information they are confronted with, and this may weaken the echo chamber but not destroy it.

The greatest threat echo chambers pose is the “disagreement-reinforcement mechanism” (Nguyen 2020, 147). AI literacy can help weaken this mechanism by educating individuals about common manipulation methods used in this context to keep agents in echo chambers from exercising their epistemic autonomy. Especially for agents already situated in echo chambers, however, it is very difficult to become aware of their epistemic situation and to act accordingly, assisting them requires drawing upon shared beliefs. It might be beneficial if persons who know someone situated in an echo chamber try to maintain or even create personal bonds as a basis for shared beliefs, as far as this is reasonable. In general experiences with fighting against cult indoctrination via civic education might inform viable strategies

against echo chambers in the digital realm (Nguyen 2020, 158).

4.2 Design Changes for Social Media

Another domain for measures that foster RE as civic virtue directly or indirectly, by preventing harmful epistemic bubbles or echo chambers, is the design of social media itself. The current setup of social media can be improved with regard to the needs of liberal democracies.

One of these changes could be to reduce the opacity of AI technology (such as those employed in filtering algorithms). Social media platforms could make their applications and mechanisms sufficiently transparent to the users and the general public, which would complement AI literacy efforts. Transparency of recommendation algorithms might alleviate epistemic bubbles because once one is knowledgeable about the criteria that the algorithm relies on, one might seek to include sources of information one deems relevant but which are significantly reduced by the algorithm. However, insights into how and why certain content is presented to persons on social media may not lead to these individuals seeking alternative sources to test their beliefs against. Solving transparency and explainability issues does not directly solve the impact of peer influence or social and political distrust which may influence agents to form or continue to remain in epistemic bubbles. With regards to echo chambers, we claim that having transparent and explainable algorithms is less likely to have an impact as it does not help to reduce the threat of the ‘disagreement-reinforcement mechanism’ since reducing opaque algorithms does not necessarily lead to stronger bonds of trust between citizens. The social impact of this recommendation would need to be validated via empirical studies. However, if one knows that the algorithm is optimized to increase attention for the platform and that suggestions of conspiracy content are thus intensified, then one might be more careful when encountering conspiracy content (for example), that is an entry point for echo chambers (see, for example, Gruber 2019).

Moreover, social media could also be designed in a way that encourages consideration of alternative viewpoints in a current debate, and it could include features that make revisions of posts in social media easy, identifiable and comfortable. Furthermore, there could be disclaimers about content that is related to a recognized echo chamber or the content could be even blocked.¹⁶ Such measures are already taken,

¹⁶ The suggestion of ‘disclaimers’ could help persons outside of echo chambers to be aware of or resist them, however in theory, it is not likely that this recommendation could help those already within an echo chamber to escape. This is because disclaimers do not help to break down the disagreement-reinforcement mechanism, as epistemic agents in echo chambers would likely not trust the disclaimer.

to some degree, so the question would be how efficient they are, if they should be employed more widely if one can provide adequate standards for their application and if such measures help facilitate or undermine RE as a civic virtue (see Siderius and Mostagir 2023). Below is one example of a design setting that might be beneficial with regard to strengthening RE as a civic virtue.

ConsiderIt is an online platform that encourages reflective thought of digital users by guiding them to reflect on their own views and analyse the perspectives of other users (Kriplean et al. 2012; Stiegler and de Jong 2015). This is achieved by the users creating a pro and con list of the relevant stances informing their position on a topic. Users have access to other users’ pro and con lists which they can add to their own list on a topic. *ConsiderIt* then takes the most shared pros and cons by persons holding distinct views and provides users with a constantly evolving guide to including the alternative perspectives to consider (Kriplean et al. 2012). Social media settings such as *ConsiderIt* may help foster RE as a civic virtue since it encourages users to reflect on their initial epistemic positions, by creating a list of pros/cons users are subtly encouraged to reconsider issues and include the diverse perspectives of others. Moreover, by *ConsiderIt* providing users with an inclusive representation of users’ alternative stances, it may encourage public trust between users. *ConsiderIt* cannot be effective for fostering RE as a civic virtue if users choose not to engage in the process of personal deliberation or critical reflection of their views. This discussion is to serve as a preliminary basis for how RE as a civic virtue could be met by some of the already existing design mechanisms. All the measures to strengthen RE as a civic virtue in the domain of social media itself could, of course, be also subject to regulation.

