In defense of (some) online echo chambers

Douglas R. Campbell[1]

Abstract

In this article, I argue that online echo chambers are in some cases and in some respects good. I do not attempt to refute arguments that they are harmful, but I argue that they are sometimes beneficial. In the first section, I argue that it is sometimes good to be insulated from views with which one disagrees. In the second section, I argue that the software-design principles that give rise to online echo chambers have a lot to recommend them. Further, the opposing principle, *serendipity*, could give rise to serious harm, in light of the conclusion of the first section that sometimes we are better off being insulated from some content online. In the third section, I argue that polarization can be a useful tool for inculcating the appropriate attitudes in a person.

Keywords Echo chambers · Social media · Epistemic bubbles · Polarization · Misinformation · Serendipity

Introduction

One of the most ethically troublesome developments of the Web 2.0 is the rise of the online echo chamber, also known in some philosophical work as a pernicious virtual community.[1] For the purposes of this article, an echo chamber is understood to be a community in which members are actively insulated from views with which they disagree *and* that these opposing views are actively discredited. In social-media contexts, echo chambers are often facilitated by the cost-free expulsion of dissident members and the filtering of what users see online, either manually or algorithmically.[2] Philosophers have argued that these communities generate different forms of harm. Firstly, there is an *individual* harm that is felt by a subset of participants in these communities: for instance, a person who has been misled by discussions in these communities might not trust a medical expert and might avoid life-saving medical treatment.[3] Even in communities formed around apparently benign beliefs, a person might feel a cost or a harm from their membership. Consider a person whose delusion that the Earth is flat has become so intense that they are estranged from their family members.

Secondly, there are *social* harms.[4] These harms are felt by everyone in society, even by those who do not participate in echo chambers. There is damage done to society when it becomes fractured, and there is damage done to our democratic institutions, since they rely heavily on common experiences or some shared ground to function well.[5] It is not possible to fully separate the social harms from individual

---

[1]　See Baym, 2011 for a broad ethical discussion of the Web 2.0. Parsell, 2008 examines pernicious virtual communities. Nguyen, 2020 distinguishes between echo chambers and epistemic bubbles. Epistemic bubbles are social structures in which opposing points of view have been omitted. Echo chambers, meanwhile, are communities in which opposing points of view have been discredited. This article concerns echo chambers first and foremost, and we will consider of the mechanisms by which echo chambers heap discredit on points of view, as Nguyen, 2020 catalogues them, but many of the considerations here also apply to epistemic bubbles.

✉　Douglas R. Campbell
　　campbelldr@alma.edu

1　Department of Philosophy, Alma College, 614 W Superior
　　Street, SAC 347, Alma, MI 48801, USA

[2]　See Nguyen, 2020: 142: "in epistemic bubbles, other voices are merely not heard; in echo chambers, other voices are actively undermined."

[3]　E.g., Parsell, 2008 argues for this claim.

[4]　See Parsell, 2008 and Sunstein, 2017 for discussions of social harms, especially the latter.

[5]　Sunstein, 2017: 140ff, especially the discussion of so-called *solidarity goods*, which are goods whose value increase when and because other people are enjoying them. Sunstein, 2017: 144 argues that "any well-functioning society depends on relationships of trust and reciprocity," which are undermined by echo chambers; Nguyen,

harms. For instance, the destabilization of democratic institutions hurts individuals who rely on these institutions to protect their rights and interests, and decisions to avoid vaccinations leads to bad health outcomes for the unvaccinated and for those around them who rely on herd immunization.

There is a third possible harm, although it is rarely explicitly thematized in philosophical work on echo chambers. This is the kind of harm that consists in having a false belief about something. To the extent that echo chambers inculcate false beliefs in someone and then make it difficult to disabuse a person of these beliefs, we can count this as a harm of membership in an echo chamber, too. Plato maintained in the *Phaedo* that the worst evil for a person is to hold a false belief.[6] More recently, some of the literature on misinformation has similarly identified the inculcation of false beliefs as a harm, but this position still remains rare.[7] It is not clear whether the alleged harm of holding a false belief is reducible to one of the first two kinds of harms, or whether there is something independently harmful about being wrong, which is likely what Plato meant by the claim.

This article does not dispute what has been argued about the harms of echo chambers, but it does maintain that recent research has been incomplete. Echo chambers are good in some cases and in some respects. I do not mean that they are merely *prima facie* good; I am not, for instance, arguing merely that participants in an echo chamber might *like* participating in it and, therefore, that we have some weak, *prima facie* reason for not objecting to the community. I mean that there are at least some cases of echo chambers that are, all things considered, good and worthwhile. (There are cases of merely *prima facie* good echo chambers, in addition.) Further, the underlying mechanisms of polarization and the rapid spread of information are in some respects sufficiently good that we should be grateful that they exist because they promote good outcomes, both for individuals and for society.

It is *not*, however, the goal of this article to establish that echo chambers overall are always more helpful than they are harmful. It might well be the case that the benefits outlined in this article are paltry in comparison to the harms that they cause. Nevertheless, it is important for us to appreciate these benefits in order for us to understand these communities fully. Consider an analogy with the emission of greenhouse gases. Environmental ethicists have, for decades, written about the moral problems posed by these emissions: whether individuals have any duties to limit emissions; the nature of climate change as a moral problem; the inequities concerning which counties have emitted the most greenhouse gases and which countries, in contrast, will suffer the most harm; and so on.[8] Yet, environmental ethicists are also aware of the fact that the reason why the emission of greenhouse gases poses such a difficult moral problem in the first place is that fossil fuels are such an easy and convenient source of energy that transitioning away from them too abruptly will harm those who are marginalized and vulnerable.[9] We emit so many greenhouse gases, which results in this moral catastrophe, *because* using greenhouse gases is good. In other words, researchers on this subject know that the emission of greenhouse gases is both good *and* bad, although surely, most, if not all, believe that it is *overall* bad and harmful, not good. A similar appreciation of the benefits and harms of online echo chambers is needed. The costs might not outweigh the benefits, but we cannot be sure until we have done the weighing, and we cannot do the weighing until we know what the benefits are. This article is about the benefits.

