Towards a Critical Social Epistemology of Social Media

Joshua Habgood-Coote

Draft – please ask before citing, comments appreciated!

Abstract

What are the proper epistemic aims of social media sites? A great deal of social media critique is in the grips of an *Epistemic Apocalypse* narrative, which claims that the technologies associated with social media have catastrophically undermined our traditional knowledge-generating practices, and that the remedy is to recreate our pre-catastrophe practices as closely as possible. This narrative relies on a number of questionable assumptions, and problematically narrows the imaginative possibilities for redesigning social media. Our goal in this paper is to shake off the epistemic apocalypse narrative and offer a better account of the epistemic aims of social media. I will pursue a critical approach to social epistemology that appreciates the non-ideal features of epistemic systems, and the ways in which knowledge production can be a site of domination, and apply this framework to thinking about the epistemic design of social media sites. I will argue that social systems ought to pursue three distinct epistemic goals: promoting good epistemic outcomes for users, realising epistemically good institutional features, and achieving structural epistemic justice. Although these goals are often mutually supportive, I will consider a number of cases in which these values lead to dilemmas about how to design epistemic institutions, which can only be resolved by appealing to ethical considerations. I will close by considering some ways in which social media might realise these aims.

Introduction

During the 1990s and into the early 2000s, many technologists were motivated by what Barbrook and Cameron (1995) called the *Californian Ideology*. This view promised that developments in communication technologies would lead to a blossoming of individual freedoms, a democratisation of discourse, and a renewal of democratic values. Suffice to say that these promises have not been realised. Academic and popular discourse about technology is currently gripped by something like the inverse of the Californian Ideology, which we might call the *Epistemic Apocalypse* narrative. According to the Epistemic Apocalypse narrative, the technological developments of the last twenty-to-thirty years have poisoned and twisted our knowledge generating practices in ways that have undermined freedom of speech and created the conditions for antidemocratic propaganda. These problems undermine the epistemic value of democracy, and pose a threat to the continued existence of liberal democracies. Those of us in the grips of this apocalypse are flailing, both literally and conceptually. In order to even make sense of our predicament we need a swathe of neologisms—'fake news', 'echo chamber', 'post-truth'—each embedded within an alarming epidemiological and ecological

metaphors (the 'virus of fake news' (Basol, Roozenbeek, van der Linden 2020), the 'information pollution' of misinformation (Lynch 2019), and the 'junk food diet' of the personalised newsfeed (Pariser 2011)).

While the epistemic apocalypse narrative may have its uses in bringing attention to the problems with our online lives, it also profoundly restricts the way we think about the epistemology of the internet, and carries over much of the troublesome individualism, technodeterminism, and whiggish history that Barbrook and Cameron identify in the Californian Ideology. The apocalypse narrative assumes that technological development is the harbinger of epistemic catastrophe, that our epistemic culture pre-internet (or pre-2016 internet) was a halcyon age, and that our primary goal should be to try to painstakingly recreate pre-internet forms of sociality. This narrative is undoubtedly appealing, but it is intellectually and politically restrictive. Epistemic problems are nothing new: the AIDs crisis, the myth that colonised lands were unpopulated, and resistance to the existence of human-caused climate change are epistemic catastrophes which started long before the internet. Although technology certainly plays a role in epistemic crisis, focusing attention on it as the sole source of our ills locks us into a technochauvinist mindset where all of our problems and solutions are technical ones (Broussard 2018). And, comparing our current predicament to a pre-lapsarian state distracts our attention from the important task of cultivating new forms of collective organisation that do a better job of realising the epistemic goals that we care about.

Our goal in this paper is to begin to collect together the normative tools we need to think about our collective epistemic aims, and to begin to think about how we might design social media institutions that promote those goals. This project approaches social media sites using the tools of *systems-oriented social epistemology* (Goldman 2010), in the hope that these sites can become better epistemic tools. In a way this is a hopelessly utopian project. Building better social media—either by improving current platforms, or by starting from scratch—is a hugely difficult and time-consuming task. Any amelioratory project will surely be small and exploratory. Nonetheless imagining things otherwise is a politically important tool that can help us unsettle our assumptions, and frame new options for social life (Rorty 1983, Stoetzler and Yuval-Davis 2002, Biss 2013, Dotson 2018).

This project is incomplete in several ways:

First, I will set to one side some important ethical and political issues about social media sites. A full project of redesigning social media sites will address issues of free speech in online spaces, and legal questions about the status of social media sites as platforms, publishers, or speakers. One of the lessons of this paper is that the epistemic evaluation of social institutions relies on ethical values, so it hopefully helps to see us towards a rapprochement between ethical and epistemic evaluation. I will not address the view—perhaps a hangover from the Californian Ideology (see Barlow 1996)—that social media sites ought only to promote non-epistemic values such as freedom of speech. Anyone who feels the draw of the epistemic apocalypse

narrative is already committed to evaluating social media sites on epistemic grounds.

Secondly, in the interests of pushing back against the epistemic apocalypse narrative, I will stress the continuities between social media sites and other kinds of knowledge-providing institutions. There is a lot more work to be done on the epistemic significance of the distinctive features of social media, including the algorithmic timeline (Noble 2018), content moderation (Frost-Arnold, forthcoming), the incentive structures created by technological systems (Nguyen forthcoming), and distinctive discursive practices that emerge online (Marsili 2020), (Pepp, Michaelson, and Sterken 2019, MS).

Thirdly, we will focus on the epistemic features of social media systems in isolation from wider epistemic practices. A proper consideration of the ways in which the ills of social media relate to wider socio-epistemic problems and the epistemic vices of users will have to wait for another day.

The plan of action is as follows. In the first section, I will make some ground-clearing remarks. In the second section, I will consider the proper epistemic aims of social systems in general, making the case that they ought i) to promote good epistemic states for their members, ii) to realise epistemically good institutional properties, and iii) to achieve structural epistemic justice. I will call an approach to social epistemology that integrates considerations of these three goals *Critical Social Epistemology.*[1] In the third section, I will demonstrate the importance of thinking about these aim together by presenting cases in which the different aims come into conflict, and I will argue that we ought to resolve these conflicts by appealing to ethical considerations. In the fourth section, I will apply the framework of critical social epistemology to social media sites. I will first consider several debates about the design of social media that appeal to these aims. I will then show how the conflicts between these three epistemic justice can also illuminate debates about the design of social media sites. I suggest that highly connected networks might be epistemically good for individuals, but bad for groups, consider the value of a fragmented public sphere in conditions of epistemic oppression, and explore how structural epistemic justice can conflict with the promotion of majoritarian epistemic interests. In the final section, summarise some important features of the project of a critical social epistemology of social media in more detail.

1. Social Media

---

[1] I borrow this term from (Congdon 2015) (see also Dillon 2012 on critical character theory). I intend the 'critical' modifier to pick out approaches to social epistemology that pay attention to the non-ideal aspects of epistemic practices, and the ways in which knowledge-generating practices can be a site for domination. As I will use the term, critical social epistemology can and should integrate the concerns of non-critical approaches, by combining the concern with traditional epistemic ideals (truth, understanding justification) with attention to the ways actual social practices impact the pursuit of those ideals.

What is a social media site? For the purposes of this essay, I will assume that a social media site is one whose primary service is to allow users to hang out, where the majority of content shared is generated by those users.[2] Many sites which are not social media sites host user generated content and are some are used as hang-out spaces. We might think of wikipedia, Tinder, and Quora. These sites will not count as social media under this definition because their *primary* service is not to allow people to hang out.[3] Social media sites are distinctive because they combine a large number of active (and often public or semi-public) participants with the kind of reliance on personal credibility that would be more normally found in intimate private conversations.

In order to avoid a mess of overlapping terminology I will introduce neutral termsfor the basic structure of social media sites. I will talk about sites having *users* who will typically have a profile displaying personal information, and be *connected* to various other users either through symmetric or asymmetric relationships. Users are able to *post content*, in the form of pictures, text, video, or audio, and are also able to *repost* other people's content. Posts may be *public* or *private*, and private posts will typically only be accessible to users who are connected to the poster. Most sites will combine aspects public and private communications (for example a public profile, and private messaging). Sites are organized around a *timeline* that organizes posts from other users, typically using some element of algorithmic sorting to choose what content gets presented to which users in which order.

From the point of view of systems-oriented social epistemology, a social media site is not merely a piece of communications technology; it is a socio-technical assemblage that meshes together a technology with certain communicative affordances, and a set of social practices that determine what and how people post and repost content. The social practices will be constrained—but not completely determined by—communications technology, and the effectiveness of technological systems will depend on suitable social practices. This meshing means that the epistemic outcomes of social media depend on both social and technical features of social media: a social media site which allows unlimited reposting might be epistemically disastrous if it is combined with reckless social practices for reposting, whereas the same technology might be of great value if combined with judicious social practices for reposting. Many critical discussions of social media have focused exclusively on one side of this complex system. For example, calls for media literacy education focus exclusively on social practices, whereas calls to remove the 'like' button (or its analogues) focus on technological affordances. This is not to say that either intervention is a bad idea, but rather that in order to think about the epistemic features of social media we need to think about both social and technological features in an integrated way.

2. Critical Social Epistemology

---

[2] For alternative definitions, see (boyd and Ellison 2007), (Sunstein 2017, 22), (Zuckerman and Rajendra-Nicolucci 2020).
[3] On the social epistemology of Wikipedia, see (Wray 2009, and Frost-Arnold 2018).

One place to start thinking about the epistemology of social media is by considering what social media sites themselves take their epistemic aims to be. Social media sites—especially facebook and twitter—have presented themselves as *platforms.* They exploit the ambiguity of 'platform' to present themselves to users as neutral platforms *for speech* while at the same time presenting themselves to companies as effective *commercial* platforms for effective targeted adverts (Gillespie 2010, Srnicek 2017).[4] This rhetorical strategy suggests that social media sites' explicit epistemic aim is to provide users with an online communication system in which all users are given equal opportunities (platforms for speech). This aim surfaces in a number of different debates about social media: in the way sites appeal to the value of freedom of speech to justify not taking action on harmful speech, their repeated claims that they are not 'arbiters of truth', and in legal debates about the social media sites' legal responsibilities for what users post and repost.[5]

The presentation of actually existing social media sites as epistemically neutral is deeply misleading. Major social media sites have systematically boosted extreme content in ways that supports white supremacists and anti-democratic propagandists (Tufekci 2018, Munn 2019, Alfano, Carter, Cheong 2018), and knowingly boosted unreliable ring-wing news sites in the interests of promoting 'trusted' sites (Oremus 2018, Bauerlein and Jeffery 2020). They have relied on problematic and opaque content moderation policies that are implemented by a large and outsourced workforce (Roberts 2018, Frost-Arnold forthcoming). Their failure to address systematic harassment of women and minority groups has created an unequal discursive environment (Sobieraj 2017). Setting to one side failures to achieve neutrality, it is unclear that this is even an attractive ideal. One can imagine a fastidiously neutral site with very limited content moderation and affordances that allow content to quickly spread that allows a minority of incompetent or malicious users to systematically spread false information and derail any attempts at co-operative inquiry. Such a site might be epistemically and politically neutral, but with would be epistemically disastrous.

