

Epistemic Paternalism Online

Clinton Castro
Department of Philosophy
Florida International University

Adam Pham
Division of Humanities and Social Science
California Institute of Technology

Alan Rubel
Information School and Center for Law, Society & Justice
University of Wisconsin-Madison

Abstract:

New media (highly interactive digital technology for creating, sharing, and consuming information) affords users a great deal of control over their informational diets. As a result, many users of new media unwittingly encapsulate themselves in *epistemic bubbles* (epistemic structures, such as highly personalized news feeds, that leave relevant sources of information out (Nguyen *forthcoming*)). Epistemically paternalistic alterations to new media technologies could be made to pop at least some epistemic bubbles. We examine one such alteration that Facebook has made in an effort to fight fake news and conclude that it is morally permissible. We further argue that many epistemically paternalistic policies can (and should) be a perennial part of the internet information environment.

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

1. Introduction

In 1995, Nicholas Negroponte introduced the idea of *The Daily Me*, a virtual newspaper perfectly custom-fitted to each reader's particular taste. As Cass Sunstein elaborates:

If your taste runs to William Shakespeare, your Daily Me could be all Shakespeare, all the time. [...] Maybe your views are left of center, and you want to read stories fitting with what you think about climate change, equality, immigration, and the rights of labor unions. [...] [W]ith the Daily Me everyone could enjoy an *architecture of control*. Each of us would be in fully charge of what we see and hear (Sunstein 2017, 1).

At the time, the Daily Me may have sounded like an improvement on that of traditional news. But we now know it has substantial drawbacks.

Consider the Michigan resident who received media attention for being “surprised to hear there was anything negative in the Mueller Report at all about President Trump” (Golshan 2019). The resident made her statement over a month after the Report, which indeed contained many negative revelations about the president and received much media attention for this, had been made available to the public (USDOJ 2019). When the resident made her statement, it was hard to imagine that any American could think that the Report was anything but negative for the President. Yet, she thought the Report exonerated him.

When the resident did learn that the Report was negative, it was through serendipity. At the time, the only Republican representative calling for impeachment on the basis of the Report happened to be hers. He held a townhall to share his thoughts about the Report. Were it not for

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

this, the resident would have gone on believing that there was no evidence that the President had obstructed justice.

The Michigan resident is not unique. Many of us have been made epistemically worse off through our exercise of the powers granted to us by *new media*, highly interactive digital technology for creating, sharing, and consuming information. Among its affordances, new media fosters *epistemic bubbles*, social epistemic structures that leave out relevant sources of information (Nguyen, *forthcoming*).¹ It does this by ferreting out our preferences and adapting to them automatically. As Eli Pariser explains,

The new generation of Internet filters look at the things you seem to like [...] and tries to extrapolate. They are prediction engines, constantly creating and refining a theory of who you are and what you'll do and want next. Together, these engines create a unique universe of information for each of us and [...] alters the way we encounter ideas and information (Pariser 2012, 9).

It's worse than this, though. New media also feeds us information designed to influence us based on those preferences and proclivities. That is, it doesn't merely reinforce our beliefs, it seeds and nurtures new ones (cf. Alfano et al. 2018).

¹ The term "epistemic bubble" bears some semblance to a more popular term, "filter bubble" (which was coined by Pariser). A *filter bubble* is a kind of epistemic bubble that is created by internet filters that attempt to show you what you want to see, by extrapolating from your past behavior and the behavior people like you (Pariser 2012).

YouTube, for example, will recommend videos to you based on what you have watched. This is done using a machine learning algorithm that is responsible for more than 70% of the time people spend on the site (Roose 2019). The YouTube algorithm is good at showing users what they are inclined to keep watching. For instance, watching videos from the liberal, progressive Young Turks is likely to lead to recommendations from CNN, and the channels of Barack Obama, Elizabeth Warren, and Bernie Sanders (Kaiser and Rauchfleish 2018). Viewing videos from the Fox News channel is likely to lead to recommendations from alt-right, men's rights, and conspiracy theory channels (Kaiser and Rauchfleish 2018). Given that two thirds of Americans get news through social media (with one in five getting news from YouTube) (Pew Research Center 2018b), it is no wonder we can consume a lot of news but be mis- or under-informed about current affairs of monumental importance.