Accordingly, I argue that online echo chambers are good in some cases and in some respects. Firstly, I argue that there are some such communities where it is a good thing to be insulated from those with which one disagrees. Secondly, I argue that social-media companies such as Reddit and Twitter are guided by a design principle that gives rise to echo chambers and that this principle has a lot of merit. Thirdly, I argue that the psychological mechanisms that underlie group polarization and the rapid spread of information online are both reasonable and sometimes helpful.

2020 argues that echo chambers are built on sowing distrust of non-members in members.

[6] Specifically, he puts into the mouth of Socrates the view that this is "the greatest and most extreme evil" (*Phaedo* 83c; translation from Plato, 1999).

[7] Cf. Marin, 2021: 363–364, who says, when writing about misinformation in particular, that "the individual harm is that some people may acquire misleading beliefs as result of seeing misinformation shared by their peers." Marin, 2021 follows Sunstein, 2017 and Parsell, 2008 in also seeing collective harms, but she thinks that the collective harms consist in the pollution of the general ecosystem of information.

[8] See, e.g., Gardiner, 2011 on climate change as the perfect moral storm, Sinnott-Armstrong, 2005 and Hiller, 2011 on individual responsibilities, and Shue, 1999 on international environmental justice.

[9] Cf. Hursthouse, 2007: 169, who makes this point. She argues that although some environmental virtues, such as being correctly oriented towards nature, might require completely rethinking our approach to the economy, justice might demand that we *not* do this, on account of the fact that many poor people throughout the world might be harmed.

Section 1: insulation

I now argue that there are times when people *reasonably* prefer to be insulated from those who disagree with them.[10] For the most part, this section concerns insulation from views with which one disagrees. Echo chambers, however, do not merely insulate us from such views but also actively and intentionally discredit those views. Later in this section, I shall move from defending mere insulation in some cases to defending outright discrediting opposing views.[11] The clearest examples of this reasonable preference concern groups of marginalized people. Technologies that often get maligned in traditional discussions of echo chambers are useful and valuable here.[12] For instance, it is much easier to expel someone from a Reddit community than it is to expel someone from an in-person community, and this fact makes it easier for online administrators and moderators to expel someone who is not toeing the party-line in a given community. This is something to be thankful for when we are talking about protecting marginalized people.

We can illustrate this point with an example. Consider a person who has some minority status and who lives in a small rural town. This person might not know anyone with the similar minority status offline, and it might not be reasonably easy or even possible at all for this person to get in touch with such a person offline. In this case, the features of social-media platforms that allow for similar people to find each other come in handy. In fact, the possibility that someone with hateful or mean-spirited intentions might enter the community is easily obviated on account of the fact that expelling and banning this other person is much easier online than it is offline.

This insight turns the usual discourse around echo chambers on its head. For usually we hear that the like-minded people who are congregating on social-media platforms are themselves hateful and mean-spirited, and every attempt to discourage their behavior gets a person expelled and banned

from these communities. We might also observe that, in the past, there was virtually no chance that a Flat-Earther could meet someone else with the same delusions and reinforce each other's false beliefs, but today, it is lamentable that this occurs so easily online. This is sometimes true, but in other cases, echo chambers are valuable: it is good that marginalized people can meet and discuss their experience, with an almost entirely cost-free way of banning people from their community; it is similarly good that these communities allow people to overcome large geographical distances when their demographic is so small.

Along the same lines, cost-free expulsion and banning can help when there is no reasonable expectation that a person should have to engage with some subset of views. Consider a group of Holocaust survivors assembling. It is unreasonable that Holocaust survivors should have to let Holocaust-deniers into their community. Consequently, the survivors might know very little about the pseudo-historical arguments presented by the deniers and know even less about how to refute them. They most likely epistemically discredit the deniers by thinking that their positions are entirely baseless and are nothing more than dressed-up gibberish. The survivors' community conforms to the formal definition of an echo chamber, but nobody could call it unreasonable or objectionable. If an in-person meeting of Holocaust survivors were disrupted by some deniers, we would not object to the deniers being thrown out of the meeting, and we might even be indignant at the fact that the survivors had to go through the trouble of expelling the deniers. Online communities make this easier, and when some privacy settings are enabled such that deniers could not even *find* the group, expelling and banning someone is no longer even necessary. These are the same formal features of echo chambers that are often found alarming, but they are unobjectionable in the case of Holocaust survivors. The fact that the survivors have founded their group around something true and important deserves further attention, and we shall consider this point at greater detail later.

It is also helpful that online communities allow participants to escape social pressures.[13] This is not a statement

---

[10] Insulation is a crucial feature of echo chambers. Jamieson and Cappella 2008: 76 argue that echo chambers are rigidly closed-off communities that amplify the voices of their members and insulate them from rebuttals, which builds insulation into the very definition of echo chambers.

[11] While I follow Nguyen, 2020 in thinking that active undermining of views with which one disagrees is constitutive of echo chambers, by no means do I want to downplay the central importance of insulation in echo chambers. See Jamieson and Cappella 2008: 163–236 for a study of the way that echo chambers *isolate* us from people with whom we disagree; they hold this up as a crucial feature of echo chambers.

[12] This is one of the major ways in which online echo chambers leverage the internet's resources to facilitate their insulation. It is also a feature of every platform that echo chambers arise on, as far as I can tell, whether we are talking about Discord, Reddit, or something else. See Parsell, 2008: 45 for an interesting discussion of the punishment of dissent online.