So, if they ought not to pursue neutrality, what should the epistemic aims of social media sites be? We can approach this question through recent research in *systems-oriented social epistemology*, which evaluates social systems by considering how well they achieve epistemic goals (Goldman 2010). This literature presents us with three ways to think about the epistemic aims social institutions: *veritism* has focused on how social institutions affect the accuracy their users' beliefs (Goldman 2000), *network epistemology* has focused on the epistemic properties of social networks (Zollman 2007, 2010, 2013), and work on *epistemic* has focused on how social institutions undermine peoples' epistemic agency (Dotson 2014, Fricker 2007), and

---

[4] For a sceptical take on the effectiveness of targeted online advertising, see (Hwang 2020).
[5] The ideal of neutrality might also be deployed against interventions which aim to improve epistemic outcomes. For a defence of epistemic paternalism on social media, see (Castro, Pham, Rubel 2020).

create inverted epistemologies that support domination (Mills 2007).[6] Considering these approaches together points us toward a unified way of thinking about the epistemic goals of social institutions, which combines *individualist* considerations about individual-level epistemic goods, *institutionalist* considerations about the epistemic properties of social networks, and *considerations of epistemic justice* including issues about the distribution of epistemic goods, and the ways epistemic systems enact social power.

There is a huge amount of work exemplifying each kind of approach to social epistemology, and in the rest of this section, we will introduce some views that exemplify each approach to give us the tools we need to think about the social epistemology of social media.

2.1. Individualism

Individualistic approaches to social media reduce the epistemic features of a social system to the epistemic states of some set of individuals. Most often the relevant properties are changes in the epistemic properties of peoples' beliefs after using that social system. So we might evaluate a newspaper by thinking about how it changes its readers' beliefs. Are readers' beliefs more accurate, more justified, backed by better understanding? As will become clear below, individualism has a great deal in common with consequentialist approaches in ethics, which evaluate the goodness of an action in terms of its consequences for some ethically relevant properties.

The best developed version of Individualism is Alvin Goldman's *veritism* (Goldman 2000). For veritism, the individual states of interest are degrees of belief, and these states are better the more accurate they are .We represent true propositions with 1 and false propositions with 0, and an individual's degree of belief in some proposition with a number in the interval [1,0]. What Goldman calls the *fundamental veritistic value* of a degree of belief is a matter of how close it is to the truth of the matter: so if p is true, a degree of belief 1 is maximally accurate, degree 0.9 is less accurate and so on (values for false propositions are switched). We give a linear measure of the accuracy of beliefs by reading off how close a belief is to the truth of the matter (so that credence 0.9 in a true proposition gets a score of 0.9, and a credence of 0.2 in a false proposition gets 0.8).[7]

To think about the epistemic value of social practices, Goldman introduces the notion of *instrumental veritistic value,* which corresponds to how relying on the outputs of a social practice effects the average accuracy of individuals' beliefs. If there are three news stories of interest in Tuesday's paper—A, B, and C—which are all accurately reported, then the instrumental veritistic value of reading the paper for a credulous reader who started off with a credal profile {0.4, 0.5, 0.6} and ended up

---

[6] The individualist bent of some work on epistemic oppression (Fricker 2007) has been complemented by focus on epistemic justice as a virtue of social institutions (Anderson 2012).
[7] To confuse things, Goldman identifies true beliefs with knowledge, but we will not follow this convention.

with {1.0, 1.0, 1.0} would be the average of the change in their credal state (in this case +0.5). According to veritism, the epistemic aim of social institutions is to have the best possible effects on the accuracy of individuals' beliefs. Veritism allows us to rank the epistemic value of different social practices by ranking their average effects on the accuracy of individuals' beliefs about some set of propositions. This allows us to neatly explain why a site that hosts a lot of (believable) false content is epistemically bad: individuals who rely allow the site to shape their credences will have a lower average instrumental veritistic value than those who rely on other belief-forming practices (for example, reading a newspaper, or not reading anything at all).

Two clarificatory points. First, Veritism evaluates individual beliefs, meaning that a site is only epistemically bad if it hosts false claims that have a negative effect on the accuracy of users' beliefs (either by causing users to form false beliefs, or by causing them to move from true belief to suspension). Second, Veritism does not allow us to evaluate an epistemic system simply by looking at the number of false stories; we need to weigh up true and false stories against one another. Goldman's view unifies the goals of acquiring truth and avoiding error, giving these two goals equal weight. We might want to distinguish these two goals, and to assign them different weights.[8]

Veritisim is an *individualist* approach because it evaluates social systems by reference to individuals' epistemic states. It is also a *consequentialist* approach, because it evaluates epistemic practices by reference to their consequences. Veritisim is closely related to the individualistic thread in social epistemology that takes account of the epistemic significance of social factors by considering how social forms of evidence (testimony, disagreement, scientific consensus) impinge on individuals' beliefs (Goldman 2010). Other individualist approaches to systems-oriented social epistemology are available: we might evaluate social institutions by looking at different individual outcomes (justification, rationality, understanding, epistemic agency, epistemic virtue),[9] by using a different way to measure accuracy (for example, appealing to total accuracy rather than average accuracy), by adding in additional measures of value (for example, by measuring the distribution of accuracy of beliefs between different people or across different questions (see section 3.3.)), or by using a different kind of normative framework (for example, by using a non-consequentialist value theory).[10] Each of these issues is complex, and we will take Goldman's veritism as our representative individualist approach.

2.2. Institutionalism

---

[8] Goldman implicitly does this in his discussion of the social epistemology of blogging (Goldman 2008), which appeals exclusively to error avoidance in support of the claim that traditional newspapers are epistemically better than blogging (Coady 2011).
[9] The choice between social systems that aim for accuracy or justification is particularly interesting: accuracy-focused systems will promote accurate beliefs, whereas justification-focused systems will promote both justified true beliefs and justified false beliefs.
[10] The debate about epistemic consequentialism has led to discussions about possible non-consequentialist views for individual epistemology (Sylvan 2020), but to my knowledge, there are no developed non-consequentialist approaches to doing social epistemology.

In addition to thinking about how social practices and institutions effect the epistemic states of individuals, we ought to think about the epistemic status of group-level properties of social practices and individuals. Institutionalist approaches to social epistemology investigate these group-level properties. For example, it might be important for us to to think about how reliable or well-justified a scientific consensus is. Although the reliability of individual scientists will contribute to the reliability of a collective position, results like the Condorcet Jury theorem (List and Goodin 2001), and the diversity trumps ability theorem (Hong and Page 2004) mean that the reliability of a consensus will not relate in a straightforward way to individual reliability.[11]

The divergences between the epistemic properties of groups and individuals lead to what Mayo-Wilson, Zollman, and Danks (2011: 654) call *the Independence Thesis*, the claim that methodological prescriptions for communities and those for individuals are logically independent. What might be good advice for a community might be bad advice for individuals, and what might be good advice for individuals might be bad advice for communities. This independence theorem is central to *Institutionalist* approaches to social epistemology, which focuses on group-level properties as the target of epistemic evaluation.

Network epistemology is one kind of institutionalist approach that focuses on the epistemic properties of different kinds of communication network (Zollman 2013). Research in this field has investigated what communication network leads to the most reliable consensus for agents who are engaged in ongoing investigation (Zollman 2007, 2010), whether communities of inquirers do better when their members follow established results or pursue maverick strategies (Weisberg and Muldoon 2009), and how propagandists who selectively share results can undermine the reliability of consensus without needing to mislead or produce fraudulent research (Weatherall, O'Connor, Bruner 2020, Weatherall and O'Connor 2019). Network epistemology uses agent-based modelling and decision theory, and typically involves idealising assumptions about agents within these networks.[12] These assumptions open up difficult questions about how these models relate to real-life networks, especially non-ideal ideal networks (Mills 2005, Frey and Šešelja 2018).

---

[11] The Condorcet jury theorem states that in a group which votes independently and each individual is more likely to be right than wrong, the majority position is more likely to be right than wrong, and as the size of the group increases the reliability of the consensus swiftly tends toward perfect reliability (List and Goodin 2001). The diversity trumps ability theorem states that groups of problem solvers with diverse approaches to a problem will outperform less diverse groups with higher expertise. In cases where the assumptions of either theorem are met, the reliability of a collective that reaches a position either by majoritarian voting or by collective problem-solving will be higher than the average reliability of group members (Hong and Page 2004).
[12] Some work in network epistemology takes a less idealised slant, focusing on modelling propagandistic strategies (Holman and Bruner 2017, Weatherall, O'Connor, Bruner 2018))

Network epistemology investigates certain kinds of institutional properties (centrally network structure and inquisitive strategies), and there are many other institutionalist approaches that focus on different epistemic properties. For example, much feminist work on scientific objectivity would count as a kind of institutionalist approach, since it construes objectivity as a collective-level property of a community of inquirers (Longino 1990, 2001), and work on judgement aggregation investigates the epistemic properties of different kinds of procedures for moving from an individual positions to a collective position (List 2005, List and Pettit 2011). There is also an important role for thinking about problematic group-level properties like polarisation (O'Connor and Weatherall 2017), epistemic bubbles, and echo chambers (Nguyen 2020)

2.3. Epistemic Justice

Individualist and Institutionalist approaches focus on articulating good epistemic states that ought to be pursued, typically considering idealised social structures. Alongside these idealised approaches that focus on characterising epistemic goods, it is vital that we take deidealised approaches that consider real-life social systems, and characterise the epistemic ills of our quotidian lives.[13] In this vein, Kristie Dotson (2014) has worked on the ways epistemic practices can undermine individuals' epistemic agency by excluding them from collective knowledge production, Miranda Fricker (2007) has investigated how credibility-affecting stereotypes can lead to individuals' testimony being rejected, and Charles Mills (2014, 2017) has offered an account of how institutions produce ignorance which supports a white supremacist social order.

Generalising a little, work in this tradition investigates both *epistemic oppression*— how political and institutional arrangements can undermine and restrict the epistemic agency of individuals (especially individuals from oppressed groups)— and the *inverted epistemologies* that produce ignorance in order to support oppressive social hierarchies.[14]

There may be a role for individual remedies to ameliorate these ills (Fricker 2007: C4), but an important part of the solution to oppressive systems of knowledge production will be for institutions and social practices to pursue the collective virtue of epistemic justice (Anderson 2012). As I propose to think about things, the structural virtue of epistemic justice involves at a minimum i) respecting the epistemic agency of all of its user or members, which may require taking action to protect and boost the agency of groups who are targets of epistemic oppression ii) ensuring that epistemic goods (such as education, knowledge, and credit) are not distributed in unjust ways (Fallis 2004, Coady 2010, 2017), iii) not contributing to (or being complicit in) ignorance-producing social practices that support oppressive

---

[13] On ideal and non-ideal theory in epistemology, see (Mckenna forthcoming).
[14] Mills points out that one can think of inverted epistemologies in veritist terms as social processes that reliably produce inaccurate beliefs about some subject-matter (Mills 2007).

social orders. For example, a scientific research group that realises the structural property of epistemic justice might respect the agency of its members by ensuring that all members can contribute knowledge, that their contributions are recognised and given equal credit[15], by giving all members some ability to shape the inquisitive agenda of the research group, and by avoiding contributing to inverted epistemologies by not pursuing research that contributes to racial hierarchies (for example by refusing to research facial recognition technologies (Buolamwini and Gebru 2018, Crawford and Paglen 2019)). Exactly what kinds of political institutions might realise the structural virtue of epistemic justice is abig question, but we might do worse than to follow Anderson's suggestion of looking towards well-functioning deliberative democracy (Anderson 2012 172, see also Anderson 2006).

In contrast to individualism's consequentialist framework, discussions of epistemic justice typically have a non-consequentialist spin, making reference to individuals being epistemically wronged (Fricker 2007), to inequalities of epistemic outcomes (Fallis 2004, Coady 2010, 2017), and to the role of epistemic systems in supporting political oppression (Mills 2007). We will return to this point in section 3.3.