4.3 Regulation of Social Media

In Sects. 2 and 3 of this paper, we have argued that epistemic bubbles and echo chambers undermine the ability to achieve full RE due to the unequal and restrictive epistemic conditions associated with said features. Critical thinking and AI literacy along with some of the design features of social media provide a good basis to foster full RE. Promoting critical thinking and AI literacy as essential features of civic education can help foster citizens’ capacities for engaging in and facilitating a healthy epistemic system. In addition, creating design features to reduce the threat of echo chambers and epistemic bubbles can further aid adherence to the rules of RE. However, these three methods alone may not be enough to safeguard RE as a civic virtue and public reason generally.

When it comes to the topic of social media, the two avenues of the current debate are, whether social media

should be regulated, and if so, how should it be regulated (Association for Progressive Communications 2018; Brannon 2019; Samples 2019) In terms of the first camp, we argue that social media does need to be regulated to protect RE as civic virtue. Since the digital sphere is not strictly distinct from the political public sphere and the norms of public reason are essential to the stability of society and the political autonomy of its citizens, then it might be necessary and justified to insist on an overarching regulatory framework to get social media companies to help foster the ideals of public reason (see also Smith and Niker 2021; Habermas 2023). Social media regulation is essential to ensure that the background conditions of society encourage the ideals of public reason. How social media should be regulated to achieve this and other democratic desiderata is a further area of investigation.

Markus Patberg (Patberg forthcoming, 6) develops three models for how social media could be reformed to benefit democracy. In the first model “private agents of democracy”, he suggests that privately owned social media platforms should be regulated by government and public agencies ensuring that they become facilitators for the protection and enhancement of democratic life and values. The second model namely “public service social media”, suggests creating social media platforms that are funded by the state to serve the interests of the state. These state-funded platforms would exist alongside current privately owned platforms but have a specific political function in that they facilitate digital space for public decision-making and deliberation. For the third model “platform socialism”, Patberg argues for the economic remodelling of social media platforms. The aim is to reconceptualise these platforms from privately to publicly owned corporations. This economic remodel would help to view these platforms as digital common property, which can help strengthen the notion that these platforms are here to serve the common good of democratic society.

Any one of these models could help to facilitate RE as a civic virtue since all three models are aimed at safeguarding democratic institutions and values. The first model could help foster RE as ‘hard’ regulations, in the form of laws as opposed to guidelines and recommendations can help to ensure that the running of these platforms (from content moderation to employee capacity building or algorithm auditing) is aligned with RE as a civic virtue. This model would enable specific legal requirements and legal consequences for non-compliance. This provides a strong motivation for developers, as well as legal recourse for stakeholders to report violations. The second model of state-funded platforms can help facilitate RE by ensuring political equality such as equal open access to these platforms. As Jason Stanly (2022) points out, if Elon Musk chooses to monetise user engagement on *X* (formerly known as *Twitter*) and other private entities follow suit, this puts into

question how citizens could exercise their political equality on these platforms. Monetisation would leave economically disadvantaged citizens in a further marginalised position as they would not be equally heard on these platforms. State-funded platforms could safeguard political equality with regard to public deliberation, which is necessary for citizens to exercise the disposition of RE. Finally, the third model of platform socialism could help foster RE as the content shared on social media platforms would not necessarily need to be content that optimises user engagement to the benefit of the Big Tech business models. Instead, content that helps foster democratic values and civic dispositions such as RE could be prioritised.

For now, our preliminary comment is that, with respect to regulatory proposals there would be a need to establish a positive effect of the regulatory measure on the disposition of citizens to follow the norms of public reasoning. Thus, if one accepts the normative claims of public reason liberalism, and follows our interpretation, there is a need for further interdisciplinary empirical research on this matter.

5 Can the Virtue of RE Help Reduce the Threat of Epistemic Bubbles and Echo Chambers?

In the previous two sections, we have demonstrated how and why epistemic bubbles and echo chambers undermine the method of RE as a civic virtue and hinted at recommendations for fostering RE. In this section, we examine how persons could escape epistemic bubbles and, in some cases, echo chambers by exercising the virtue of RE.