[13] I am reminded here of the academic debate concerning friendships online and offline (see, e.g., the dialectic between Cocking and Matthews, 2000, on the one hand, and Briggle, 2008, on the other hand). Part of this debate concerns whether our online or offline friends know us best. Cocking and Matthews argue that our offline friends know us best on account of the fact they have access to aspects of ourselves that we filter out online. However, Briggle argues in response that we filter out things about ourselves in in-person interactions in order to conform to social pressures. Here is Briggle, 2008: 74's insightful example: "we can imagine an accountant who does not feel completely at home in any single life-context. Her office mates are friendly enough, but there is a great deal of political posturing and half-hidden competition. She likes her friends on her volleyball team, but here too there is historical baggage from awkward romantic relationships that became crossed.

that can be made universally, of course: some communities might enforce very stringent rules that mirror in-person social pressures, but some do not. In some cases, this is unfortunate. There is no doubt that sexist communities online flourish because the participants are made anonymous by default. Being a sexist offline generally but not always comes at a cost. Being a sexist online does not, especially in those communities that allow sexism to flourish. This is a result of the disapproval of sexist behavior in offline spaces, such as the workplace, whereas this disapproval might not exist on some social-media platforms. By the same token, there is sometimes a disapproval of marginalized lifestyles and identities offline. The absence of this disapproval on some social-media platforms is a *good* thing because it means that these marginalized lifestyles can flourish.[14] Again, we are noting the same formal features: the absence of disapproval means that the disapproved-of thing can flourish online; just as this is a bad thing in some cases, it is a good thing in others.

It behooves me, at this point, to clarify what I mean by 'good' and 'bad'. In this context, in which we are reflecting on the possibility of harms for individuals, a good thing is one that promotes our well-being, either mental or physical, and a bad thing is one that reduces our well-being, either mental or physical. For instance, having to listen to the Holocaust-denier is *bad* for the Holocaust survivor because doing so triggers, say, their post-traumatic stress disorder. It is *good* to be insulated from pro-anorexia communities as someone recovering from an eating disorder because these communities threaten the possibility of a full recovery.[15]

The last point worth making in this section is a reminder: echo chambers often come into being because of the distrust of institutions, professionals, and elites. This is true of virtually every prominent echo chamber: anti-vaccination communities are against the medical establishment; pro-anorexia communities are against the psychiatric establishment; QAnon communities are against the political establishment; Flat-Earthers are against NASA. We can multiply examples easily.[16] The dangers of this distrust are

clear: there are serious medical costs incurred, and violence is not far behind. The technology that makes expulsion and banning virtually cost-free is used to disastrous effect when, for example, medical doctors are expelled from a community that is denying the legitimacy of life-saving treatment.

However, we ought to remember that these communities often have good reason to be wary of certain institutions. The distrust did not come out of nowhere and nothing. The pro-anorexia online community is an example of a bad and harmful echo chamber, and it seems right that everyone in the community would do well to listen to their health-care providers. It also seems important to sympathize with members of the pro-anorexia community and to hear why they distrust the medical establishment. Their echo chamber is not a good or worthwhile one, all things considered, but to this extent it *is* good: namely, to the extent to which people with anorexia can find support among people who understand what they are going through, without having to engage with a medical establishment that they believe has failed them or made them feel crazy. They might be worried, specifically, about the possibility of compulsory treatment.[17] We can say the same thing for other echo chambers, *mutatis mutandis*. We might find, for instance, that communities that vigorously deny some fact that scientists have discovered were failed by the scientific establishment at some point in their education; maybe they were made to feel like that they had no place learning about science.

Earlier in this article, I explained that one constitutive feature of echo chambers is that they actively undermine opposing views. So far, I have merely argued that being *insulated* is a good and unobjectionable thing in some cases. This is a weaker notion than what we find in echo chambers, in which views are discredited, but it does not take far to get there from here. Let us return to the example of people with anorexia. We want such people to be insulated from pro-anorexia views; it is, all things considered, good to be so insulated. Moreover, it would be *even better* if they believed that pro-anorexia communities had nothing to offer or even that such communities were toxic, unhealthy, and predatorial. In many (and perhaps all) cases, this is true. It would be better for people who are recovering from eating disorders to think that there is simply no evidence for the beliefs of pro-anorexia communities. We ought to heap epistemic discredit upon such beliefs. We would probably achieve better health outcomes more reliably if we primed

---

Furthermore, there is a tenderness about her that she feels uncomfortable expressing in this group always alert to signs that a certain unspoken minimal level of toughness may be compromised. She likes her companions in her poetry club, but she feels restrained from expressing other aspects of her identity lest she tread on their air of serious contemplation."

[14]   In this way, we might say that it is good to be insulated *with* those whom one agrees, not just *from* those with whom one disagrees.

[15]   Here, I mean to reverse the arguments from Parsell, 2008, who argues that echo chambers, which he calls pernicious virtual communities, are harmful on the grounds that they promote bad health outcomes. They surely do sometimes, but I have argued that they can also sometimes promote *good* health outcomes.

[16]   One exception that has occurred to me is that of the community of *incels*, or involuntary celibates, or perhaps any sexist online

community in general. However, one of my undergraduate students at the University of Toronto insightfully pointed out that perhaps in these communities, women as a class of people or people who *are* having sex, unlike the incels themselves, are the elites that the community is reacting against.

[17]   For discussions of compulsory treatment of anorexic patients, see Giordano, 2010 and, especially, Griffiths and Russell, 1998.

the patients in question to be such that *if they do encounter pro-anorexia communities*, then they immediately dismiss them, instead of trying to take their views seriously and ending up engaging regularly with these communities. Virtually all of us live in communities in which Flat-Earther beliefs are seen as baseless in every way, and it is hard to see what is wrong with actively undermining the epistemic status of Flat-Earthers. What is remarkable about the anorexia example is that it shows that there are cases where not only is it permissible to actively undermine opposing views but that there are cases where it is *better*, to the extent that it promotes better health outcomes more reliably. The same could be said for other cases, too, such as anti-vaccination communities: we might want parents of young children to be insulated from people that say their children should not be vaccinated against polio; we would be even more satisfied if parents believed that the anti-vaccination communities were completely foolish and epistemically worthless.