3. Individualism, Institutionalism, and Epistemic Justice Involve Competing Goals

If social systems ought to promote the epistemic goals of individual epistemic goods, valuable institutional properties, and structural epistemic virtue, we might ask how these goals relate to one another. Is the pursuit of these different goals mutually supportive, or are there cases where pursuing one of the goals undermines the pursuit of another? In this section I will show that the pursuit of individual epistemic goods, good epistemic structures, and epistemic justice are in tension, creating three dilemmas for the critical epistemologist.

3.1. Individualism and Institutionalism

The independence thesis (Mayo-Wilson, Danks, Zollman 2011) states that the goals of good epistemic outcomes for individuals and valuable group-level properties can come apart.[16] To get a grip on some of the cases in which individual and collective outcomes come apart, let's consider the Zollman effect, and the social dilemmas that emerge with information cascades.

In a series of papers, Kevin Zollman has used agent-based modelling to explore the epistemic properties of different kinds of communication networks (Zollman 2007, 2010). Zollman considers the following situation: a group is gathering probabilistic

---

evidence about a question, which they have to act on in a series of tests where in each test they will apply the option they think is most likely to be successfully. One might think of medical researchers who are deciding whether to use a new treatment where the researchers engage in rounds of treatment where they decide to use the old or new treatment, and afterwards share results with some of their colleagues. The members of the group are connected by symmetric relations that allow them to share their results, and they update their beliefs about which option is better by applying Bayes' rule to their and their neighbours' results. Using computer modelling to simulate these networks, Zollman found that the degree of connectivity in the network creates a trade-off between the speed with which consensus is reached, and the reliability of this consensus. More connected networks (like the fully connected network on the right in figure 1) reach a consensus answer more quickly than less connected networks (like the wheel on the left), but are less reliable at reaching the correct result.[17]
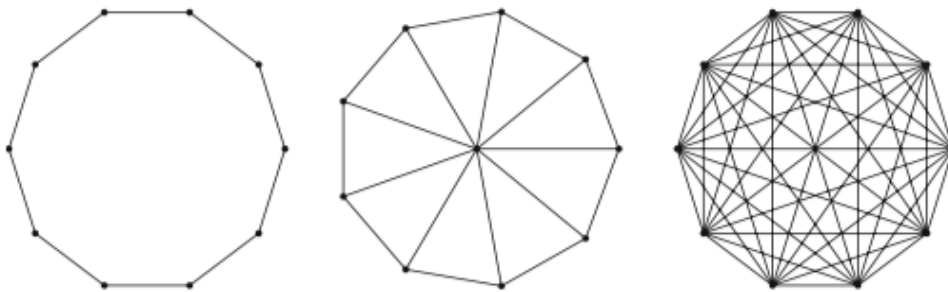

Fig 1. A cycle, wheel, and fully connected social networks (from Zollman 2007 p 579).

Zollman's gloss on this result is that more connected networks are less reliable because they are more likely to reach a preemptive consensus based on a stretch of misleading evidence. This result demonstrates the epistemic importance of transient diversity (temporary disagreement), which slows down the process of reaching consensus enough that inquirers collect enough evidence to form a reliable consensus (Zollman 2010). This result illustrates the independence theorem because what is better for individuals (having more connections and hence more evidence) is worse for the collective (because more connections lead to less reliable results).

A social dilemma occurs when what it is epistemically best for each individual to take one option, but every individual taking that option creates a worse outcome for the group.[18] There are various candidates for epistemic versions of social dilemmas. Kummerfeld and Zollman (2015) show that social dilemmas can occur for the choice of strategies of inquiry, creating situations in which it is better for individuals to pursue conservative strategies, and better for the group to include at least some risk-

---

[17] Some caveats: this result appears to only hold up in certain kinds of difficult problem situations (Rosenstock, Bruner, O'Connor 2017), and there are general reasons to take modelling results as hypothesis-generating for normative advice (Frey and Šešelja 2018).
[18] The possibility that information cascades involve free-rider dynamics was first noted by (Banarjee 1992, 816), and is explored by (List and Pettit 2004), and (Dunn 2018) (who argues that these situations ought not to be described as social dilemmas).

takers.[19] The literature on information cascades provides another kind of social dilemma where the choice for individuals is how many other people to be connected to.[20] Consider a model where a group is deciding on a binary factual question by majority vote. Each individual starts with the same prior belief, and then receives different evidence which is equally likely to be correct. They update on this evidence via Bayesian reasoning to reach a view about the probable answer to the question, and express this credence by voting for the option they think is most likely. Each individual desires that their own vote is correct. If none of the individuals in the model are connected, we meet the assumptions of the Condorcet Jury theorem and the majority position will be more reliable than the average accuracy of individuals within the group. However, if we take the same model and introduce social learning where individuals can share their votes, things look rather different. In a roll-call situation in which individuals vote in sequence, and each individual can see all of the prior votes, the group faces the threat of information cascades. If the first two voters have received evidence that favours option A, then even if the third voter has evidence that favours B, she will vote for A because the evidence from their peers outweighs their private evidence. In the roll-call situation the group's position an be determined by the luck of who gets to vote first, not by the aggregate of the group's evidence. In the roll-call setting, groups will systematically under-perform the unconnected groups. The choice between the unconnected and roll-call scenarios leads to a social dilemma: individuals will want to be in the roll-call scenario since being able to see previous votes increases their chances of voting correctly, but as the group becomes more connected its performance will degrade due to the increased chance of information cascades, meaning that everyone is worse off. This dilemma generalises: for models with intermediate levels of connection, the more connected the groups, the more reliable individuals are, and the less reliable the collective position is (Zollman forthcoming).

We've illustrated the possibility of social dilemmas using a mathematical model, and it might be that other factors complicate the picture for real-life groups. However, given the empirical evidence for information cascades,[21] and jury-theorem-type results[22], we ought to take the social dilemmas seriously, in social groups in general, and with social media platforms in particular. The important upshot for critical social epistemology is that when we design social systems with epistemic goals in mind, in some cases we will have to choose between promoting the epistemic states of individuals, and the epistemic states of the system considered as a collective entity.

3.2. Institutionalism and Epistemic Justice

---

[19] See (Dunn 2018) for a different kind of social dilemma involving conciliatory and steadfast responses to disagreement.
[20] In this paragraph and the next, I am drawing heavily on (Zollman MS, chapter 3). For a discussion of online information cascades, see (Sunstein 2017 C4).
[21] (Anderson and Holt 1997, Hung and Plott 2001, Kübler and Weizsäcker 2005, Ziegelmeyer et al 2010)
[22] (Miller 1996)

Let's now turn to consider the relation between institutional epistemic properties and epistemic justice. Here the conflict emerges because the ideally best institutional arrangements may not be the best response to non-ideal conditions. This point is a cousin of the problem of the second-best in economics: in non-ideal political conditions, trying to approximate the epistemically ideal institutional structures may lead to sub-optimal outcomes.

Nancy Fraser's critique of the Habermasian public sphere tradition in political theory presents us with a clear example of the problems with pursuing ideal epistemic structures in non-ideal situations (Fraser 1990). The public sphere tradition claims that the ideal for liberal democratic societies is a *unified* public sphere: an open forum for debate in which all citizens discuss issues of public concern on equal footing, bracketing their social power. Habermas's hope is that an open public sphere allows the unforced force of the better argument to determine the course of collective political action. The public sphere is a *political* ideal, concerned with the legitimate grounds of political authority, but it is also an *epistemic* ideal, concerned with the institutions that support the application of collective intelligence to political problems.

Fraser charges Habermas with the characteristic intellectual vice of ideal theorists (Mills 2005): reading their view of the ideal into their description of social reality. Drawing on revisionist histories of public discourse, she points out that Habermas' description of the bourgeois public sphere fails to reckon with the exclusionary aspects of the public sphere in actually existing democracy, leading him to overlook alternative discursive arenas in which marginalised groups organised to discuss and articulate issues of specific concern to these groups.[23] Fraser discusses a number of formal and informal barriers to participation in the public sphere, including restrictive conceptualisations of the political, discursive inequalities, and conceptual lacunae for expressing thoughts.[24] Although she doesn't put it in these terms, her central point is that in conditions on substantial epistemic oppression, a unified public sphere will compound prior political inequalities. This public sphere might be a laudable epistemic ideal for societies that have already achieved a substantial degree of political equality (although, see Fraser 1990, 68-69 on non-epistemic reasons for public*s* in egalitarian societies), but it is not an useful epistemic ideal for non-ideal political circumstances. The failure of the public sphere is both *political*—because it will not ameliorate political inequalities—and *epistemic*—because discourse in conditions of epistemic oppression, the public will fail to act on the political knowledge of people from subordinated social groups.

---

[23] Fraser draws from feminist organizing, and Squires (2006) broadens the scope to articulate the Black American tradition as a series of different kinds of counterpublics.

[24] Fraser is writing nearly twenty years before the mainstream literature on epistemic injustice but anticipating several of the central themes of Fricker (2007). Of course, work on epistemic oppression in Black Feminism antedates Fraser by at least a hundred years, and has been ignored by white feminists (Dotson 2012, Berenstain 2020). On Fraser's failure to deal with black counterpublics and for a more detailed typology of subaltern public spheres, see (Squires 2006).

Fraser contends that in non-ideal circumstances the epistemic ideal is precisely the subaltern counterpublic spheres which are overlooked by the public sphere tradition. In a well-functioning counterpublic, members of a marginalised group organise together to discuss issues of common concern which have been shunted out of the 'public' domain, developing the conceptual and epistemic resources that they need to express their concerns, contesting mainstream interpretations of their identities, and articulating the knowledge available from their social position. Although they are semi-public spaces of refuge, counterpublics are not enclaves (at least not by choice). The aim of a properly functioning counterpublic is to influence the public sphere by ensuring uptake for conceptual and epistemic resources.

There are two lessons here. First, that the goodness of epistemic systems needs to be evaluated relative to real-world conditions. Second, that mechanisms which are epistemically good in ideal conditions can become mechanisms of epistemic oppression in non-ideal conditions. This doesn't necessarily mean that institutional arrangements ought to be rejected because of their oppressive character (sometimes oppression is unavoidable, and sometimes it can be justified by appeal to other goods), but in order to properly evaluate epistemic systems we need to take epistemic oppression into consideration.

3.3. Individualism and Epistemic Justice

Individualism and epistemic justice come into conflict because as we noted above individualism presupposes a kind of consequentialism that is blind to epistemic oppression. Individualistic consequentialist approaches like veritism flatten out differences between different people and different questions and ignores important constraints on the pursuit of epistemic goods.[25] By contrast, the ideal of epistemic justice require us to think about the distribution of epistemic goods between different people and different questions, and how respect for peoples' epistemic agency might constrain the pursuit of epistemic goods. Just as consequentialism in the ethical domain fails to account for issues about the distribution of moral goods, and struggles to account for side-constraints to the pursuit of good outcomes, consequentialism in the epistemic domain fails to account for issues about the distribution of *epistemic* goods, and struggles to deal with side-constraints to the pursuit of epistemic goods. In the interests of simplicity, we will focus on how distributional and rights-based issues show up for veritism, but similar points could be made about other individualist approaches.