Firstly, in terms of epistemic bubbles, they are likely to undermine the rule of ‘Holism’ since an agent is exposed to a narrow set of information thus preventing the agent from relying on a diverse set of information to test their beliefs against (see Sect. 3.1 for discussion). However, if the disposition to take into account all arguments concerning a topic upon reflection is strongly cultivated through civic education, then in theory an agent in an epistemic bubble may have the motivation and ability to search explicitly for counter-evidence thus choosing to actively expose herself to a wide range of information. As Nguyen (2020, 155) suggests, as long as people engage with information that was previously excluded from their bubble, they could easily ‘shatter’ their bubble. Thus, the implementation of RE as a virtue in civic education could lead to an epistemic bubble dissolving in theory.¹⁷

¹⁷ We have used terminology such as ‘in theory’ and ‘could’ as we are showing the normative connection between the virtue of RE and how cultivating the disposition to follow RE could be a method to reduce the threat of epistemic bubbles. However, this theoretical claim would require empirical investigation.

As discussed in Sect. 3.1 the rule of ‘Fallibilism’ is also undermined by epistemic bubbles since agents generally have an inflated self-confidence and due to confirmation bias and selective exposure they do not see the need to revise their beliefs and consequently are not accustomed to do so. If the disposition to treat all beliefs as defeasible and revisable is cultivated through civic education, then agents in an epistemic bubble may not fall prey to confirmation bias as individuals who are educated to treat all beliefs as defeasible and revisable may be more inclined to question and reassess their own convictions even within the confines of an epistemic bubble. Epistemic humility is a disposition associated with the rule of fallibilism since agents are required to recognize that their beliefs may be flawed or incomplete. If agents are encouraged and taught via civic education to be aware of the need to revise their beliefs, then in theory once in an epistemic bubble they could easily pop the bubble by being willing to revise their beliefs. This willingness for revision could help to prevent the false sense of epistemic self-confidence agents in epistemic bubbles experience since they are aware that their beliefs are defeasible.

RE as a method and correspondingly also as a civic virtue includes a strong social dimension and epistemic division of labour (see also Jäger and Malfatti 2020). If citizens actively choose to exercise the civic virtue of RE, this could lead an agent to consider alternative arguments and the opinions of experts, thus, increasing their epistemic circle, and allowing for the right kind of epistemic cooperation to be possible. Consequently also ‘Rationality’, according to which agents have to choose the most plausible system of beliefs, could help to burst bubbles, in the course of reflection.

Secondly, in terms of echo chambers, the disposition to follow all four rules of full RE is undermined (see Sect. 3.2). Echo chambers threaten the civic virtue of RE mainly because of the epistemic distrust members of echo chambers have towards non-members. As stated earlier, one of the most dangerous features of echo chambers is what Nguyen (2020, 147) calls the “disagreement-reinforcement mechanism”. To recap, the disagreement-reinforcement mechanism accounts for how echo chambers use past beliefs as a form of epistemic “inoculation” so that when members come into contact with contrary beliefs that are supposed to undermine the epistemic creditability of the group, these contrary beliefs end up strengthening the group’s epistemic credibility (Begby 2013). Nguyen suggests that a “social epistemic reboot” is a plausible method for individuals to remove themselves from echo chambers however, it is a taxing epistemic requirement as it demands the agent to (re) evaluate every belief they hold (Nguyen 2020, 157–58). The social epistemic reboot may be facilitated by the incorporation of AI literacy and critical thinking skills as they provide the agents with cognitive and technical capabilities to

be self-critical. Similarly, insofar as agents are encouraged to view their beliefs as revisable and encouraged to expose themselves to alternative information then the virtue of RE could help facilitate Nguyen’s social epistemic reboot since RE is the disposition of (self)critical citizens. Some episodic evidence suggests that emotional connections with other persons outside of the echo chamber can motivate a kind of epistemic reboot (Nguyen 2020, 158). Since emotions and corresponding beliefs are supposed to play a prominent role within the method of RE according to important proponents of the method (DePaul 1993, 180–82; Elgin 1996, 146–69) these episodes of epistemic reboot might be also reconstructed as RE processes. In general, the disposition to search for commonly shared beliefs as a basis for justification and the corresponding epistemic humility, as included in the civic virtue of RE, might be safeguarding agents from entering echo chambers from the start.