Section 2: consumer sovereignty

Social-media platforms are built under the guidance of a design value that Cass Sunstein calls *consumer sovereignty*.[18] It is precisely the application of this design value by today's information technology that has given rise to echo chambers.[19] The best statement of consumer sovereignty comes from what turned out to be an accurate prediction made by Bill Gates in 1995:

> Customized information is a natural extension […]. For your own daily dose of news, you might subscribe to several review services and let a software agent or a human one pick and choose from them to compile your completely customized "newspaper." These subscription services, whether human or electronic, will gather information that conforms to a particular philosophy and set of interests.[20]

Eric Schmidt from Google similarly predicted: "it will be very hard for people to watch or consume something that has not in some sense been tailored for them."[21] The precise definition of consumer sovereignty is hard to pin down, but the general idea is that the consumer who is using information

technology such as social-media platforms should be *sovereign* in the sense that everything that they see online should be either explicitly chosen by them or chosen for them algorithmically on the basis of past revealed preferences. Another key feature of consumer sovereignty is that when something designed with this principle in mind decides what content to show a user, it does not take into account anything but the user's own preferences (or perceived preferences). It does not take into account what is good for society, for instance. Accordingly, consumer sovereignty as a design principle legislates that software ought to populate a site or an app with user-tailored content. In other words, everything that we see should be *tailored* to our preferences, as Schmidt predicted it would be. It would violate the idea of consumer sovereignty for some subset of what we see online to be something other than what we wish to see.

Mark Zuckerberg got at the idea of consumer sovereignty well when he stated that "a squirrel dying in front of your house may be more relevant to your interests than people dying in Africa," with the result that Facebook would show users the former, rather than the latter.[22] Similarly, two developers, explaining their design goals for Facebook in a blog post entitled "News Feed FYI: More Articles You Want to Spend Time Viewing," clarified that:

> We are adding another factor to News Feed ranking so that we will now predict how long you spend looking at an article in the Facebook mobile browser or an Instant Article after you have clicked through from News Feed. […] With this change, we can better understand which articles might be interesting to you based on how long you and others read them, so you'll be more likely to see stories you're interested in reading.[23]

This perfectly reflects consumer sovereignty as a design value. Sunstein does a commendable job explaining the way that the proponents of this idea are often "unself-conscious" about it, adopting it without much scrutiny on account of the way that our economic system is built around giving people what they say they want, and the way that this principle is destructive in many cases.[24] Sunstein's focus is particularly on the way that this undermines democratic institutions whose functioning requires shared experiences and values, whereas consumer-sovereignty-oriented technology allows people to have only those experiences that fit their own narrow set of preferences and values since it populates our apps with user-tailored content.

---

[18]   See Sunstein, 2017: 52 ff.

[19]   For the role of technology in the creation of echo chambers, see Pariser, 2011 and Watson, 2015.

[20]   Gates 1995: 167–168. For a discussion, see Sunstein 2017: 52–53.

[21]   Quoted in Holman W. Jenkins Jr., "Google and the Search for the Future," *Wall Street Journal*, August 14, 2010, http://www.wsj.com/articles/SB10001424052748704901104575423294099527212 (accessed January 10, 2023). For a discussion, see Sunstein, 2017: 53 ff.

[22]   From Pariser, 2011: 1.

[23]   See Sunstein 2017: 117–118.

[24]   Sunstein, 2017: 54 ff.

There are other problems with consumer sovereignty as a design goal, too. Firstly, it inherits all of the drawbacks of echo chambers since the principle facilitates these communities. With consumer sovereignty in mind, for instance, an ill person will not see content online that encourages treatment that the ill person is opposed to. So, just as there are the social costs that Sunstein describes, there are also the individual costs. Secondly, consumer sovereignty is advanced in practice by the use of privacy-violating technologies such as cookies.[25] Websites need to track us in order to record our revealed preferences, and while, hypothetically, they could do so in a way that respects our informed consent, in practice many of them do not, as ethicists who have written on this subject know well.[26]

However, even though consumer sovereignty as a design value has much speaking against it, there is also much in favour of it. In some cases and respects, it is valuable and worthwhile.

The clearest illustration of its appeal is in many of the cases described in the previous section. Holocaust survivors have a reasonable and unobjectionable preference to not see Holocaust-denying content. If we have the technology to satisfy this preference, it does not seem objectionable to promote content on their social-media apps that is about something other than denying the reality and evils of the Holocaust.[27] If these survivors want to search on a social-media platform specifically *for* Holocaust content, nothing in the idea of consumer sovereignty says that this should be prevented or impeded. The same goes for the other examples, such as those recovering from an eating disorder.

Moreover, the satisfaction of consumers' preferences would also lead marginalized people to avoid seeing hateful and bigoted content. This is a good thing about consumer sovereignty: if someone does not wish to see hate speech, then this preference is satisfied in line with the design value. We evaluate the idea of consumer sovereignty unfairly when we have in mind only those cases in which we are considering the satisfaction of preferences of vicious people. It is true that these cases exist, but consumer sovereignty also protects marginalized people by satisfying their own preferences. If a member of racial minority wants to specifically

search *for* racist content, then they have the ability to do that. However, it does not seem objectionable to satisfy their preferences for not seeing racist content when they hold these preferences.