Don Fallis (2004), and David Coady (2010, 2017) have argued that Goldman's veritism fails to account the importance of distributions of epistemic outcomes between individuals. To get a grip on how distributional issues can emerge in epistemology, consider a toy example. Imagine two secondary school systems: system A assigns students to four schools based on location, and divides up its

---

[25] But see footnote 4.

funding equally, whereas system B assigns students to schools based on their performance in an entrance exam, and gives the 'high-performing' school half of its funding. The mixing of students in system A means that students in all schools gain reasonable increases to accuracy (in the range of +0.1 to +0.3), whereas the streaming in system B means that the students in the 'high performance' school gain huge increases to accuracy (+0.9), whereas the students in the 'low performance' school have less accurate beliefs when they leave the school (-0.1). On graduation, the students at each of the eight schools have the following average instrumental veritistic values with respect to the questions in the school curriculum:

System A	Instrumental Veritistic value	System B	Instrumental Veritistic value
A1	+0.2	B1	+0.9
A2	+0.3	B2	+0.3
A3	+0.1	B3	+0.1
A4	+0.2	B4	-0.1
Average accuracy	+0.2	Average accuracy	+0.3

Although system B has a negative effect on the students in its 'low-performance' school (B4), it is outweighed by the extremely effective education it delivers in the 'high-performing' school (B1). Because veritism evaluates based on a mean average, without paying any attention to the distribution of accuracy, it ranks system B as better than system A.[26] There is perhaps a sense in which school system B is better (just as an unequal society with high average welfare is *in some sense* better than a more equal society with lower average welfare) but system B is considerably worse from the perspective of political equality and epistemic justice.[27]

Fallis (2004) and Coady (2010, 2017) suggest that veritism ought to take account of the way changes in accuracy (Goldman's instrumental epistemic value) are distributed between agents. There are several different ways to represent the importance of distributional considerations: a strict egalitarian position would require the most equal distribution of veritistic value, a prioritarian position would require maximising the veritistic value of the least accurate agents, a sufficientarian position would require getting all agents above a threshold of accuracy, and an epistemic version of the difference principle would require maximising the veritistic value of the worst off.[28] Evaluating these options is beyond the scope of this paper, but we can assume that some kind of distributional consideration will matter for epistemic justice (Anderson 2012).

---

[26] We could get similar results if we switched to consider total veritistic value.
[27] Are constraints of epistemic equality generated by epistemic or ethical considerations? Following Fricker (2007), philosophers have situated epistemic justice in a liminal zone that overlaps epistemic and ethical considerations. If it turned out that epistemic justice was an purely an issue for ethics or politics, these conflicts would no longer be epistemic dilemmas, but they would remain dilemmas.
[28] To complicate things, several of these positions can be applied either to fundamental or instrumental veritistic value: strict egalitarianism might aim to ensure everyone's credences are equally accurate (fundamental value), or it might aim to equally distribute increases in accuracy (instrumental value).

Similar distributional issues emerge when we consider accuracy with respect to different questions. Consider two possible ways to teach the history of the United Kingdom: curriculum A is extremely accurate about medieval history, and the second world war, but doesn't cover the imperial history of Britain, and curriculum B covers all three issues with a reasonable but lower average accuracy. Veritism claims that curriculum A is better, but if we are interested in distributing knowledge about different subject-matters equally, curriculum B might be better. Considering active ignorance can also show us how inaccuracy on certain questions can exemplify epistemic oppression. Curriculum B's silence on British Imperial history would plausibly be part of the system of white ignorance that is endemic in the British education system (Bain 2018).

Individualism also faces problems dealing with side-constraints to the pursuit of epistemic goods. Structural epistemic justice requires respecting the epistemic agency of inquirers, generating a right on the part of individuals to participate in collective knowledge-production. This requirement can conflict with the goal of maximising average accuracy (or other epistemic goods). One way to ensure that an epistemic practice has higher average veritistic value is to stop people from participating in that practice if their participation will lower the average accuracy. Another toy example: we are in charge of a community debate group, and we are interested in increasing the veritistic value of that group. There are lots of benign ways to increase average accuracy: we might introduce debate rules, giving out a precis of research on the discussion topics. However, if including members of a minority group in discussions leads to other participants sharing persuasive misinformation about that group, one way in which we could increase accuracy would be to exclude the minority group from discussion. Exclusion from collective inquiry is not always significant—it's not a big deal if I am excluded from a private discord group about the best Buffy episode—but systemic exclusion from collective knowledge production does constitute an important infringement of epistemic agency.

The conflict between the consequentialist axiology of individualism and the egalitarian and rights-based aspects of epistemic justice demonstrates that the systems-oriented social epistemologist needs to pay attention not only to the amount of epistemic goods generated by social systems, but also to the way those goods are distributed between people and questions, and to the basic rights to participate in collective knowledge-production. Although I have presented examples that broke in favour of epistemic justice, in other cases maximising individual epistemic interests may win out. If the sole means of pursuing substantial epistemic improvement for individuals is to undermine the epistemic agency of a small group who are not otherwise marginalised, considerations of epistemic justice might be overridden by the epistemic interests of the majority.
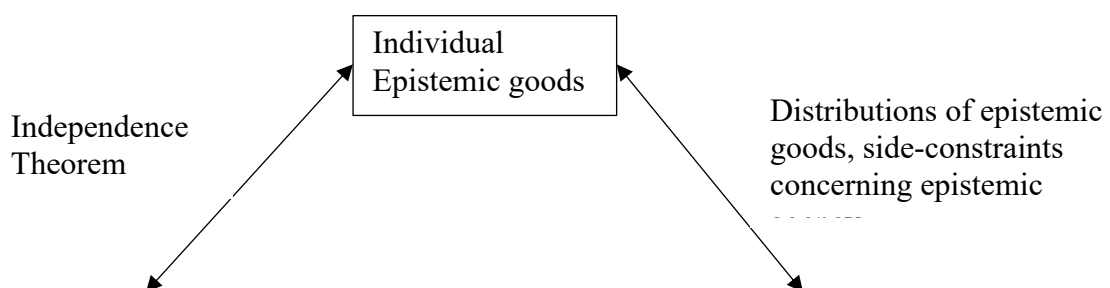
3.4. Epistemic Dilemmas?

The conflicts between the epistemic goals associated with individualism, institutionalism, and epistemic justice might be taken as a reductio of the view that institutions ought to aim at all three epistemic goals. At least at first pass, this objection is not convincing: these dilemmas instantiate general tensions in normative theory: i) between individual goods and collective goods, ii) between social arrangements in ideal and non-ideal situations, and iii) between consequentialism and constraints on the pursuit of maximising goods. The tensions we find here in the domain of epistemology are examples of much more general normative tensions.

One might take these tensions as evidence that social epistemology involves an element of non-rational choice between competing values. This inference is too quick. While these tensions cannot be resolved by appealing to epistemic values, we can appeal to extra-epistemic considerations – for example to ethical considerations. We should prioritise institutional epistemic goods over individual epistemic outcomes when the institutional epistemic outcomes are ethically important. For example, having an accurate predictions of climate change is extremely morally important, so there might be a good case for limiting communication amongst climate researchers if doing so would increase the accuracy of the consensus prediction. Similarly, applying ideal practices to a non-ideal situation is bad, when the bad epistemic outcomes are also ethically problematic, and the ethical-epistemic character of epistemic justice means that exclusions from collective knowledge-production, and unequal distributions of epistemic goods are bad (in part) because they are ethically bad.

I am not suggesting that epistemic value reduces to ethical value, but rather that ethical considerations can provide us with balancing considerations to resolve some epistemic dilemmas. Some epistemic dilemmas may not have an ethical dimension: whether we prioritise individual or collective epistemic outcomes regarding trivial questions is ethically unimportant. In these cases I think we should fall back on incommensurability, and accept that there is not rational way to resolve these choices. Appealing to moral considerations is not a simple way out of these epistemic dilemmas, but it makes clear what kind of choice we are dealing with.

4. The Social Epistemology of Social Media

Our discussion of the epistemic goals of social systems suggests that the epistemic evaluation of social systems involves a trio of epistemic dilemmas in which we weigh up the value of individual outcomes, institution-level properties, and structural epistemic justice. The dilemmas between these three values cannot be resolved by sole appeal to epistemic considerations, but only by appealing to ethical considerations.

Individual
Epistemic goods

Independence
Theorem

Distributions of epistemic
goods, side-constraints
concerning epistemic

```
┌─────────────────┐                              ┌─────────────────┐
│ Institutional   │ ◄──────────────────────────► │ Structural      │
│ Epistemic Virtues│                              │ Epistemic Justice│
└─────────────────┘                              └─────────────────┘
              Problem of the
              second-best
```
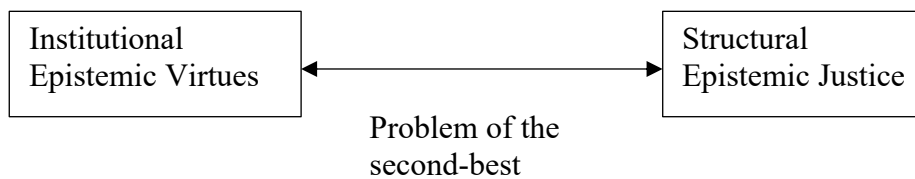
Figure 2: the Epistemic goals of social institutions.

With this picture of the epistemic aims of social institutions in place, we can consider how these goals and the tensions between them can help us to illuminate debates about the epistemology of social media. I will do this in two stages: first showing how discussions of the epistemology of social media appeal to these different epistemic aims (4.1), and secondly demonstrating how the tensions between these aims can illuminate some of the dilemmas involved in designing social media (4.2).

4.1. Individualism, Institutionalism, and Epistemic Justice

First, we will consider some debates about the design of social media that demonstrate that concerns about individual epistemic goods, collective epistemic virtue, and structural epistemic justice play a role in debates about how to design social media. These examples are picked because they are neat examples of appeals to different kinds of values, and because they help to illustrate the conflicts between different epistemic aims. I do not mean to endorse any of these arguments.

4.1.1. Individualism

Many changes to the rules and functionality of social media sites make appeals to individual epistemic goods.[29] We will consider two examples: Facebook's changes to the newsfeed in April 2019 to reduce the amount of 'false news', and Twitter's changes to the retweet function in October 2019.

In a press release from April 2019, Facebook announced a slew of changes to the platform that were designed "to fight false news".[30] These changes included: exploring the possibility of crowdsourced fact-checking, expanding the role of the associated press in fact-checking stories, reducing the reach of Facebook Groups that repeatedly share false stories, and incorporating a 'Click-Gap' signal into the newsfeed algorithm which promises to reduce the reach of domains which are more linked on Facebook than they are on the rest of the Internet. None of these

---

[29] (Silverman 2019), (Rosen and Lyons 2019).
[30] Facebook has a policy of using 'false news' as an umbrella term and avoiding 'fake news', but news sources coverage inevitably switch to 'fake news'.

interventions involve straightforwardly removing false content—a politically dicey proposal given right-wing concerns about social media censorship—rather they are designed to reduce the impact of false stories on Facebook users. Setting to one side blustering rhetoric about 'fighting' misinformation, the substance of this proposal is to implement changes to the newsfeed algorithm that demote certain groups and domains, in order to reduce the amount of false claims that show up on the newsfeed, with the proximal aim of ensuring that people who use Facebook come away with fewer false beliefs. Here Facebook is presupposing a kind of error-avoiding veritism according to which social media platforms are better, the less inaccurate they make their users' beliefs. It's important to remind ourselves that a developed veritist framework will consider both avoiding error and possessing the truth. A more consistent application of veritism would also consider what affects these changes have on the amount of *true* beliefs that people come away from Facebook with.