Combine this with new media's immense popularity², and there is cause for concern. Consider "vaccine hesitancy." The reluctance or refusal to vaccinate has surged in recent years and made its way to the World Health Organization's list, "Ten threats to global health in 2019" (World Health Organization 2019). Much of the popularity of "anti-vaxxing," a movement that aims to spread vaccine hesitancy, is owed to new media (specifically, sites such as Facebook, Twitter, and YouTube) (Hussain et al. 2018). And now, two decades after the federal

² To give a sense of this: In the U.S., 69% of adults are on Facebook, 73% use YouTube, and 45% are on Twitter, and the median American uses three social media sites (Pew Research Center 2018a; 2019). Usage is even higher among teens. As of 2015, 97% of 12th graders used social media sites (Twenge 2017).

government declared it “eliminated,” measles has returned to the United States (Joy 2019).

Epistemic bubbles can have dangerous material consequences.

This raises questions about the responsibility new media developers have to alter the architecture of control that its users enjoy. This, in turn, raises questions about the latitude that social media developers have here. On the one hand, we might think developers should lead their users to consider a more diverse array of content than they currently do, even if that is not in line with users’ wishes. On the other, we might think that there is something objectionably paternalistic about this: Users should decide their information diets for themselves.

We will argue that because so much of the internet information environment is epistemically noxious, there is lots of room and opportunity for epistemic paternalism (acting to improve the epistemic lot of another, regardless of the others’ wishes). In fact, we argue, epistemically paternalistic policies should be a perennial part of the internet information environment. We proceed as follows. First, we motivate a framework for guiding developers’ changes to their technologies. We then use the framework to show that an epistemically paternalistic policy Facebook enacted to combat fake news on its site is permissible. We close with reasons for thinking that epistemically paternalistic policies like the one we discuss should be a common feature of internet information environments.

2. Unilateral Change

Developers make unilateral changes to their technologies all of the time. They must in order for their products to function in the dynamic social, economic, and technological environments in which they operate. But which unilateral changes are developers permitted to make?

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

Begin by considering two cases.

Mundane Update. In April of 2018, Facebook updated their plan to restrict data access on Facebook. This involved eliminating the Events API³, which enabled users to grant apps access to events they attended or hosted (which made it easy to add these events to one's calendar). The update phased out this functionality because the Events API allowed applications to access private information about users without their consent. The functionality will be accessible again, to applications that meet stricter requirements than the site previously had (Schroepfer 2018).

Emotional Contagion. For a week in 2012, Facebook ran a psychological experiment on 689,003 unwitting users. Some of these users had content with “positive” emotional content filtered from their experience of the site. Others had content with “negative” emotional content filtered out. The study showed that “that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness” (Kramer et al. 2014, 8788).

In thinking about the limits of unilateral decision-making, it is helpful to keep both of these cases in view. We take it that Emotional Contagion was morally wrong and that Mundane Update was not, and we think most will agree. We'll explain our judgments of these cases to shed some light on the ethics of making unilateral changes to large social media platforms.

³ API stands for Application Programming Interface. An API is a program that allows different applications to communicate to each other.

Why, exactly, was Mundane Update not morally wrong? We think that a good explanation is that it exposes no one to treatment they can reasonably reject.⁴ The only parties that might have a claim to rejecting the treatment they receive are the users of apps that depend on the Events API and the developers of those apps. Both parties are inconvenienced and have options taken away from them (the users cannot add events to their calendars automatically; the developers will eventually have to accommodate Facebook’s stricter requirements or cease to use the API). But these inconveniences are minor. No one seems to be harmed (the inconvenience does not rise to this level in either case), nor is anyone’s freedom or autonomy threatened (they are still able to do the underlying thing that matters, i.e., put important events in calendars, interface with Facebook, etc.). Further, the reasons that favor the update—i.e., protecting third parties from having their private information exposed without their permission or awareness—are weighty. Given that the imposition is minor and the claims in favor substantial, the update exposes no one to a treatment they can reasonably reject. So, it is permissible.

Using the same framework, we can explain why Emotional Contagion was morally wrong. Begin by asking what reasons the experimental subjects have for rejecting the treatment. Much of the criticism of the experiment focused on the idea that users may have been harmed by the experiment. One critic conjectured that Facebook could very well have killed users with their experiments, stating, “At their scale and with depressed people out there, it’s possible” (Goel, 2014). Perhaps Facebook did harm some users; at the very least it exposed them to the

⁴ This approach is inspired by Scanlon (1998).

possibility. Perhaps experimental subjects have reasons for rejecting the experiment in virtue of this. Perhaps. But this is not our central complaint.