The design value of consumer sovereignty also leads to a generally better e-commerce landscape. It is makes sense that consumers online should see products, including music, promoted that, in fact, fit with their preferences. This leads to better outcomes for all parties: the consumer benefits by being shown products that they have demonstrated some interest in; the vendor benefits by having their products shown to people who have demonstrated some interest in it; and the advertising platform benefits by being able to provide more value to the consumer through more-appropriate advertisements. Perhaps users do not always know what they want and so cannot reveal their preferences fully; in that case, the software designers might make predictions for them. Perhaps the software designers cannot make accurate predictions because the evidence that they have access to is too limited; in that case, we might have to suffer sub-par recommendations until better information-gathering tools are made available. At any rate, none of these objections yield the conclusion that the design value of consumer sovereignty does *not* lead to a generally better e-commerce experience.

There is a well-known objection to this view, based on a competing design value called *serendipity*, which I shall address shortly in this section. For now, let us note that there would be something absurd about *deliberately* promoting worse outcomes for everyone in the advertisement transaction when we could use our technology to promote better ones, simply by being guided by the idea of consumer sovereignty.

With that point in mind, it is worth reflecting on the fact that it does not seem like any other state of affairs could realistically obtain, even if we thought it was good to have some technology that occasionally showed people content that they *did not* want to see. Consumer-sovereignty-oriented technologies have become so popular precisely because they *did* satisfy people's preferences. These technologies make it easier for us to live with respect to media exactly how we live in every other respect: we try to satisfy our preferences. If we changed these technologies such that they showed people things that they did *not* want to see (or even *wanted to not see*), it is unlikely that these technologies would stay around much longer because the experience would be less pleasant; there would then be some resulting market demand *for* some technology that satisfies our preferences. Technologies that fail to satisfy our preferences have no real staying power in comparison with their competitors. (After all, *ex hypothesi*, these technologies would no longer be what we preferred.) It is odd that we would take

---

[25] For a discussion of cookies and privacy, see Spinello, 2011.

[26] See, e.g., Tavani, 2007.

[27] Campbell (forthcoming) discusses the promotion of content online as an example of a *nudge*. Nudges are features of contexts that promote the likelihood that we will make a certain choice but do not coerce us. Nudges are closely related to echo-chamber phenomena, because social-media platforms will often promote content that we agree with in order to nudge us to stay on their platform longer. Campbell cites examples of nudges that promote what is bad for us, such as the way in which Instagram nudges young girls towards eating-disorder content online because the software engineers know how powerful this content is at keeping girls hooked on their platforms.

seriously the view that preferences ought to *not* be satisfied, and it is hard to see how such a situation would be sustained: people *want* their preferences satisfied; that is what it means to hold a preference in the first place. The problem is that sometimes our preferences are bad, and their satisfaction would be evil, but it is unfair to lay the blame for this at the feet of consumer sovereignty as if were responsible for this, and it seems unrealistic to think that we could sustainably respond to bad preferences by no longer being guided by the principle that we ought to satisfy preferences. The principle of consumer sovereignty reflects how we live the rest of our lives.

This point gets at how we should respond to those people who champion the competing design value of *serendipity*, though serendipity is not without value.[28] To hold that serendipity has value is to hold that it is valuable to be exposed to something that one would not have wished or chosen to be exposed to ahead of time. For instance, people might find it valuable to be exposed to songs that belong to a genre that they have never listened to before. Serendipity is a design value that competes with consumer sovereignty, but it is possible to see both of them reflected in the same software. We might design a music-streaming service that mostly promotes songs that belong to a genre that the consumer most often listens to, which is the extent to which consumer sovereignty is reflected in the design. We might also add to the same streaming service the caveat that every tenth song is chosen from a genre that the listener has never listened to. So, one might get mostly country songs but also some opera. The operatic suggestions reflect the design value of serendipity. The listener would not have chosen to listen to opera, simply because he or she had never listened to it before; therefore, there was no basis on which a preference could be formed. The serendipitous suggestion had the effect of introducing the listener to something new.

It is not my goal to dispute the claim that serendipity has *some* value. It is plausible that the world is made a more interesting place by being exposed to what one would not have chosen to be exposed to. There is a sense in which everyone has some grounds for finding serendipity valuable: no matter what one is currently passionate about, there was a time in one's life before this passion had been discovered, and one could never have chosen to be exposed to it – because the ignorance was so great that there was no way that one could have had such a preference. However, I think that there are generally two reasons to react negatively to the discussions of serendipity that abound in the literature.

The first is that it is easy to overstate the value of serendipity. People such as Sunstein believe that serendipity

– more precisely, building into software platforms serendipitous encounters with the unfamiliar – will solve the harms of echo chambers, but this flies in the face of the research that his view about the harms of echo chambers is built on.[29] Imagine a staunch opponent of vaccines serendipitously stumbling upon pro-vaccination information. Defenders of serendipity in this context predict that this will effect a change of mind in the anti-vaccination user. For instance, Urbano Reviglio has said that serendipity has the "potential to prevent the threats of filter bubbles and echo chambers."[30] However, there is, to my knowledge, no research supporting this hope. In fact, the psychological research that I shall discuss shortly predicts just the opposite: people do not give up their views when presented with evidence that contradicts them, but they instead *double down* on their views.[31] That is why it is so crucial that we appreciate that echo chambers *discredit* opposing views: even when a person stumbles upon an opposing view, he or she will not recognize in it any legitimacy and will not investigate it.[32] What makes Sunstein's defense of serendipity so unusual in this respect is that he himself cites the psychological evidence in his work on this subject, and this aspect of human psychology is well-cited in the research on echo chambers.