An important point to note, is that RE as a civic virtue can help agents prior to being in and those who are already caught in epistemic bubbles, whereas it may only help prevent agents from entering echo chambers but not help remove those who are already caught within one.¹⁸ This is because RE as a civic virtue serves rather as a preventative disposition for agents to exercise, as opposed to a cure for them to rely on in flawed epistemic environments. One crucial element of prevention is the skill set to assess the dangers of entering an epistemic community (such as by developing critical thinking which can help safeguard against the harmful epistemic and political effects of the disagreement-reinforcement-mechanism). Here the disposition to follow RE is helpful, along with recommendations suggested in Sect. 4.1.

The aim of this section is to interrogate the two-way channel between RE, epistemic bubbles and echo chambers. The methods of fostering RE discussed in the previous section could potentially offer a solution to reducing the threat of epistemic bubbles and echo chambers as the cultivation of the virtue of RE helps to instil epistemic practices that could encourage the fallibility of one’s belief system, an openness to new information, systematic thinking, and a recognition that one’s political understanding may be incomplete and is always provisional.

¹⁸ The disposition to follow the method of RE might also be helpful for an agent who has developed the disposition to a minimal degree before entering an echo chamber. This minimal devolvement of RE as a civic virtue may make them aware that the distrust of outside sources and persons (arising from disagreement-reinforcement mechanism) is justified inadequately. Thus, this minimal devolvement of RE as a civic virtue may help epistemic agents to reflect and seek counter-evidence to challenge their belief, this might lead them to leave the echo chamber. Moreover, if one can build up a shared base of emotionally tied beliefs that ground the effort to debunk the echo chamber, the disposition to follow the method of RE could help in the debunking process. However, empirical evidence would be needed to substantiate these claims further.

6 Conclusion

We proposed a novel analysis of how epistemic bubbles and echo chambers which are fuelled by the current design of social media and digital platforms are harming liberal democracies by undermining the civic virtue of RE. This analysis is based on a Rawlsian perspective on justice, that stresses the importance of public reason and reasonable justification for engagement on political issues in the public sphere. As shown, the digital sphere and the public dissolve into one another in liberal democracies. Thus, we claim that the Rawlsian conditions for public reason must extend into the digital sphere to social media platforms. The conditions for public reason are spelled out with reference to a rule-based analysis of the method of (full) RE. Having the disposition to follow these rules adequately thus amounts to the civic virtue of RE. We show how the civic virtue of RE might be affected negatively by epistemic bubbles and echo chambers. We further provide ideas for fostering RE as a civic virtue via civic education by the state and civil society, changes in the design of social media and its regulation. The aim of calling for RE as a civic virtue is to acknowledge that the cultivation of the four rules can help develop citizens' capacities for justice by fostering healthy epistemic practices that can encourage citizens to exercise their political virtues when engaging in public debate. There are, of course, other urgent problems that liberal democracies currently face, like growing inequality, disruptive technologies and climate change. However, the findings highlight important normative implications of public reason liberalism and provide initially plausible hypotheses for interdisciplinary empirical inquiry that might help safeguard basic democratic norms.

Acknowledgements We would like to thank the two anonymous reviewers of this special issue whose comments helped make valuable additions to this paper. In addition, we thank Veli Mitova, Boaz Miller for their comments on an earlier version of this paper in February 2024. Lastly, we would like to thank the audiences at the Forum on Philosophy, Engineering, and Technology (fPET) in 2023 Delft, and the general conference of The European Consortium For Political Research (ECPR) 2023 in Prague for their comments and suggestions.

Author Contribution Both authors contributed equally to the paper.

Funding Open access funding provided by University of Johannesburg.

Data availability The authors of this paper have not used data sets this is in line with the methods of a philosophical paper. Thus we can't provide a data availability statement. All the sources we have used in the construction of this paper have been noted in the in-text citations and the reference list.