Another, weightier reason experimental subjects have to reject the treatment is that it used them without their permission. There is a common idea, call it *Consent*, that there is a strong prima facie reason to avoid using people in ways they cannot consent to (cf. Korsgaard 1996; cf. O’neill 1989; cf. Kant 1998). Facebook’s experiment is clearly in tension with Consent. The experiment denied users the opportunity to consent by exploiting their assumption that they would be interacting with a normal version of the site and surreptitiously exposing them to a faux version of the site in order to run the experiment. It did this without asking them whether they wanted to be part of an experiment and without letting them know about the experiment, thus it used them in ways they could not consent to.

A possible objection to our assessment of Emotional Contagion is that *A/B testing*, the process of experimenting with different options in order to measure the relative success of each one (Deswal 2012), is known to be a common method for developing web-based products; thus, the experimental subjects were not used without their consent. According to this objection, Facebook didn’t exploit users’ assumption that they would be interacting with a normal version of the site. Rather, it did something they consented to by using the site, i.e., being exposed to different versions of the site in order to gain information about how it affects them and improve the site’s functionality.

We find this response unconvincing. In ordinary A/B testing, users are exposed to slightly different versions of a site, *each version of which is a candidate version of the site*. In the

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

case of the Emotional Contagion experiment, users were not exposed to candidate versions of the site. Instead, they were exposed to defective versions of the site. So, even savvy users who interacted with Facebook in full awareness that the site might be in A/B testing would have had their expectations upset in Emotional Contagion.

Now, the experimentee's claims against experimentation could be outweighed if the reasons for the experiment were weighty enough. But the reasons in favor of Emotional Contagion do not seem to be weighty at all. Facebook did not run the experiment because its users were at any substantial risk of harm (contrast this with Mundane Update, for example). Rather, the company was simply curious about how much positive and negative posts influenced their users.⁵ Further, Facebook had options available to them that did not involve surreptitious testing. As Facebook CTO Mike Schroepfer would later announce, “we should have considered other non-experimental ways to do this research” (Schroepfer 2014). Facebook also could have asked users if they wanted to volunteer for experimentation. Of the roughly one billion users it had at the time, 689,003 would have been happy to oblige.

Now that we have explained our judgments of Mundane Update and Emotional Contagion, we are in a position to step back and highlight a few features of the framework we

⁵ In a vague apology for the experiment, Facebook’s COO said, “This was part of ongoing research companies do to test different products, and that was what it was” (Gibbs 2014). In a blogpost promising more transparency and greater care in research, Facebook’s CTO said that the emotional contagion hypothesis was “important to look into” (Schroepfer 2014).

have used. In thinking about whether some intervention is permissible, it is helpful to think about whether anyone who will be affected by it will be exposed to treatment they can reasonably reject. When thinking about others' reasons for rejecting some treatment, it is helpful to think of whether the treatment will harm them, whether it respects their personhood, and what the reasons for the intervention are.⁶ These factors can push and pull against each other to make a treatment permissible or impermissible.

3. Demoting Fake News

Let us now consider a unilateral decision Facebook has made in their fight against fake news.

Demoting Fake News. In April of 2019, Facebook announced that it would use a new metric, Click-Gap, to determine where to rank posts in one's News Feed⁷. Click-Gap measures the gap between a website's traffic from Facebook, compared to its traffic from the internet at large. The idea here is that if the website has a lot of traffic from Facebook and nowhere else, the site probably has low-quality content. Click-Gap is part of a measure to demote low-quality content, such as fake news, in the News Feed and prevent it from going viral on the website (Rosen 2019).

⁶ There, of course, can be other factors and we cannot mention them all here (e.g., we have made no mention of considerations of justice or special relationships).

⁷ "News Feed is a personalized, ever-changing collection of photos, videos, links, and updates from the friends, family, businesses, and news sources you've connected to on Facebook" (Facebook 2019).

In this section, we argue that this measure is an instance of permissible epistemic paternalism. We begin by explaining why it is an instance of epistemic paternalism. We then argue that it is permissible, despite being paternalistic. We close with reasons for thinking that epistemically paternalistic policies like this one should be common.