Social media companies are interested in other individual epistemic outcomes besides the accuracy of users' beliefs. In October 2019, as part of its preparations for the US election, Twitter announced changes to the how users could repost other posts. At this point, Twitter users could click the retweet button and choose to retweet without comment, or quote tweet with a prefatory message. In order to encourage more 'thoughtful amplification', the retweet function was changed so that pressing the retweet button led straight to the quote tweet screen (although it was still possible to retweet without a comment by leaving the text box empty). The hope was that by adding a nudge that added friction to retweeting would encourage users to add their own commentary, in order to "increase the likelihood that people add their "own thoughts, reactions and perspectives to the conversation.."[31] This change was reversed in December 2020. In a blogpost, twitter employees report that "we observed that prompting Quote Tweets didn't appear to increase context: 45% of additional Quote Tweets included just a single word and 70% contained less than 25 characters."[32] We can see this change as motivated by an individualistic interest in users making more justified contributions to conversation. By adding friction into the retweet function, twitter users were discouraged from quick and unjustified retweets, and were encouraged to take a second to think about retweeting by adding supporting (or rebutting) reasons to the claim that they were retweeting.[33] This example highlights the fact that although veritisitic concerns are central to individualism, individualists can be interested in other epistemic goods besides accuracy, such as whether beliefs are justified, and the quality of contributions to collective inquiry.

4.1.2. Institutionalism

---

[31] (Gadde and Beykpour 2020a)
[32] (Gadde and Beykpour 2020b)
[33] Although it is tempting to assimilate retweeting to assertion (Rini 2017), retweeting appears to be a distinct speech act that only sometimes involves indirect endorsement of the embedded claim (Marsili 2019), (Pepp, Michaelson, Sterken forthcoming).

Alongside concerns for individual epistemic goods, discussions of social media demonstrate concerns with institutional epistemic properties. We will focus on two particularly clear examples of institutional epistemic goods: Cass Sunstein's view that the ideal for social media is a unified public sphere, and Facebook's appeal to the value of connectedness.[34]

In a series of remixes of his book *Republic.com* (Sunstein 2001, 2007, 2017), Cass Sunstein raises worries about the ways in which a information technologies that involve 'filtering' information creates a fractured public sphere that undermines the epistemic workings of deliberative democracy. In *Republic.com* from 2001 his focus was on personalised news and niche websites. By 2007, *Republic.com 2.0* changed focus to blogs, and 2017's *#Republic* was squarely focused on social media.

In Sunstein's view, a well-functioning deliberative democracy requires supporting institutions; general-interest venues for disseminating information that establish a common ground of accepted beliefs and enable the serendipitous discovery of novel views. His concern is that information technologies that allow for personalised filtering herald the death of these institutions (he vacillates between concerns with current technologies, and worries about future developments). He worries that a media system organised around our interests as consumers will focus people's attention on special-interest forums, undermining the common ground through the filtering of information through search engines, hashtag-based social media, and personalised newsfeeds and replacing them with partisan news sources, and individualised information diets. Sunstein argues that this fragmentation of our epistemic landscape leads to a number of problems, including echo chambers, group polarisation, persistent false beliefs, cybercascades, a lack of shared experiences, and violent extremism. He canvasses a number of policy proposals for the internet including websites designed for civil deliberation, subsidies for public internet websites, encouraging viewpoint diversity on partisan websites, and 'opposing viewpoint' and 'serendipity' buttons for social media sites (Sunstein 2017: 213-233) .

Sunstein situates his work in a Republican tradition of American legal scholarship that grounds the value of free expression in the epistemic value of public discourse, but there is a strongly Habermasian flavour to Sunstein's critique (2017 46-7, 153-5, see also (Habermas 2006 fn3)). Like Habermas, Sunstein claims that the ideal for a deliberative democracy is a unified public sphere in which all citizens participate on a relatively equal footing, have shared experiences that support a common ground of beliefs, and citizens have chance encounters with a wide range of fellow citizens with different views.[35] From this perspective, the problem with special-interest websites, blogs, and social media is that they lead to a fractured public sphere—or perhaps a networked set of distinct public spheres (boyd 2010)—that undermine collective democratic problem-solving. Sunstein's critique of the contemporary

---

[34] For a book that applies tools from network epistemology to problems around misinformation, see (O'Connor and Weatherall 2019, especially chapter4).

[35] In other work, Sunstein has criticized Habermas, arguing that prediction markets are preferable to democratic deliberation (Sunstein 2006).

internet is robustly institutionalist. He frames his topic in with concerns about the availability of public space on the internet, and his central question is "what are the social preconditions of a well-functioning system of democratic deliberation? (Sunstein 2017, 5).

Institutional epistemic features also play a role in how social media companies present themselves to the public. Since at least 2010 Facebook has presented itself as promoting the social purpose of *connecting* users (van Dijck 2013). Here's Mark Zuckerberg's celebrating Facebook reaching two billion users in 2015:

> I'm so proud of our community for the progress we've made. Our community stands for giving every person a voice, for promoting understanding and for including everyone in the opportunities of our modern world. A more open and connected world is a better world. It brings stronger relationships with those you love, a stronger economy with more opportunities, and a stronger society that reflects all of our values.[36]

In this passage, Zuckerberg highlights the political and epistemic benefits of connectivity, for both individuals and societies. Over the years, Facebook's use of 'connectivity' has acquired an expansive set of associations, referring to many people are using Facebook, projects to provide free internet access, and a sense of social closeness. However, appeals to the value of connecting each user to a large number of other users remains central.[37] In a post from 2017 entitled *Bringing the World Closer Together*, Zuckerberg presents Facebook's history as centred around connectedness:

> I started Facebook to connect my college. I always thought one day someone would connect the whole world, but I never thought it would be us. I would have settled for connecting my whole dorm. We were just college kids. But we cared so much about this idea -- that all people want to connect. So we just kept pushing forward, day by day, just like you.[38]

It is no accident that Facebook is designed for users to become highly connected. José van Dijck argues that Facebook's rhetoric of social connectedness belies their connectivity-centred economic model (van Dijck 2013, see also Zuboff 2019). By designing a platform that incentivises users to connect with many others, along with features like the 'like' button which tracks users across the rest of the internet,

---

[36] (Zuckerberg 2015)

[37] In a memo from 2016, Facebook executive Andrew Bosworth claimed that connectedness was a fundamental value for Facebook, and appeared to present connectedness as a fundamental moral value:
> The ugly truth is that we believe in connecting people so deeply that anything that allows us to connect more people more often is *de facto* good. It is perhaps the only area where the metrics do tell the true story as far as we are concerned […]It is literally just what we do. We connect people. Period. (quoted in Mac, Warzel, Kantrowitz 2018)

[38] (Zuckerberg 2017)

Facebook is able to harvest more information about their users, which it can use to offer its customers highly targeted adverts.

Notwithstanding Facebook's commercial interest in a highly connected userbase, there is a substantive issue for epistemologists: is a highly connected group of people an epistemically virtuous social structure?[39] Zuckerberg suggests that highly connected groups promote 'understanding', and a stronger society, but as we shall see below (in section 4.2.1.), connectedness raises difficult issues about the relation between individual and collective epistemic value.

4.1.3. Epistemic Justice

In section 2.3., we distinguished three issues that come under the heading of epistemic justice: inverted epistemologies that produce politically-charged ignorance, issues of epistemic oppression, wherein people are undermined in their capacities as knowers and inquirers, and issues about the distribution of epistemic goods. Social media sites are socio-technical assemblages that combine technological systems with social practices, and that issues of epistemic justice show up both in the social and technological sites of these systems. In thinking about social and algorithmic oppression that occurs through social media sites, we should draw on the rich thread of work that uses an intersectional lens to explore the ways that other kinds of technological systems perpetuate both material and epistemic oppression.[40]

Starting with inverted epistemologies, we have an embarrassment of examples of social media sites abetting inverted epistemologies, both by permitting certain kinds of organising, and by providing the technological basis of white ignorance. As Safiya Noble points out in the epilogue of *Algorithms of Oppression* (Noble 2018), a significant part of the issue of misinformation—which is often ignored in discussions of 'fake news'—concerns technologically-abetted ignorance about marginalised groups. If sites fail to moderate racist hate speech (as the majority of social media companies have), they are effectively providing an organising service for white supremacists who have a history of being 'innovation opportunists' (Daniels 2009, 2018). White supremacist organising on social media is not merely permitted, but tacitly abetted by algorithmic systems which appear to promote racialised ignorance. For example, it has been suggested YouTube (which for our purposes can count as a social media site) has a bias towards 'extreme' content, creating a situation in which the recommended videos for fairly banal videos will quickly lead users down a pipeline of increasingly white supremacist content (Tufekci 2018, Alfano, Carter, and Cheong 2018, Munn 2019). If social media sites are designed for algorithmic self-

---

[39] There is an open question about how connected Facebook actually is, since the algorithmic timeline means that many Facebook users only see posts from a minority of people they are connected to.
[40] For example, work on technological redlining in search (Noble 2018), connections between technology and the carceral state (Benjamin 2019), inverted epistemologies in facial recognition (Boulamwani and Gebru 2018, Crawford and Pagin 2019).

radicalisation, the technologies undergirding social media are complicit in the inverted epistemology of white supremacy.[41]

Turning to epistemic oppression, we find no shortage of examples of marginalised groups on social media having their capacity to contribute to inquiry being undermined. Social media companies permit significant amounts of online harassment, which has disproportionate harms on women, especially women of colour, lesbian women, and trans women.[42] Sarah Sobieraj argues that the online harassment of women should be understood as "patterned resistance to women's public voice" (2018, 1701). She argues that this resistance employs common misogynistic strategies of intimidation, shaming, and discrediting that make gender—and in particular women's bodies—salient in online spaces.[43] She argues that the cumulative effect of misogynistic harassment is to "silence women, undermine their contributions to digital discourse, press them out of valued digital publics, and create a climate of self-censorship that mirrors the calculations women make in physical public spaces about what is safe and what is risky."(Sobieraj 2018 1709-10). In the jargon of epistemic justice, we can say that social media platforms are sites of epistemic oppression, and suggest that a significant proportion of the responsibility accrues to the companies who have failed to moderate harassment. Given the prevalence of technological redlining (Noble 2018) in other algorithmic systems, it would be no surprise to find that the algorithmic systems that sort timelines, suggest connections, and contribute to content moderation further undermine the epistemic agency of marginalised groups. For example automated systems for detecting hate speech produce high rates of false positives when used on African American Vernacular English (AAVE) (Sap et al 2019). These errors seem to emerge because the datasets used to train these systems incorrectly identified linguistic markers of AAVE as hate speech (Sap et al 2019, Davidson, Bhattacharya, and Weber 2019, Kim et al 2020).

We have only scratched the surface of the inverted epistemologies and epistemic oppression on social media, but these example hopefully indicate that achieving structural epistemic justice is a substantive challenge for social media, and that an important part of existing critiques of social media platforms is that they abet and enact epistemic oppression and inverted epistemologies.

4.2. Conflicts between Epistemic Aims in Social Media Platforms

4.2.1. The Independence theorem

In section 3.1., we saw that the independence theorem means that the methodological prescriptions and groups can come apart, and used two models to

---

[41] For an alternative take that stresses the role of social networks in radicalization, see (Lewis 2018).
[42] For an overview of statistics on online harassment in the united states, see https://www.womensmediacenter.com/speech-project/research-statistics
[43] Sobieraj focuses on the strategies of misogyny, but there will be further stories to be told about misogynoir and transmisogyny.

illustrate the idea that highly connected networks can be good for individuals, but bad for collective-level performance. We shouldn't mistake mathematical models for reality, but I think that there are fairly good reasons to think that currently existing social media sites instantiate this dilemma, in the sense that social media networks are good for individual epistemic outcomes, but bad for collective-level epistemic performance.