Declarations

Conflict of interest The authors declare that there are no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Association for Progressive Communications (2018) 'Content regulation in the Digital Age: submission to the United Nations Special Rapporteur on the right to Freedom of Opinion and expression'. Association for Progressive Communications. https://www.apc.org/sites/default/files/APC_submission_Content_regulation_in_the_digital_age_FINAL.pdf
- Bail CA, Lisa P, Argyle TW, Brown JP, Bumpus H, Chen MB, Fallin Hunzaker J, Lee M, Mann F, Merhout, and Alexander Volfovsky (2018). 'Exposure to Opposing Views on Social Media Can Increase Political Polarization'. *Proceedings of the National Academy of Sciences* 115 (37): 9216–21. <https://doi.org/10.1073/pnas.1804840115>
- Ball B (2021) On Journalism, Knowledge, and Democracy'. *Moral Philos Politics* 8(1):5–26. <https://doi.org/10.1515/mopp-2019-0033>. 'Defeating Fake News
- Bardon Aurélie, Bonotti M, Zech ST, and William Ridge (2023) Disaggregating civility: politeness, Public-Mindedness and their connection. *Br J Polit Sci* 53(1):308–325. <https://doi.org/10.1017/S000712342100065X>
- Bardon Aurélie, Bonotti M, Zach ST (2024) Civility, contentious monuments, and Public Space. *The Self, Civic Virtue, and Public Life*. Routledge
- Battaly H (2018) Closed-mindedness and Dogmatism. *Episteme* 15(3):261–282. <https://doi.org/10.1017/epi.2018.22>
- Baumberger C (2020) and Georg Brun. 'Reflective Equilibrium and Understanding'. *Synthese*, February. <https://doi.org/10.1007/s11229-020-02556-9>
- Begby E (2013) The Epistemology of Prejudice. *Thought: J Philos* 2(2):90–99
- Beisbart C, and Georg Brun (2024) Is there a defensible conception of reflective equilibrium? *Synthese* 203(3):79. <https://doi.org/10.1007/s11229-024-04495-1>
- Benton P (2023) 'The Complementary Relation Between the Right and the Good in Justice as Fairness: Implications for Liberal Democracies'. PhD Thesis, University of Pretoria
- Blake M (2014) 'Burdens of Judgment'. In *The Cambridge Rawls Lexicon*, edited by David A. Reidy and Jon Mandle, 74–78. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139026741.026>
- Boettcher J (2014) 'Political Virtues'. In *The Cambridge Rawls Lexicon*, edited by David A. Reidy and Jon Mandle, 631–34.

- Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139026741.163>
- Bozdog E, Jeroen, van den, Hoven (2015) Breaking the Filter bubble: democracy and design. *Ethics Inf Technol* 17(4):249–265. <https://doi.org/10.1007/s10676-015-9380-y>
- Brändle C, Schmidt MW (2021) ‘Autonomous Driving and Public Reason: A Rawlsian Approach’. *Philosophy & Technology*, August. <https://doi.org/10.1007/s13347-021-00468-1>
- Brannon VC (2019) ‘Free Speech and the regulation of Social Media Content’. R45650. Congressional Research Service. <https://crsreports.congress.gov/product/pdf/R/R45650>
- Coeckelbergh M. ‘Democracy, Agency E (2023) and AI: Political Epistemology in Times of Artificial Intelligence’. *AI and Ethics* 3 (4): 1341–50. <https://doi.org/10.1007/s43681-022-00239-4>
- Daniels N (1996) Justice and Justification: reflective equilibrium in theory and practice. Cambridge Studies in Philosophy and Public Policy. Cambridge, New York: Cambridge University Press
- Daniels N (2020) ‘Reflective Equilibrium’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2020. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2020/entries/reflective-equilibrium/>
- DePaul MR (1993) Balance and refinement beyond coherence methods of Moral Inquiry. Routledge, London
- Dillahunt TR, Brooks CA, Gulati S (2015) ‘Detecting and Visualizing Filter Bubbles in Google and Bing’. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, 1851–56. Seoul Republic of Korea: ACM. <https://doi.org/10.1145/2702613.2732850>
- Dubois E, and Grant Blank (2018) The Echo Chamber is overstated: the moderating effect of political interest and diverse media. *Inform Communication Soc* 21(5):729–745. <https://doi.org/10.1080/1369118X.2018.1428656>
- El-Bermawy MM (2016) ‘Your Filter Bubble Is Destroying Democracy’. *Wired*, 2016. <https://www.wired.com/2016/11/filter-bubble-destroying-democracy/>
- Elgin CZ (1996) Considered Judgment. Princeton University Press
- Elgin CZ (2017) True enough. The MIT Press, Cambridge, MA
- Furman K (2023) Epistemic bunkers. *Social Epistemology* 37(2):197–207. <https://doi.org/10.1080/02691728.2022.2122756>
- Gaus GF (2015) ‘Public Reason Liberalism’. In *The Cambridge Companion to Liberalism*, edited by Steven Wall, 112–40. Cambridge Companions to Philosophy. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139942478.008>
- Gettier EL (1963) ‘Is Justified True Belief Knowledge?’ *Analysis* 23 (6): 121–23. <https://doi.org/10.1093/analys/23.6.121>
- Grade Band Progression Charts (2021) AI4K12. 29 June 2021. <https://ai4k12.org/gradeband-progression-charts/>
- Gruber T (2019) ‘What can AI do for humanity’. Seoul, June 25. <https://vimeo.com/562566762/9351aedf5e>
- Habermas Jürgen (2023) *A New Structural Transformation of the Public Sphere and Deliberative Politics*. Translated by Ciaran Cronin. English edition. Cambridge Hoboken: Polity
- Hannák Anikó, Sapieżyński P, Khaki AM, Lazer D, Mislove A, Christo Wilson (2017) ‘Measuring Personalization of Web Search’. arXiv. <http://arxiv.org/abs/1706.05011>
- Hardwig J (1985) Epistemic Dependence. *J Philos* 82(7):335–349
- Hitchcock D (2022) ‘Critical Thinking’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Winter 2022. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2022/entries/critical-thinking/>
- Ichikawa JJ, Matthias Steup (2018) ‘The Analysis of Knowledge’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2018. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/>
- Introna L, Nissenbaum H (2000) Shaping the web: why the politics of search engines matters. *Inform Soc* 16(3):169–185. <https://doi.org/10.1080/01972240050133634>
- Jäger C (2020) and Federica Isabella Malfatti. ‘The Social Fabric of Understanding: Equilibrium, Authority, and Epistemic Empathy’. *Synthese*, July. <https://doi.org/10.1007/s11229-020-02776-z>
- Jamieson KH, and Joseph N. Cappella (2008) Echo Chamber: Rush Limbaugh and the conservative media establishment. Oxford University Press
- Jason Stanley (2022) ‘Elon Musk’s Not-so-Hidden Agenda’. Project Syndicate. 2022. <https://www.project-syndicate.org/commentary/musk-using-twitter-to-undermine-democracy-by-jason-stanley-2022-11>
- Kauppinen A and Jaakko Hirvelä. forthcoming. ‘Reflective Equilibrium’. In *The Oxford Handbook of Normative Ethics*, edited by David Copp, Tina Rulli, and Connie Rosati. Oxford University Press
- Kinkead D and David M. Douglas (2020) ‘The Network and the Demos: Big Data & the Epistemic Justifications of Democracy’. In *Big Data and Democracy*, edited by Kevin Macnish and Jai Galliot. Edinburgh University Press
- Krasnoff L (2014) ‘The Reasonable and the Rational’. In *The Cambridge Rawls Lexicon*, edited by David A. Reidy and Jon Mandle, 692–97. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139026741.179>
- Kriplean T, Morgan J, Freelon D, Borning A, Lance Bennett (2012) ‘Supporting Reflective Public Thought with Considerit’. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 265–74. Seattle Washington USA: ACM. <https://doi.org/10.1145/2145204.2145249>
- Lackey J (2021) ‘Echo Chambers, Fake News, and Social Epistemology’. In *The Epistemology of Fake News*, edited by Sven Bernecker, Amy K. Flowerree, and Thomas Grundmann, 1st ed., 206–27. Oxford University Press Oxford. <https://doi.org/10.1093/oso/9780198863977.003.0010>
- Lewandowsky S, Ullrich KH, Ecker CM, Seifert NS, and John Cook (2012) Misinformation and its correction: continued influence and successful debiasing. *Psychol Sci Public Interest* 13(3):106–131. <https://doi.org/10.1177/1529100612451018>
- Long D, Magerko B (2020) ‘What Is AI Literacy? Competencies and Design Considerations’. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. Honolulu HI USA: ACM. <https://doi.org/10.1145/3313831.3376727>
- Miller B, Isaac Record (2013) Belief in a Digital age: on the Epistemic Implications of Secret Internet Technologies’ Justified. *Episteme* 10(2):117–134. <https://doi.org/10.1017/epi.2013.11>
- Nelson JL, Webster JG (2017) The myth of partisan selective exposure: a portrait of the online Political News Audience. *Social Media + Soc* 3(3):205630511772931. <https://doi.org/10.1177/2056305117729314>
- Nguyen C, Thi (2020) Echo chambers and Epistemic Bubbles. *Episteme* 17(2):141–161. <https://doi.org/10.1017/epi.2018.32>
- Pariser E (2012) *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Reprint Edition. London: Penguin Publishing Group
- Patberg M Forthcoming. ‘What is Social Media’s place in democracy?’ *The Review of Politics*
- Pham A, Rubel A, and Clinton Castro (2022) ‘Social Media, Emergent Manipulation, and Political Legitimacy’. In *The Philosophy of Online Manipulation*, by Fleur Jongepier and Michael Klenk, 1st ed., 353–69. New York: Routledge. <https://doi.org/10.4324/9781003205425-21>
- Quong J (2011) Liberalism without perfection. Oxford University Press, Oxford; New York
- Quong J (2022) ‘Public Reason’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2022. Metaphysics

- Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2022/entries/public-reason/>
- Rawls J (1974) ‘The Independence of Moral Theory’. *Proceedings and Addresses of the American Philosophical Association* 48:5. <https://doi.org/10.2307/3129858>
- Rawls J (1995) Political liberalism: reply to Habermas. *J Philos* 92(3):132. <https://doi.org/10.2307/2940843>
- Rawls J (1997) The idea of public reason revisited. *Univ Chic Law Rev* 64(3):765–807. <https://doi.org/10.2307/1600311>
- Rawls J (1999) *A theory of justice: revised Edition*. Belknap, Cambridge, Massachusetts
- Rawls J (2001) *Justice as Fairness: A Restatement*. Edited by Erin Kelly. Cambridge, Massachusetts: Harvard University Press
- Rawls J (2005) *Political liberalism: expanded Edition*. Columbia classics in philosophy edition. Columbia University
- Rini R (2019) ‘Social Media Disinformation and the Security Threat to Democratic Legitimacy’. *NATO Association of Canada: Disinformation and Digital Democracies in the 21st Century*, 10–14
- Samples J (2019) ‘Why the Government Should Not Regulate Content Moderation of Social Media’. Cato Institute. 9 April 2019. <https://www.cato.org/policy-analysis/why-government-should-not-regulate-content-moderation-social-media>
- Schmidt MW (2022) *Das Überlegungsgleichgewicht als Lebensform: Versuch zu Einem Vertieften Verständnis Der Durch John Rawls bekannt gewordenen rechtfertigungsmethode*. Brill, Mentis, Paderborn
- Schmidt MW (2024) Defining the method of reflective equilibrium. *Synthese*, 203(5), 170. <https://doi.org/10.1007/s11229-024-04571-6>
- Siderius J (2023) and Mohamed Mostagir. ‘When Should Platforms Break Echo Chambers?’ <https://siderius.lids.mit.edu/when-should-platforms-break-echo-chambers/>
- Smith L, and Fay Niker (2021) What social media facilitates, Social Media should regulate: duties in the New Public Sphere. *Political Q* 92(4):613–620. <https://doi.org/10.1111/1467-923X.13011>
- Stiegler H, Menno DT, de Jong (2015) Facilitating personal deliberation online: Immediate effects of two ConsiderIt variations. *Comput Hum Behav* 51(October):461–469. <https://doi.org/10.1016/j.chb.2015.05.018>
- Sunstein C (2007) *Infotopia: how many minds produce knowledge*. Oxford University Press.
- Sunstein C (2018) *Republic: Divided democracy in the age of social media*. Princeton University Press.
- Zuiderveen Borgesius, Frederik J, Trilling D, Möller J, Bodó Balázs, Claes H, De Vreese, Natali Helberger (2016) Should we worry about Filter bubbles? *Internet Policy Rev* 5(1). <https://doi.org/10.14763/2016.1.401>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.