Let me quickly survey some of this research. For rhetorical and dialectical purposes, I shall appeal to research that Sunstein himself is aware of. Consider one experiment by Brendan Nyhan and Jason Reifler in which participants in 2005 incorrectly believed that George Bush banned stem-cell research.[33] They then had this belief corrected by an article either from Fox News or by *The New York Times*. Conservatives adjusted their belief that George Bush had done this, but liberals did not. They would not change their position according to the new evidence. It made a significant difference, too, what the source of the correction was: conservatives tended to distrust *The New York Times*, and liberals tended to distrust Fox News. Nyhan and Reifler present this as evidence that attempts to correct misperceptions fail. Another experiment that they did supports a stronger conclusion: attempts to reduce misperceptions *backfire*. One such study involved measuring the effect that news of the Duelfer Report would have on conservative

---

[28] Sunstein, 2017 stands out as a leading champion of serendipity; so does Reviglio, 2019, who specifically cites serendipity's power to combat epistemic bubbles and echo chambers.

[29] Sunstein, 2017 is my target in the main text, but Reviglio, 2019's view falls prey to the same problem. Nguyen, 2020: 152 makes a point similar to mine: increasing the range of points of view to which we are exposed through serendipity is "useless or worse" because people will simply ignore or distrust these new points of view.

[30] Reviglio, 2019: 155.

[31] Sunstein, 2017: 93ff himself is aware of this evidence and discusses it.

[32] Note that this is precisely what we would want someone recovering from an eating disorder to do when stumbling upon a pro-anorexia community: ignore it and move on.

[33] This study was done by Nyhan and Reifler (2010: 320–322).

and liberals.[34] The Duelfer Report documented that Iraq did *not* have weapons of mass destruction before the United States of America invaded the country. Nyhan and Reifler found that news of the Duelfer Report caused conservatives, unlike liberals, to agree *more* strongly with the claim that Iraq had stockpiles of weapons of mass destruction. In other words, *the correction backfired*. Nyhan and Reifler infer that this result supports their view that information processing is goal-directed: we tend to "evaluate information with a directional bias toward reinforcing [our] pre-existing views."[35] In light of the way that corrections backfire, this effect has been called, by Nyhan, Reifler, and others, *the backfire effect*. This effect is observed in research settings in which participants are required to read the corrections; one wonders how little a difference a correction might make to changing someone's mind when a user of a social-media site serendipitously stumbles upon something that flies in the face of their beliefs. We are warranted to conclude that serendipity does not confer the benefits that defenders such as Sunstein hope it does.

Consequently, there just seems to be no reason to be optimistic about serendipity in a discussion of echo-chamber harms. There is one important qualification: Urbano Reviglio has argued convincingly that serendipity could reduce what he calls the *redundancy* of information in user-tailored environments.[36] This is important because one way that echo chambers actively discredit viewpoints is by skewing the pool of available evidence: a user believes that opposing views are baseless because he or she is never exposed to the fact that there *are* arguments for the opposing views. Exposure to opposing evidence might not cause one to change one's mind, and the correction might even backfire, but it might be better than believing that there is no countervailing evidence at all. It is more plausible to hold that serendipity might have some value in aesthetic contexts, such as when someone is exposed to a new genre of movies or music. Even then, I am not confident that this would amount to much beyond users being recommended things that they do not act on; it might well be the case that just as people dismiss views that they disagree with, they dismiss media that they are unfamiliar with.

The second reason to react negatively to serendipity is that it clearly in some cases leads to bad outcomes. We have already seen this above: there are some cases in which it is reasonable and unobjectionable for communities to be insulated. If we can avoid having Holocaust survivors see Holocaust-denying content unless they *prefer* to see that content, then it does not seem that there is any reason to expose them to this content. Yet, just as Holocaust survivors might serendipitously stumble upon an interesting documentary that they had never heard of, they might also serendipitously stumble upon some Holocaust-denying content. Those who defend the value of serendipity overlook the fact that some people have a reasonable and unobjectionable preference to *not* see some content. Serendipity seems like a poor pretext for recommending a marginalized person bigoted content. It also seems like a poor pretext for showing someone with an eating disorder pro-anorexia content because it would have been better to let someone recovering from anorexia *not* see that content.[37] Serendipity appears valuable when we consider the way that people often have unreasonable preferences to not see some content, such as when someone prefers not to see opera music suggested to them on their streaming app only on the grounds that he or she has never listened to it before, never really considered it, and does not really know what opera music is like. It seems to us that this person would benefit from having their worldview expanded, but it is helpful not to generalize from such cases because there are cases in which the preference for being insulated is reasonable and unobjectionable.

As I said, I do not want to dispute the value of serendipity in some respects: for instance, it truly is plausible that the world is a more interesting place when one is exposed to things that one would not have chosen to be exposed to ahead of time. However, the research on changing one's mind that people cite in discussions of echo chambers undermines the claim, often made by the same people, that serendipity will dissolve echo chambers, and serendipitously stumbling upon something bad for us is just as much a possibility as stumbling upon something good for us.

The last point to make about consumer sovereignty is a clarification: we should not think that being guided by the principle of consumer sovereignty will lead to outcomes that are good for everyone or that are, all things considered, good. One of the conclusions of this section has been just that there are some respects in which this design principle is good: for instance, it might lead parents of young children to join groups that strongly support vaccinating children against polio and that condemn as foolish anyone who thinks that children should not be vaccinated against polio. It might be bad for society overall for these factions to not take each other seriously. It might lead to the erosion of

[34] See Nyhan and Reifler (2010: 310–315).

[35] Nyhan and Reifler (2010: 307), who say that they "expect that liberals will welcome corrective information that reinforces liberal beliefs or is consistent with a liberal worldview and will disparage information that undercuts their beliefs or worldview (and likewise for conservatives)." Kunda, 1990 and Molden and Higgins 2005 are earlier reviews of this same effect.

[36] See Reviglio 2023.