We saw in section 4.1.2. that Facebook presents itself as motivated by the desire to promote *connectedness* amongst its users, but at least involves providing each user with connections to many other users. This feature serves the commercial aim of extracting the maximum amount of behavioural data from users, but it also has epistemic virtues. In a highly connected network, individuals have access to a greater amount of information, can expect to hear of interesting news in a timely manner, and will be able to make use of coverage-supported reasoning (see Goldberg 2010 C6). However, the fact that highly connected networks provide epistemic goods for individuals does not mean that they are without problems at the institutional level. In the models from section 3.1., we saw that highly connected networks can reach consensus too quickly, and lose reliability due to informational cascades. This might lead us to speculate that part of the reason why highly connected social media sites like Facebook and Twitter remain popular despite their well-known epistemic ills is that they trade off individual epistemic goods, against our collective epistemic interest. We gain fast and reliable information about what our friends are up to, and about breaking news, but at the cost of information cascades and false but surprising stories spreading quickly across closely connected networks.[44]

We would need much more evidence from both modelling and social science to back up the suggestion that highly connected social media networks are good for individuals, but there is something appealing about this suggestion. By the time of writing it is widely accepted that social media have serious epistemic problems. Given this fact, why do so many people continue to use social media sites? Although there may be some mileage in explanations that appeal to addiction or irrationality, our more charitable suggestion is that individuals who are using highly connected social media sites are trading off the collective epistemic good against their own epistemic advantage.

4.2.2. The problem of the second-best

The conflict between designing social media for collective epistemic virtue and for structural epistemic justice shows up in debates about the online public sphere. We

---

[44] We can also find examples of the reverse of this tradeoff. The best candidates for sites that take advantage of the Condorcet Jury theorem are prediction markets and betting exchanges. While the outputs of prediction markets might be more reliable than the average user, these sites allow very restricted (or no) communication between users to ensure that bets are made as independently as possible, meaning that their collective performance comes at the cost of restricting individual-level evidence.

saw in section 4.1.2. that Sunstein suggests that the ideal for social media in a deliberative democracy is a unified public sphere in which citizens shared a common ground of accepted beliefs, and are exposed to an overlapping range of perspectives. The unified public sphere might be a good ideal for a society that has already achieved conditions of structural epistemic justice. But, as Fraser points out (see 3.2), in actually existing societies which are characterised by endemic exclusion and silencing in the public sphere, a unified public sphere would only compound epistemic oppression (see (Soberiaj 2018) for a version of this point). Applying Fraser's analysis to online discourse, we might suggest that for societies characterised by epistemic oppression, the proper ideal is not a unified public sphere realised by one social media site, but rather a network of counterpublic spheres representing an overlapping set of groups' interests.[45] Designing sites to cultivate a unified public sphere might promote the collective epistemic ideal, but it fails to reckon with the epistemic conditions of actual societies. If we want to promote the value of structural epistemic justice, then we ought to be thinking about how to design

Work in media studies has been heavily influenced by the public sphere tradition in political theory, and critical internet studies provides us with a rich set of potential examples of online counterpublics, including curated blocklists on twitter (Geiger 2016), Black twitter (Graham and Smith 2016, Brock 2020 C3), hashtag networks (Jackson, Bailey, Foucault Welles 2020), blogs (Knight Steele 2016a, 2016b, Gabriel 2015), and BlackPlanet (Byrne 2007). Online counterpublic spaces can play a role in resisting epistemic oppression by providing safe havens and support networks (Geiger 2016, Sobieraj 2018), allowing marginalised groups to contest media narratives (Knight Steele 2016a, 2016b, Gabriel 2015), and allowing marginalised groups toorganise (Jackson, Bailey, Foucault Welles 2020).

For the social epistemologist of social media the question is: is how can social media sites be designed to cultivate well-functioning counterpublics, enclaves, and satellite publics? This is partly a practical question, requiring experiments, drawing on case studies of successful and unsuccessful counterpublics, alongside normative work in social epistemology. Some of the questions that social epistemology can contribute to answering are:

1. How should content moderation function for counterpublic spaces, bearing in mind that dominant conceptual resources may misinterpret the speech of marginalised groups (Frost-Arnold, forthcoming), and that members of counterpublics may be targeted for systematic harassment?
2. How can sites designed for counterpublic spaces balance the need for privacy (to develop epistemic resources and avoid harassment), with the desire of counterpublics to influence the main public sphere?

---

[45] To be fair to Sunstein, he does note the importance of deliberative enclaves to public discussions, especially in context of epistemic oppression (2017 85-89, 254). However, his policy proposals focus on promoting a unified public sphere, and neglect the question of how to promote well-functioning counterpublics.

3. How can we distinguish between well-functioning counterpublic groups, and deliberative enclaves which stifle collective epistemic progress?

Identifying the tension between ideal epistemic structures and structures which contribute to ameliorating epistemic oppression is very much a first step, but it helps us to hold open space for projects to design social media sites that can support counterpublics, as well as addressing epistemic oppression in other ways.

4.2.3. Individualism and Epistemic Justice

In section 3.3., we saw that the consequentialist structure of individualist approaches to social epistemology means that individualism is unable to distinguish between different people and questions in ways that matter for structural epistemic justice, and that individualism countenances undermining the epistemic agency of marginalised groups if doing so is in the interests of the majority. In order to correct this gap, we need to think about epistemic justice, in particular about the distribution of epistemic outcomes, and the importance of protecting the epistemic agency of marginalised groups. As we noted above, conflicts between majority epistemic interests and epistemic justice can in principle break either way, and there may be some cases in which the epistemic interests of the majority justify some undermining of epistemic agency. However, in the majority of realistic cases the badness of epistemic oppression and very unequal distributions of epistemic outcomes trump maximising aggregate epistemic outcomes.[46] In this section, we focus on demonstrating the importance of enriching the maximising perspective of individualism with considerations of epistemic justice, by showing that in order to think about the design of social media we need to think about the distribution of epistemic outcomes between individuals and questions, and the ways in which social practices that have good epistemic outcomes can nonetheless undermine the epistemic agency of minority groups.

Above I suggested that issues about the distribution of epistemic outcomes—especially accuracy—between people can be an issue for epistemic justice. Social media companies give everyone who uses them access to the same basic services in the form of the public website, but these services may work differently for different groups of people, or be differentially accessible for different groups of people. Many social media sites are designed primarily for English and other European languages, meaning that they offer importantly different services to non-English speakers. These differences can have epistemic consequences: if anti-misinformation efforts are focused on posts in English, then users who do not speak English or primarily speak

---

[46] An analogy: public buildings which are inaccessible for people who use wheelchairs or other mobility aids are a bad thing from the point of view of inclusion and equal access. In the majority of cases ensuring equitable access comes at a reasonable cost, and does not decrease the ability of people who do not use wheelchairs from using public buildings. However, in some special cases the benefits of ensuring accessibility for minority groups decreases the overall accessibility of a building. If the belltower of a gothic cathedral can only be accessed by a steep circular staircase, making the tower accessible for wheelchairs (for example by replacing the stair with a lift) might mean that overall fewer people are able to access the tower.

another language will in aggregate come away from Facebook use with less accurate beliefs than English-speaking users. Many social media sites are also poorly designed for users with sensory impairments. For example, images on Twitter do not have automatic descriptions, and YouTube's automated captions have frequent errors and don't distinguish between different speakers. Failures of accessibility on social media means that users with accessibility needs will simply not be able to access information that is presented in inaccessible ways. Issues about the distribution of epistemic goods should be situated as part of a broader concern with design justice, concerning the ways in which technological systems distribute benefits and burdens differentially between different groups of people (Constanza-Chock 2020).

Besides concerns with the distribution of epistemic goods between individuals, we might also be concerned with the distribution of epistemic goods regarding different *questions.* Algorithmic newsfeeds typically promote certain kinds of content—posts from news sources, posts from friends, posts from smaller groups—and by so doing, they will prioritise users having accurate beliefs about different kinds of questions. Facebook's reorganisation of their newsfeed in 2018 prioritised posts from connected users over posts from public sources including news media and businesses.[47] One of the effects of this change is that using Facebook's newsfeed will have a greater impact on users' beliefs about what their friends are up to than their beliefs about current events. While it is possible that this change might be a good thing from the point of view of average accuracy (assuming that content about one's friends is more accurate that news stories), we might be concerned that this change means that Facebook fails to provide accurate information relative to the set of important questions we collectively care about.[48]

Finally, we ought also to be concerned about the ways in which pursuing majoritarian epistemic interests can indirectly lead to the epistemic exclusion of marginalised groups. There are an important class of social practices—which we might call epistemic accountability practices—which lead to good epistemic outcomes by allocating non-epistemic costs to people who act in ways that aren't socially beneficial. For example, we might promote the goal of people sharing accurate information by allocating social sanctions to people who are discovered to have shared false information. Epistemic accountability practices are extremely common online: we might think of call-outs that aim to embarrass a unreliable user, calls to unfollow persistently deceptive users, as well as actions taken by sites to restrict users who share false information. While well-designed accountability practices may be good from the point of view of maximising epistemic goods, they may have an oversized effect on users who are already marginalised. These users may be less able to bear the costs of accountability practices, may receive disproportionate negative effects for flouting shared epistemic norms, and may be

---

[47] (Mosseri 2018)
[48] The ethics of question-choice is an important theme in feminist epistemology of science, see (Anderson 1995).

more likely to be incorrectly accused of flouting epistemic norms. This unequal distribution of costs can have a number of negative effects: marginalised users may smother their own speech to avoid the costs of perceived norm flouting (Dotson 2011), avoid contentious topics, or simply move to different social media sites without such rigorous accountability practices. Self-exclusion from spaces of public discourse and testimonial smothering owing to accountability practices are doubly problematic: they undermine users epistemic agency by making discursive spaces inhospitable environments, and it removes the specialised knowledge of marginalised people, meaning that epistemic communities become less diverse (Frost-Arnold 2014). While I don't want to suggest that epistemic exclusion always trumps the pursuit of maximising good epistemic outcomes, acknowledging the costs of practices that are good for the majority of users complicates the assessment of social practices, highlighting the importance of centring marginalised groups in systems-oriented social epistemology.[49]

5. How to Design an (Epistemically) better Social Media

Critics of social media who are in the grips of the epistemic apocalypse narrative will often structure their critique around a specific epistemic vice of social media (the large amount of false posts, the existence of filter bubbles, algorithmic boosting of certain kinds of content), which their amelioratory proposals will centre around fixing that vice in a way that appeals to offline epistemic practices (add reputation systems, 'mix up' filter bubbles by adding random mixing of users, prioritise human editing over algorithmic filtering). While some of these proposals may be beneficial, our discussion of the conflicts between the different epistemic aims of social media highlights the importance of thinking about a number of different epistemic goals in a connected way. Adding a reputation system to a social media might help to reduce the amount of false information posted (Rini 2017), but it might also reduce the amount of true information posted, have chilling effects on the speech of marginalised groups, and undermine valuable transient diversity (Frost-Arnold 2014). Mixing up information may break up problematic epistemic bubbles (in the sense of Nguyen 2019), but it also runs the risk of undermining the formation of well-functioning counterpublics that rely on filtering out certain kinds of posts (on the role of blocklists in maintaining counterpublics, see Geiger 2016). Prioritising human editorial work over algorithmic filtering might help to ameliorate the algorithmic boosting of certain kinds of content, but it runs the risk of replacing it with human-driven inverted epistemologies (a particular problem given current labour practices for content moderation, see Frost-Arnold forthcoming). One of the takeaways from this discussion is that the epistemic evaluation of social systems is complex, and has an important ethical aspect pertaining both to the moral importance of structural epistemic justice, and the ethical issues involved in the choice of epistemic outcomes.