A *prima facie* intuitive, standard definition of paternalism, owed to Gerald Dworkin (2017), is as follows:

Paternalism. *S* acts paternalistically towards *R* by doing (omitting) *Z* iff:

(C1) *Z* (or its omission) interferes with the liberty or autonomy of *R*.

(C2) *S* does so without the consent of *R*.

(C3) *S* does so just because *Z* will improve the welfare of *R* (where this includes preventing his welfare from diminishing), or in some way promote the interests, values, or good of *R*.

Were we to accept this definition, Demoting Fake News would not be an instance of paternalism because (C1) is not met.

It's worth pausing for a moment to explain why. For the purposes of this discussion, we will follow John Christman (2018) in understanding an individual as *autonomous* when she is not directed by doxastic or conative attitudes that are not simply imposed externally on her, but are hers, authentically (i.e., she can reflectively endorse them), and *free* when she can act without external or internal constraints and has the resources to effectuate her desires. So, to determine

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

whether Demoting Fake News are paternalistic, we must ask whether they undermine users' autonomy or freedom.

Begin with autonomy. If the policy works, it will play a role in shaping users' doxastic attitudes (e.g., their beliefs about whether vaccines are safe). The crucial question, then, is whether the resulting attitudes will be authentic. It seems that they will be. Whatever effect the interventions will have on users' attitudes, it will be by way of shielding users from misinformation masquerading as information. Users affected by the policy will then be *prevented* from forming inauthentic attitudes. This is because people (for the most part) will think that their doxastic attitudes are justified and accurate. Now, there might be users who want to be anti-vaxxers, come what may. Perhaps they value being a member of the conspiracy-theorist community more than they value good epistemic hygiene. However, such users are not going to have their considerations changed by an intervention like Demoting Fake News.⁸ The touch of this intervention is far too light for that. So, (C1) is not met in the case of Demoting Fake News in virtue of the effects that it has on autonomy.

⁸ This isn't to say that the architecture of social media sites can't influence users in important ways. We think that the "technological seduction" that sites like YouTube exhibit (cf. Alfano et al. (2019)) *can* encroach on autonomy by, for example, seeding and nurturing convictions that either can't be endorsed upon reflection *or* have been seeded and nurtured through methods that agents are alienated from.

We take it that a similar conclusion holds with respect to liberty. Demoting Fake News does not constrain what users can do. The policy may bring about a state of affairs that some might not like, namely one where an unvarnished experience of Facebook is not possible. But bringing about a state of affairs that is not desired hardly consists in constraining what one can do; and under this policy users can still post what they were able to post before, follow whomever they were able to follow before, and so on. So, the claim that Demoting Fake News limits users' freedom strains credulity.

Despite failing to meet (C1), we think that Demoting Fake News is an instance of a kind of paternalism. This is because we reject the common definition of paternalism. To see why, consider the following case from Shane Ryan (2016):

Smoke Alarm. Suppose that a mother is worried about the safety of her son in his new apartment. It's his first time living away from home and his mother knows that there is no smoke alarm in his apartment. She thinks that if she were to suggest that he get one, then he would agree, but knowing him as she does, she doesn't believe that he would actually get one. She knows that he is very proud of his new-found independence, and she thinks that if she were to ask him whether he would like her to buy him one he would say no. She decides to buy him one anyway and, by offering it to him already bought, tries to make his acceptance of it a *fait accompli* (Ryan 2016, 126).

In this case, (C1) does not obtain and, yet, the mother acts paternalistically towards the son.

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

That Paternalism is susceptible to counterexamples like Smoke Alarm is the primary reason⁹ that we are motivated to accept a less standard account of paternalism, owed to Ryan (2016):

Paternalism*. *S* acts paternalistically towards *R* by doing (omitting) *Z* iff:

(C1*) *S* does so irrespective of what *S* believes the wishes of *R* may be.

(C2*) *S* does so just because *S* judges that *Z* may or will advance *R*'s ends (her welfare, interests, values or good).

By the lights of this definition, Demoting Fake News could qualify as a paternalistic intervention, so long as it is motivated by Facebook's judgment that improving people's epistemic lot improves their welfare.