[37] Much is often made of the way that echo chambers lead to bad health outcomes (e.g., by Parsell, 2008), but this is an example of how being serendipitously exposed to content that one did *not* want to see leads to bad health outcomes.

democratic institutions and norms. However, *in this respect* it is good: it leads to better health outcomes for the children of pro-vaccination parents, and it leads to better health outcomes for neighbors who rely on herd immunization against polio for some degree of protection. We might decide that the satisfaction of the preferences of individuals is bad for society or for democracy overall, but let us not overlook the way that failing to satisfy some preferences could lead to bad outcomes, even if the trade-off is worthwhile overall.

Section 3: polarization

There is a lot to dislike about polarization, which, as I understand it, is the process by which a person's beliefs become so strongly held that they will not even consider opposing views or the possibility of changing their minds. The specific form of polarization that I am concerned with here is *issue polarization*, which is the polarization of beliefs concerning particular issues, such as vaccinations. Many of our liberal-democratic institutions rely on people being open-minded and holding their beliefs not so strongly that they are resistant to compromise and discussion. Much of contemporary liberal-democratic theory relies on people of all backgrounds being able to collectively recognize something as providing us with a reason for action.[38] Those who warn us of the dangers of echo chambers do a good job marshalling the evidence that polarization is increasing, and it is increasing in a way that threatens the bases of our democratic societies.[39]

Polarization is not always bad. In some cases and respects, it is good. For instance, consider Western society's deep and universal aversion to slavery. This is a remarkable about-face for the Western world both in terms of its speed and its thoroughness. There is no major political party in the West that is in favor of slavery, and there is not even any minor political party in favor of it, to the best of my knowledge. There is no room at all on the political spectrum for a pro-slavery stance, and this is true even in light of the fact that it was not so long ago that Western countries openly practiced slavery, let alone were open-minded about it.

Few laypeople could state their opposition to slavery in a way that would stand up to philosophical scrutiny; fewer people could correctly report that they inferred their anti-slavery stance from evidence and reasons; and fewer still could muster arguments against, say, Aristotle's defense of slavery in the first book of the *Politics*. We would immediately distrust someone who takes a pro-slavery position seriously.[40] We heap epistemic demerit on such a person.[41] This speaks in favor of the view that some echo chambers are good, and that polarization is sometimes useful, not just *prima facie* but all things considered. For the world is a better place in virtue of the fact that societies that once openly practiced slavery now no longer even think that arguments in favor of it are worth taking seriously. It does not follow, of course, that such intense polarization regarding *every* true belief is desirable, but this example illustrates that there are some true beliefs that are so important to the correct treatment of human beings that intense polarization is worth having, and the more intense the polarization, the better.[42] This indicates that the drivers of polarization might be usefully incorporated into moral education: we could inculcate appropriate attitudes, appropriately intensely.

The example of slavery establishes that the content of the polarization matters for assessing whether the polarization in question is good or bad. It is also worth attending to the formal mechanisms of polarization in the absence of any content. These mechanisms often involve people acting reasonably.

One such mechanism is that in online communities, people tend to present evidence in favor of views that they already hold. People then form their beliefs and modulate the intensity of their beliefs in light of the pool of available arguments. The problem is that the pool of available arguments is skewed by a selection bias: only arguments that yielded a certain conclusion become available for people's consideration in the first place. The result is unfortunate epistemically because people get a mistaken sense of what the pool of available arguments looks like; however, the generation of the belief is formally acceptable because they are reasoning the same way that people reason whenever they form a belief on the basis that some consensus exists.[43]

---

[38]	I have in mind the sprawling literature on public reason; see, e.g., Billingham and Taylor, 2022.

[39]	See Sunstein, 2017 for a philosophical treatment of this subject, whereas Pariser, 2011 discusses this from the perspective of software technologies enabling the polarization, and Mason, 2018 conducts a wide-ranging survey of the nature and extent of polarization in the 21st century. Mason, 2018 in particular argues that the nature of political polarization in Western countries is destabilizing.

[40]	In fact, some philosophers have argued that we have moral duties to avoid holding and expressing some beliefs, which they call *contemplative, expressive, or doxastic duties*. Brennan and Freiman, 2020: 191–192 survey some reasons for thinking that this is the case, while arguing that moral philosophers run the risk of violating these duties.

[41]	Nguyen, 2020 unpacks echo chambers such that *epistemic demerit* is constitutive of them: non-members are regarded by members of an echo chamber as epistemically discredited or demerited; their points of view are not trustworthy.

[42]	To the extent that my argument here is that belief-formation practices that might sometimes strike us as objectionable are not, in fact, objectionable, see the recent research done on why belief in conspiracy theories is not as irrational as it is often made out to be; e.g., see Dentith (2016, 2017).

[43]	Sunstein, 2017 discusses this point in the context of misinformation; Nguyen, 2020: 144 calls this *bootstrapped corroboration*, putting the point well when he says that "in general, corroboration is often a

Ultimately, it is not a bad thing that people have some polarized beliefs, such as, again, about the evils of slavery. More to the point, we might consider harnessing the features of human psychology that lead to polarization to promote similarly intense polarization about similarly important topics. I have not attempted to refute the arguments that polarization on some topics is bad: of course, it leads to harm when, e.g., someone is so closed-minded about some medical treatment that they will not be moved to even consider it, with the result that they suffer some bad health outcome. Such a case does not suffice to show that polarization or even closed-mindedness is always bad. The widespread and abrupt anti-slavery polarization provides an important case study in moral education.

Conclusion

Online echo chambers are good in some cases and in some respects. There are three reasons for thinking that this is true. The first is that there are cases in which being insulated from certain views is reasonable and unobjectionable. The second is that echo chambers are built with a certain design value in mind – namely, consumer sovereignty – and this design value is appealing for many of the same reasons that make it unobjectionable for people to prefer to not be exposed to some content online. For instance, it is good for someone recovering from an eating disorder to not be exposed to pro-eating-disorder content online. Being exposed to such content, perhaps due to the alleged value of serendipity or some other reason, could lead to bad health outcomes, just as people who are concerned about echo chambers say that echo chambers lead to bad health outcomes. The third reason for thinking that echo chambers are good in some cases and respects is that the fundamental psychological mechanisms that underlie them are not always objectionable in form, and sometimes, the world benefits by being intensely polarized, such as by being intensely polarized against slavery.