---

[49] This paragraph is heavily influenced by Karen Frost-Arnold's discussion of the epistemic benefits of anonymity (Frost-Arnold 2014). However, whereas Frost-Arnold articulates the harm of excluding marginalized groups in veristic terms, I focus on understanding this harm in terms of epistemic exclusion.

Looking beyond simple fixes, what lessons can we draw for the projects of ameliorating existing social media sites, and (more realistically) designing new kinds of social media? Our focus has been on understanding the normative framework for evaluating social media sites, and we are now in a position to offer a characterisation of what a (re)designing project that draws on critical social epistemology ought to look like.

- **The critical social epistemology of social media is an *ethical* project.** Choosing epistemic outcomes to promote presupposes ethical choices, and poorly designed epistemic systems can cause significant ethical harms. Many features of actually existing social media sites perpetuate significant epistemic oppression, and promote inverted epistemologies.
- **The critical social epistemology of social media is a *conflictual* project.** There are important tradeoffs between promoting individual epistemic outcomes, creating epistemically good institutional structures, and respecting epistemic justice. The designers of social media platforms need to be upfront about which aims they are aiming to promote, and about the tradeoffs that they are making, and they need to be accountable to the groups who stand at risk of being harmed by their designs (Constanza-Chock 2020). Accountability will require that social media companies are much more open about what users do on their sites, and about the black-boxed algorithmic features of their services.
- **The critical social epistemology of social media is a *pluralistic* project.** Social media sites can either be designed to balance off different epistemic aims, or they can be designed to promote specific epistemic aims leaving space for other sites to address the other aims. For example, it might be that highly connected and closely moderated sites are a good way to promote individual epistemic goods, whereas less connected sites that promote virtuous adverseriality, or involve voting mechanisms are a good way to realise collective epistemic goods, and more private sites aimed at different marginalised groups are a good way to promote counterpublic spaces. With that said, the exclusive pursuit of one value ought to have constraints: it would be bad if a site managed to realise group-level epistemic virtues by excluding all members of marginalised groups.
- **The critical social epistemology of social media is both a *social* and a *technological* project.** The epistemic features of a social media platform supervene on the interaction between its technological features and the social practices of its users (see section 1). Predicting these interactions will be complex, and will likely involve some degree of path-dependence, so there will be an important element of experimentation and exploration in (re)designing projects.
- **The critical social epistemology of social media is an *interdisciplinary* project.** This paper presupposes that normative theory, modelling of communication networks, and work in epistemology all have a role to play in (re)designing social media, but successful (re)design projects will need to

draw not only on philosophy, but also on (at least!) sociology, political theory, science and technology studies, critical internet studies, computer science, social psychology, as well as the work of advocacy groups. If we want to think about social epistemology in a non-ideal way, we need to grapple with the problems of actually existing epistemic systems.

- **The critical social epistemology of social media is a project of** *imagination.* The social media sites that are at the centre of public discourse (Facebook, Twitter, Instagram, YouTube) operate on a common model that promotes a highly connected user-base, algorithmic newsfeeds which are designed to maximise use, and minimal content moderation.[50] Part of our project in this paper has been to gather conceptual resources to frame different kinds of social media sites, and to go beyond the conservative project of reproducing offline practices online, but we have only really scratched the surface of possible design principles. What would a site look like which was designed around exploiting the Condorcet Jury theorem? What would a site look like which was organised around feminist approaches to objectivity? (Frost-Arnold forthcoming) What would a social media site designed around combatting inverted epistemologies look like? These are all open and important questions.

Contributors: Jazz Artemis, Zara Bain, Alex Barber, Chris Bertram, Ken Boyd, Liam Kofi Bright, Dan Cavedon-Taylor, Christina Chimisso, Luis Couto, Ana-Maria Crețu, Caterina Dutilh Novaes, Alexis Elder, Laura Engel, Alex Fernandez, Karen Frost-Arnold, Javier Gonzalez De Prado Salas, Moti Gorin, Samuel Hobbs, Maria Jimenez Buedo, Laura Kane, Geoff Keeling, Jason Konek, Aadil Kurji, Neri Marsilli, Alex Meaden, Rachel McKinney, Seiriol Morgan, Jesús Navarro, Charles Pence, Richard Pettigrew, Jon Pike, Mark Pinder, Chris Ranalli, Jeanine Schroer, Martin Sticker, Davide Taila, Tzuchein Tho, Sara Wheeler, Drazen Varga, Jesús Vega-Encabo, Alan Wilson, and Kevin Zollman.

References

Alfano, Mark, Carter, Joseph Adam & Cheong, Marc (2018). Technological Seduction and Self-Radicalization. *Journal of the American Philosophical Association* (3):298-322.
Anderson, Elizabeth (1995) Knowledge, Human Interests, and Objectivity in Feminist Epistemology, *Philosophical Topics*, 23: 27–58.
Anderson, Elizabeth (2006). The epistemology of democracy. *Episteme* 3 (1-2):8-22.
Anderson, Elizabeth (2012). Epistemic Justice as a Virtue of Social Institutions. *Social Epistemology* 26 (2):163-173.
Anderson, Lisa R, and Holt, Charles A (2007) Information Cascades in the Laboratory, *American Economic Review,* 87(5), 847-862.

---

[50] But see (Zuckerman and Rajendra-Nicolucci 2020) for an alternative ranking of the popularity of social media sites that highlights the popularity of other models.

Bain, Zara (2018). Is there such a thing as 'white ignorance' in British education? *Ethics and Education* 13 (1):4-21.

Barbrook, R., & Cameron, A. (1996). The Californian ideology. *Science as Culture*, *6*(1), 44-72.

Barlow, John Perry (1996) A Declaration of the Independence of Cyberspace, *Electronic Frontier Foundation* https://www.eff.org/cyberspace-independence

Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good news about Bad News: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition 3*(1), 1-9.

Bauerlein, Monika, and Jeffery, Clara (2020) Facebook Manipulated the News You See to Appease republicans, Insiders Say, *Mother Jones* published 21 October 2020, https://www.motherjones.com/media/2020/10/facebook-mother-jones/

Benjamin, Ruha. (2019). Race after technology: Abolitionist tools for the new Jim Code. *Social Forces*.

Berenstain, Nora (2020). White Feminist Gaslighting. *Hypatia* 35 (4):733-758.

Bishop, Michael A. (2005). The Autonomy of Social Epistemology. *Episteme* 2 (1):65-78.

Biss, Mavis (2013). Radical Moral Imagination: Courage, Hope, and Articulation. *Hypatia* 28 (4):937-954.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, *13*(1), 210-230.

Brock Jr, A. (2020). *Distributed Blackness: African American Cybercultures* (NYU Press.

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. Mit Press

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91).

Byrne, Dara N. (2007) Public Discourse, Community Concerns, and Civic Engagement: Exploring Black Social Networking Traditions on BlackPlanet.com, *Journal of Computer-Mediated Communication*, Volume 13, Issue 1, 1 October 2007, Pages 319–340, https://doi.org/10.1111/j.1083-6101.2007.00398.x

Castro, Clinton ; Pham, Adam & Rubel, Alan (2020). Epistemic Paternalism Online. In Guy Axtell & Amiel Bernal (eds.), *Epistemic Paternalism*. London: Rowman & Littlefield. pp. 29-44.

Coady, David (2010). Two Concepts of Epistemic Injustice. *Episteme* 7 (2):101-113.

Coady, David (2011). An Epistemic Defence of the Blogosphere. *Journal of Applied Philosophy* 28 (3):277-294.

Coady, David (2017) Epistemic Injustice as Distributive Injustice, in *The Routledge Handbook of Epistemic Injustice,* Ian James Kidd, José Medina Gaile Pohlhaus Jr (ads), Routledge pp.61-68

Congdon, Matthew (2015). Epistemic Injustice in the space of reasons. *Episteme* 12 (1):75-93.

Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press.

Crawford, Kate and Paglen, Trevor (2019) Excavating AI: The Politics of Training Sets for Machine Learning (September 19, 2019) https://excavating.ai

Crawford, Kate and Trevor Paglen, (2019) "Excavating AI: The Politics of Training Sets for Machine Learning (September 19, 2019) https://excavating.ai

danah boyd. (2010). Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In *Networked Self: Identity, Community, and Culture on Social Network Sites* (ed. Zizi Papacharissi), Routledge pp. 39-58

Daniels, J. (2009). *Cyber racism: White supremacy online and the new attack on civil rights*. Rowman & Littlefield Publishers

Daniels, J. (2018). The algorithmic rise of the "alt-right". *Contexts*, *17*(1), 60-65.

Davidson, T., Bhattacharya, D., & Weber, I. (2019). Racial bias in hate speech and abusive language detection datasets. *arXiv preprint arXiv:1905.12516*.

Dotson, Kristie (2011). Tracking Epistemic Violence, Tracking Practices of Silencing. *Hypatia* 26 (2):236-257.

Dotson, Kristie (2012). A Cautionary Tale: On Limiting Epistemic Oppression. *Frontiers: A Journal of Women Studies* 33 (1):24-47.

Dotson, Kristie (2014). Conceptualizing Epistemic Oppression. *Social Epistemology* 28 (2):115-138.

Dotson, Kristie (2018). On intellectual diversity and differences that may not make a difference. *Ethics and Education* 13 (1):123-140.

Dunn, Jeff (2018) Epistemic Free Riding in *Epistemic Consequentialism*.Ahlstrom-Vij, Kristoffer & Dunn, Jeff (eds.)   Oxford: Oxford University Press 310-332

Fallis, D. (2003, October). Social epistemology and the digital divide. In *ACM International Conference Proceeding Series* (Vol. 101, pp. 79-84).

Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social text*, (25/26), 56-80.

Frey, Daniel & Šešelja, Dunja (2020). Robustness and Idealizations in Agent-Based Models of Scientific Interaction. *British Journal for the Philosophy of Science* 71 (4):1411-1437.

Fricker, Miranda (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.

Frost-Arnold, Karen (2014). Trustworthiness and truth: The epistemic pitfalls of internet accountability. *Episteme* 11 (1):63-81.

Frost-Arnold, Karen (2018) Wikipedia, in Coady, D., & Chase, J. (Eds.). *The Routledge Handbook of Applied Epistemology*. Routledge. pp. 28-40

Frost-Arnold, Karen (forthcoming) *Who Should we be Online? A Social Epistemology for the Internet.*

Gabriel Deborah. (2015) Alternative Voices, Alternative Spaces: Counterhegemonic Discourse in the Blogosphere. In: Thorsen E., Jackson D., Savigny H., Alexander J. (eds) *Media, Margins and Civic Agency*. Palgrave Macmillan, London. https://doi.org/10.1057/9781137512642_2

Gadde, Vijaya and Beykpour, Kayvon (2020a) Additional Steps we're taking ahead of the 2020 US Election, *Twitter Blog* 9th October 2020, https://blog.twitter.com/en_us/topics/company/2020/2020-election-changes.html

Gadde, Vijaya and Beykpour, Kayvon (2020b) An Update on our work around the 2020 Us elections, *Twitter Blog* 12th November 2020 https://blog.twitter.com/en_us/topics/company/2020/2020-election-update.html

Geiger, R. Stuart (2016) Bot-based collective blocklists in Twitter: the counterpublic moderation of harassment in a networked public space, *Information, Communication & Society,* 19:6, 787-803, DOI: 10.1080/1369118X.2016.1153700

Gillespie, T. (2010). The politics of 'platforms'. *New media & society*, *12*(3), 347-364.