Now, the relationship between having true (or accurate, etc.) or justified (or rational, etc.) doxastic attitudes and one's welfare is complicated (for a survey of these issues see (Hazlett

⁹ But it is not the only reason. As Ryan (2016) points, there are other difficulties with Dworkin (2017)'s definition. (C3) states that *S* *Z*'s because *Z*'ing will improve the welfare of *R*. But, it seems more apt to say that *S* *Z*'s because *S* judges that *Z*'ing *may* improve *R*'s welfare. Further, (C2) seems to stand in need of correction, because we seem to be able to act paternalistically in cases where the paternalism is welcomed. To illustrate this, Ryan (2016) gives the example of a Victorian wife who has internalized sexist norms and is happy to let her husband make decisions for her. That the husband handles the wife's matters for her and for her own good seems paternalistic, even though she sees his handling of her matters as legitimate.

2013)). Knowing a fact can be irrelevant to our ends (e.g., knowing how far two randomly selected grains in the Sahara are from one another (Sosa 2000)). It can even be detrimental (e.g., knowing how a movie ends before you've seen it (Kelly 2003)). So, we find it helpful to couch the remaining discussion in terms of epistemic paternalism.

Epistemic Paternalism. *S* acts epistemically paternalistically towards *R* by doing

(omitting) *Z* iff¹⁰:

(C1**) *S* does so irrespective of what *S* believes the wishes of *R* may be.

(C2**) *S* does so just because *S* judges that *Z* may or will make *R* epistemically better off.

We will follow Kristoffer Ahlstrom-Vij (2013, esp. 50-61) in understanding someone as *epistemically better off* when she undergoes an epistemic Pareto improvement with respect to a question that is of interest to her. An *epistemic Pareto improvement* consists in an improvement along at least one epistemic dimension of evaluation without deterioration with respect to any

¹⁰ Note that our definition of epistemic paternalism breaks from the common understanding of the concept, owed to Ahlstrom-Vij (2013), which states that “a practice is epistemically paternalistic if and only if it interferes with the freedom of inquirers to conduct inquiry in whatever way they see fit [...] without consulting those interfered with on the issue of whether they should be interfered with in the relevant manner [...] for the purpose of making those interfered with epistemically better off (Ahlstrom-Vij 2013, 61). Our break from Ahlstrom-Vij is less a philosophical dispute than a difference of interest in focus.

other epistemic dimension of evaluation. Following Ahlstrom-Vij, we take it that *reliability*, having a high ratio of true-to false beliefs, is a dimension of epistemic evaluation.

By the lights of Epistemic Paternalism, Demoting Fake News is epistemically paternalistic. As Adam Mosseri, VP of News Feed explains, Demoting Fake News aims to fight fake news's effect of making the world less informed (Mosseri 2017). That is, the policy has been enacted because Facebook believes that they will make users epistemically better off (i.e. (C2**) is met). In explaining the responsibility Facebook has for doing its share in the fight, Mosseri likens Facebook's responsibility to the responsibility teachers have to fight misinformation. This, we think, is apt. Teachers are often epistemically paternalistic towards their students—they guide students' education in ways that are insensitive to students' wishes, in order to make the students epistemically better off. And, like teachers, Facebook would not change course were it to learn that Demoting Fake News was unwelcomed (i.e., (C1**) is satisfied). This is because of the kind of pressure Facebook is reacting to when it demotes fake news, i.e., pressure from the public and, increasingly, governments¹¹ to fight fake news. Having articulated the ways in which Demoting Fake News is paternalistic, let us now turn to the moral question: is it permissible?

To answer, consider the two constituencies that might have a claim against the policy: users of Facebook and purveyors of fake news.

¹¹ Germany is proposing a law to fine Facebook for advertisements containing fake news (Olsen 2016).

Begin with the users. What claim might they have against the policy? Well, it's not plausible that they are harmed, so they can't appeal to that. Nor is the policy in tension with Consent. Facebook has been open about its campaign to fight fake news, and in any event, it is reasonable to expect curators of news feeds to exercise discretion, especially against misleading or otherwise low-quality content.

Disgruntled users might retort that the opportunity to see more fake news than they would under this policy has been taken away without their consent. But, importantly, Consent does not speak against *affecting* people without their consent. Rather, it speaks against *using* them in ways they are not able to consent to. And using someone—as opposed to merely affecting them— involves intending their presence or participation to contribute to achieving the end (Kerstein 2013; Kerstein 2019; Scanlon 2008; Guerrero 2016). Removing the option to see more fake news does not violate this standard, because eliminating it does not intend the presence or participation of users who would like to have the option.