References

Baym, N. (2011). Social Networks 2.0. In M. Consalvo, & C. Ess (Eds.), *The handbook of Internet Studies* (pp. 384–305). Wiley-Blackwell.

Billingham, P., & Taylor, A. (2022). A Framework for analyzing public reason theories. *European Journal of Political Theory*, *21 4*, 671–691.

Brennan, J., & Freiman, C. (2020). Moral Philosophy's Moral Risk. *Ratio*, *33*(3), 191–201.

Briggle, A. (2008). Real friends: How the internet can Foster Friendship. *Ethics and Information Technology*, *10*, 71–79.

Campbell, D. R. "Nudging and social media: The Choice Architecture of Online Life," *Giornale Critico di Storia Delle Idee Forthcoming*.

Cocking, D., & Matthews, S. (2000). Unreal friends. *Ethics and Information Technology*, *2*, 223–231.

Dentith, M. R. X. (2016). When inferring to a conspiracy might be the best explanation. *Social Epistemology*, *16*, 572–591.

Dentith, M. R. X. (2017). The Problem of Conspiracism. *Argumenta*, *5*, 1–16.

Gardiner, S. (2011). *A Perfect Moral Storm: The ethical tragedy of Climate Change*. Oxford University Press.

Gates, B. (1995). *The Road ahead*. Penguin Books.

Giordano, S. (2010). Anorexia and refusal of life-saving treatment: The Moral Place of competence, suffering, and the family. *Philosophy Psychiatry & Psychology*, *17*(2), 143–154.

Griffiths, R., & Russell, J. (1998). Compulsory treatment of Anorexia Nervosa Patients. In W. Vandereycken, & P. J. V. Beumont (Eds.), *Treating eating Disorders: Ethical, legal and personal issues* (127 vol.). New York University Press.

Hiller, A. (2011). Climate change and individual responsibility. *The Monist*, *94*(3), 349–368.

Hursthouse, R. (2007). Environmental Virtue Ethics," in Rebecca L. Walker and. In P. J. Ivanhoe (Ed.), *Environmental Ethics* (pp. 155–172). Oxford University Press.

Jamieson, K. H., & Capella, J. N. (2008). *Echo Chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press.

Jenkins, W. Jr. (2010). "Google and the search for the future," *Wall Street Journal*, August 14, http://www.wsj.com/articles/SB10001424052748704901104575423294099527212. (accessed January 10, 2023).

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Marin, L. (2021). Sharing (Mis)information on Social networking Sites. An exploration of the norms for distributing content authored by others. *Ethics and Information Technology*, *23*(3), 363–372.

Mason, L. (2018). *Uncivil Agreement: How politics become our identity*. University of Chicago Press.

Molden, D. C., Higgins, E. T., "Motivated, & Thinking (2005). In K. J. Holyoak, & R. G. Morrison (Eds.), *The Cambridge Handbook of thinking and reasoning* (pp. 295–317). Cambridge University Press.

Nguyen, C., & Thi (2020). Echo Chambers and Epistemic bubbles. *Episteme*, *17*(2), 141–161.

Nyhan, B., & Reifler, J. (2010). When Corrections fail: The persistence of political Misperceptions. *Journal of Political Behavior*, *32*(2), 303–330.

Pariser, E. (2011). *The Filter bubble: How the New Personalized web is changing what we read and how we think*. Penguin Press.

Parsell, M. (2008). Pernicious virtual Communities: Identity, Polarisation, and the web 2.0. *Ethics and Information Technology*, *10*, 41–56.

Plato. (1999). *Complete Works*. Edited by John Cooper and D.S. Hutchinson. Indianapolis: Hackett.

Reviglio, U. (2019). Serendipity as an Emerging Design Principle of the Infosphere: Challenges and Opportunities. *Ethics and Information Technology*, *21 2*, 151–166.

Reviglio, U. "Serendipity as a Design Principle of Personalization Systems: Theoretical distinctions," S. Copeland, W. Ross, & M. Sand (Eds.), *Serendipity Science: An emerging field and its methods* Cham, Switzerland: Springer, 145–166.

very good reason to increase one's confidence in the relevant beliefs." Both Sunstein, 2017 and Nguyen, 2020 are interested in cases where this process goes wrong. I agree that sometimes this is bad, but sometimes bootstrapping corroboration by ensuring that one sees only those whom one agrees with is not bad.

Shue, H. (1999). Global environment and International Inequality. *International Affairs*, *75*(3), 531–545.

Sinnott-Armstrong, W. (2005). It's Not My Fault: Global Warming and Individual Moral Obligations. In W. Sinnott-Armstrong, & R. Howarth (Eds.), *Perspectives on Climate Change* (pp. 285–307). Emerald Publishing Press.

Spinello, R. A. (2011). Privacy and social networking technology. *International Review of Information Ethics*, *16 12*, 41–46.

Sunstein, C. (2017). *#Republic: Divided Democracy in the age of Social Media*. Princeton University Press.

Tavani, H. (2007). Philosophical theories of privacy: Implications for an adequate online privacy policy. *Metaphilosophy*, *38*(1), 1–22.

Watson, J. C. (2015). "Filter Bubbles and the Public Use of Reason: Applying Epistemology to the Newsfeed," in F. Scalambrino (ed.), *Social Epistemology and Technology: Toward Public Self-Awareness Regarding Technological Mediation*, 47–58.