Goldberg, Sanford C. (2010). *Relying on Others: An Essay in Epistemology*. Oxford University Press.

Goldman, Alvin (2008). The social epistemology of blogging. In M. J. van den Joven & J. Weckert (eds.), *Information Technology and Moral Philosophy*. Cambridge University Press. pp. 111--122.

Goldman, Alvin (2010). Systems-oriented social epistemology. *Oxford Studies in Epistemology* 3:189-214.

Goldman, Alvin I. (2000). *Knowledge in a Social World*. Oxford University Press.

Graham R, Smith 'Shawn. (2016) The Content of Our #Characters: Black Twitter as Counterpublic. *Sociology of Race and Ethnicity* 2(4):433-449. doi:10.1177/2332649216639067

Habermas, J. (2006). Political communication in media society: Does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research. *Communication theory*, *16*(4), 411-426.

Holman, Bennett & Bruner, Justin (2017). Experimentation by Industrial Selection. *Philosophy of Science* 84 (5):1008-1019.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, *101*(46), 16385-16389.

https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html

Hung, Angela A, and Plott, Charles R (2001) Information Cascades: Replication and an Extension to majority rule and conformity-rewarding institutions, *American Economic Review,* 91(5) 1508-1520

Hwang, T. (2020). Subprime Attention Crisis: Advertising and The Time Bomb at The Heart of The Internet. *FSG Originals x Logic, Farrar, Straus and Giroux*.

Jackson, Sarah J., Bailey, Moya, and Foucault Welles, Brooke (2020) *#Hashtag Activism: Networks of Race and Gender Justice*, MIT press, Cambridge MA

Kim, J. Y., Ortiz, C., Nam, S., Santiago, S., & Datta, V. (2020). Intersectional Bias in Hate Speech and Abusive Language Datasets. *arXiv preprint arXiv:2005.05921*.

Kübler, Dorothea, and Weizsäscker, Georg (2004) Limited depth of reasoning and failure of cascade formation in the laboratory, *Review of Economic Studies* 71(2) 425-441

Kummerfeld, E., & Zollman, K. J. (2015). Conservatism and the scientific state of nature. *The British Journal for the Philosophy of Science*, *67*(4), 1057-1076.

Ledwich, M., & Zaitsev, A. (2019). Algorithmic extremism: Examining YouTube's rabbit hole of radicalization. *arXiv preprint arXiv:1912.11211*.

Levy, Neil & Alfano, Mark (2020). Knowledge From Vice: Deeply Social Epistemology. *Mind* 129 (515):887-915.

Lewis, Rebecca (2018) Alternative Influence: Broadcasting the Reactionary Right on YouTube, *Data&Society Reports,* https://datasociety.net/wp-content/uploads/2018/09/DS_Alternative_Influence.pdf

List, Christian (2005). Group Knowledge and Group Rationality: A Judgment Aggregation Perspective. *Episteme* 2 (1):25-38.

List, Christian & Pettit, Philip (2004). An epistemic free-riding problem? In Philip Catton & Graham Macdonald (eds.), *Karl Popper: Critical Appraisals*. Routledge. pp. 128-158.

List, Christian & Pettit, Philip (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press.

List, Christin & Goodin, Robert E. (2001). Epistemic democracy: Generalizing the Condorcet jury theorem. *Journal of Political Philosophy* 9 (3):277–306.

Longino, Helen E. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.

Longino, Helen E. (2001). *The Fate of Knowledge*. Princeton University Press.

Lynch, M. P. (2019). *Know-it-all society: Truth and arrogance in political culture*. Liveright Publishing.

Mac, Ryan, Warzel, Charlie, Kantrowitz Alex (2018) Growth At Any Cost: Top Facebook Executive Defended Data Collection In 2016 Memo — And Warned That Facebook Could Get People Killed, *Buzzfeed News,* March 29, 2018, https://www.buzzfeednews.com/article/ryanmac/growth-at-any-cost-top-facebook-executive-defended-data#.upw3jdyR8tw

Marsili, Neri (2020). Retweeting: its linguistic and epistemic value. *Synthese*:1-27.

Mayo-Wilson, Conor ; Zollman, Kevin & Danks, David (2013). Wisdom of the Crowds vs. Groupthink: Learning in Groups and in Isolation. *International Journal of Game Theory* 42 (3):695-723.

Mayo-Wilson, Conor ; Zollman, Kevin J. S. & Danks, David (2011). The Independence Thesis: When Individual and Social Epistemology Diverge. *Philosophy of Science* 78 (4):653-677.

Merton, R. K. and [Zuckerman, Harriet] (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56-63.

Miller, N. R. (1996). Information, individual errors, and collective performance: Empirical evidence on the Condorcet Jury Theorem. *Group Decision and Negotiation*, *5*(3), 211-228.

Mills, C. W. (2005). " Ideal theory" as ideology. *Hypatia*, *20*(3), 165-184.

Mills, C. W. (2007). White ignorance. *Race and epistemologies of ignorance*, *247*, 26-31.

Mills, C. W. (2014). *The racial contract*. Cornell University Press.

Munn, Luke (2019) Alt-right pipeline: Individual journeys to extremism online *First Monday*, 24(6) ttps://firstmonday.org/ojs/index.php/fm/article/download/10108/7920 doi: http://dx.doi.org/10.5210/fm.v24i6.10108

Mosseri, Adam (2018) Bringing People Closer Together, *facebook Newsroom* https://about.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/

Nguyen, C. Thi (2020). Echo chambers and epistemic bubbles. *Episteme* 17 (2):141-161.

Nguyen, C. Thi (forthcoming). How Twitter gamifies communication. In *Applied Epistemology*. Oxford University Press.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. nyu Press.

O'Connor, Cailin & Weatherall, James Owen (2017). Scientific polarization. *European Journal for Philosophy of Science* 8 (3):855-875.

O'Connor, Cailin & Weatherall, James Owen (2019). *The Misinformation Age*. New Haven, CT, USA: Yale University Press.

Oremus, Will (2018) The Great Facebook Crash, *Slate,* published June 27 2018, https://slate.com/technology/2018/06/facebooks-retreat-from-the-news-has-painful-for-publishers-including-slate.html

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

Pepp, Jessica ; Michaelson, Eliot & Sterken, Rachel (2019). What's New About Fake News? *Journal of Ethics and Social Philosophy* 16 (2).

Pepp, Jessica ; Michaelson, Eliot & Sterken, Rachel (forthcoming). On Retweeting
Rini, Regina (2017). Fake News and Partisan Epistemology. *Kennedy Institute of Ethics Journal* 27 (S2):43-64.

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.

Rorty, Amelie Oksenberg (1983). Imagination and power. *Social Science Information* 22 (6):801-816.

Rosen, Guy, Lyons, Tessa (2019) Reduce, Remove, Inform: New Steps to Manage Problematic Content, *Facebook Newsroom* April 10th 2019
https://about.fb.com/news/2019/04/remove-reduce-inform-new-steps/

Rosenstock, Sarita ; O'Connor, Cailin & Bruner, Justin (2017). In Epistemic Networks, is Less Really More? *Philosophy of Science* 84 (2):234-252.

Sap, M., Card, D., Gabriel, S., Choi, Y., & Smith, N. A. (2019, July). The risk of racial bias in hate speech detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1668-1678).

Shapin, S. (1989). The invisible technician. *American scientist*, *77*(6), 554-563.

Silverman, Henry (2019) The Next Phase in Fighting Misinformation, *Facebook Newsroom* April 10th 2019, https://about.fb.com/news/2019/04/tackling-more-false-news-more-quickly/

Smart, Paul (2018). Mandevillian Intelligence. *Synthese* 195 (9):4169-4200.

Sobieraj, S. (2018). Bitch, slut, skank, cunt: Patterned resistance to women's visibility in digital publics. *Information, Communication & Society*, *21*(11), 1700-1714.

Sobieraj, S. (2020). *Credible Threat: Attacks Against Women Online and the Future of Democracy*. Oxford University Press.

Squires, C. R. (2002). Rethinking the black public sphere: An alternative vocabulary for multiple public spheres. *Communication theory*, *12*(4), 446-468.

Srnicek, N. (2017). *Platform capitalism*. John Wiley & Sons.

Steele Catherine Knight (2016a) The Digital Barbershop: Blogs and Online Oral Culture Within the African American Community. *Social Media + Society*. October-December 1-10 doi:[10.1177/2056305116683205](10.1177/2056305116683205)

Steele Catherine Knight (2016b) Signifyin', Bitching, and Blogging: Black Women and Resistance Discourse Online, in *The Intersectional Internet: Race, Sex, Class, and Culture Online,* Safiya Umoja Noble and Brendesha M Tynes (eds) Peter Lang New York pp. 73-94

Sunstein, C. R. (2001). *Republic. com*. Princeton university press.

Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton university press.

Sunstein, C. R. (2018). *# Republic: Divided democracy in the age of social media*. Princeton University Press.

Sunstein, Cass R. (2006). Deliberating Groups versus Prediction Markets. *Episteme* 3 (3):192-213.

Tollefsen, D. P. (2009). Wikipedia and the Epistemology of Testimony. *Episteme*, *6*(1), 8-24.

Tufekci, Zeynep (2018) Youtube, The Great Radicaliser, *New York Times,* March 10, 2018

Van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press.

Weatherall, James Owen, O'Connor, Cailin & Bruner, Justin P. (2020). How to Beat Science and Influence People: Policymakers and Propaganda in Epistemic Networks. *British Journal for the Philosophy of Science* 71 (4):1157-1186.

Weisberg, Michael & Muldoon, Ryan (2009). Epistemic Landscapes and the Division of Cognitive Labor. *Philosophy of Science* 76 (2):225-252.

Wray, K. B. (2009). The epistemic cultures of science and Wikipedia: A comparison. *Episteme*, *6*(1), 38-51.

Yuval-Davis, Nira & Stoetzler, Marcel (2002). Standpoint theory, situated knowledge and the situated imagination. *Feminist Theory* 3 (3):315-333.

Ziegelmeyer, Anthony et al (2010) Fragility of Information cascades: An Experimental study using elicited beliefs, *Experimental Economics* 13(2), 121-145.

Zollman, Kevin J. S. (2007). The communication structure of epistemic communities. *Philosophy of Science* 74 (5):574-587.

Zollman, Kevin J. S. (2010). The Epistemic Benefit of Transient Diversity. *Erkenntnis* 72 (1):17-35.

Zollman, Kevin J. S. (2013). Network Epistemology: Communication in Epistemic Communities. *Philosophy Compass* 8 (1):15-27.

Zollman, Kevin J. S. (MS) *Network Epistemology*

Zuboff, Shoshana (2017) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Profile Books.

Zuckerberg, Mark, (2015). "1 Billion Facebook Users: An Important Milestone". *Zuckerberg Transcripts*. 443. https://epublications.marquette.edu/zuckerberg_files_transcripts/443

 Zuckerberg, Mark, (2017) "Facebook Post on "Bringing the World Together"". *Zuckerberg Transcripts*. 281. https://epublications.marquette.edu/zuckerberg_files_transcripts/281

Zuckerman, Ethan and Rajendra-Nicolucci, Chand (2020) Beyond Facebook Logic: Help us map Alternative Social Media! *Knight First Amendment Institute, Columbia University Blog* https://knightcolumbia.org/content/beyond-facebook-logic-help-us-map-alternative-social-media