As a last recourse, users might appeal to autonomy or liberty. But as we argued above, the policies do not undermine users' autonomy or liberty; they do not instill in users inauthentic attitudes, nor do they limit users' abilities to effectuate their desires. In light of these considerations, it does not seem that there are users who could reasonably object to Demoting Fake News.

Let's, then, turn to the purveyors of fake news. They may claim that this policy harms them (or their businesses), and that this is morally relevant. It is hard to know how much weight, if any, this appeal should have. This is for several reasons. One is that fake news is an affront to

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

users' autonomy: it is misinformation masquerading as information, and thus doxastic attitudes based on it are inauthentic. Another reason is that the purveyors are left with alternatives under this policy. It leaves them free to promote vaccine hesitancy (or whatever cause they seek to advance), so long as their methods do not subvert user autonomy in the way just described. So, they are left with the options that allow them to minimize (or eliminate) any harm done to them. They just have to do it in ways that is more respectful to users.

Purveyors may try to appeal to Consent. We do not think they will get any traction here either. Consent, recall, stated that there a strong prima facie reason to avoid *using* people in ways they cannot consent to. Demoting Fake News does not use purveyors. It does not require the presence or participation of purveyors to contribute to achieving its end. To demonstrate this, it might be helpful to consider a policy that would use purveyors of fake news. Facebook could institute a policy where any person or organization caught promoting fake news was digitally strung up and shamed for reasons of deterrence. The deterrence policy in that case requires using the purveyors of fake news.

Finally, purveyors may appeal to considerations of autonomy or liberty. As with the appeal to Consent, it is difficult to see how purveyors' autonomy might be compromised, as Demoting Fake News does not involve anything that might affect purveyors' attitudes. The liberty of purveyors is not compromised either. We have already stated why: under Demoting Fake News, purveyors are free to advance the ideas they support or the profit they seek (they just might not be able to pursue these aims quite as effectively, if they do not change their current approaches).

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

Let us now turn to the reasons that speak in favor of Demoting Fake News. It is reasonable to think that it will prevent significant harms to individuals. Policies like Demoting Fake News have proved to be quite effective. For instance, in November 2016 Facebook, updated its Facebook Audience Network policy, which banned ads that contain or promote fake news (Wingfield et al. 2016). As a result, the sharing of fake news among users on the site fell by roughly 75% (Chiou and Tucker, 2018). Given that fake news is a driver of harmful movements, such as vaccine hesitancy, there is a strong consideration in favor of the policy. After all, those who wind up sick because of vaccine hesitancy are significantly harmed.

Harm reduction is not the only effect Demoting Fake News has. It also supports user autonomy. Consider:

A growing body of evidence demonstrates that consumers struggle to evaluate the credibility and accuracy of online content. Experimental studies find that exposure to online information that is critical of vaccination leads to stronger anti-vaccine beliefs, since individuals do not take into account the credibility of the content (Nan and Madden, 2012; Betsch et al., 2010, 2013; Allam et al., 2014). Survey evidence [...] shows that only half of low-income parents of children with special healthcare needs felt “comfortable determining the quality of health websites” (Knapp et al., 2011). Since only 12% of US adults are proficient in health literacy with 36% at basic or below basic levels (Kutner et al., 2006), Fu et al. (2016) state that [...] “low-quality antivaccine web pages [...] promote compelling but unsubstantiated messages [opposing vaccination]” (Chiou and Tucker, 2018).

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

Far from being disrespectful to agents' autonomy (something paternalistic policies are often accused of), epistemically paternalistic policies like Demoting Fake News are an important element of respecting it; the policy, if successful, will protect users from internalizing attitudes that would be inauthentically held.

So, Demoting Fake News is a policy no one could reasonably reject; further, there are compelling reasons to enact it. So, it is justified. In fact, it is for these reasons that Facebook *must* engage policies like Demoting Fake News: users who adopt unwarranted beliefs because of fake news and individuals who contract illnesses because of vaccine hesitancy have a very strong claim against Facebook's taking a laissez-faire approach to combating fake news on its site.

Since much of the internet information environment is epistemically noxious, there is lots of room and opportunity for interventions, such as Demoting Fake news, that are epistemically paternalistic. Hence, many epistemically paternalistic policies can (and should) be a perennial part of the internet information environment. What should we conclude from that? Well, one thing is that we should recognize that developers should engage in epistemic paternalism as a matter of course. Another is that our focus in evaluating epistemically-relevant interventions should not be on whether such actions are epistemically paternalistic. Rather, it should be on how they relate to other values (such as well-being, autonomy, freedom, and so on).

4. Conclusions and Caveats

What we have offered here are really only the opening moves of a longer discussion about how, exactly, to manage new media in ways that appropriately balances the interests of everyone involved. The policy we have discussed governs with a fairly light touch. But initiatives like

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

Demoting Fake News may be limited by the same fact that makes them justifiable, i.e., that they govern with a light touch. These policies may be able to pop epistemic bubbles, but they may not be able to dismantle sturdier structures, such as *echo chambers* (epistemic structures that, like epistemic bubbles, leave relevant sources of information out; but, unlike epistemic bubbles, actively discredit those sources (Nguyen *forthcoming*)). There is good reason to look into what we may do to chip away at these structures. Social media sites—8chan¹², which has been

¹² “an online message board that [...] characterizes itself as the “darkest reaches” of the online world in its tagline and has fostered a reputation as a nearly lawless space for free speech” (Stewart, 2019).

References

Ahlstrom-Vij, Kristoffer. 2013. *Epistemic Paternalism: A Defense*. Basingstoke: Palgrave

Macmillan.

Alfano, Mark, Joseph Adam Carter, and Marc Cheong. 2018. “Technological Seduction and

Self-Radicalization.” *Journal of the American Philosophical Association*, 4(3), 298-322.

doi:10.1017/apa.2018.27

Chiou, Lesley and Tucker, Catherine E. 2018. “Fake News and Advertising on Social Media: A

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

Study of the Anti-Vaccination Movement” SSRN, July 6, 2018.

<https://ssrn.com/abstract=3209929> or <http://dx.doi.org/10.2139/ssrn.3209929>

Christman, John. 2018. "Autonomy in Moral and Political Philosophy." *The Stanford Encyclopedia of Philosophy* Spring 2018 Edition.

<https://plato.stanford.edu/archives/spr2018/entries/autonomy-moral/>.

Deswal, S. 2012. “A Web Designer’s Introduction to A/B Testing.”

<https://webdesign.tutsplus.com/articles/a-web-designers-introduction-to-ab-testing--webdesign-9056>

Facebook. 2019. “News Feed.” Accessed August 1, 2019.

<https://www.facebook.com/facebookmedia/solutions/news-feed>.

Gerald Dworkin. 2014. “Paternalism.” *The Stanford Encyclopedia of Philosophy* Fall 2019.

<https://plato.stanford.edu/entries/paternalism/>.

Gibbs, Samuel. 2014. “Facebook Apologizes for Psychological Experiments on Users.” *The Guardian*. <https://www.theguardian.com/technology/2014/jul/02/facebook-apologises-psychological-experiments-on-users>

Guerrero, Alexander A. 2016. “Appropriately Using People Merely as a Means.” *Criminal Law and Philosophy* 10(4): 777–794. doi:10.1007/s11572-014-9346-x.

Hazlett, Allan. 2013. *A Luxury of the Understanding*. New York: Oxford University Press.

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

Hussain, Azhar, Syed Ali, Madiha Ahmed, and Sheharyar Hussain. “The Anti-vaccination Movement: A Regression in Modern Medicine.” *Cureus* vol. 10,7 e2919. 3 Jul. 2018, doi:10.7759/cureus.2919

Goel, Vindu. 2014. “Facebook Tinkers With Users’ Emotions in News Feed Experiment, Stirring Outcry.” *The New York Times*.
<https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html>.

Golshan, Tara. 2019. “Republicans’ Successful Campaign to Protect Trump from Mueller’s Report, in One Quote.” *Vox*. <https://www.vox.com/2019/5/30/18646048/republican-protect-trump-mueller-report-amash>.

Joy, Kevin. 2019. “What’s Causing the Latest Measles Outbreak?” *Michigan Health Blog*.
<https://healthblog.uofmhealth.org/whats-causing-2019-measles-outbreak-symptoms-perfcon>

Kaiser, Jonas and Adrian Rauchfleisch. 2018. “Unite the Right? How YouTube’s Recommendation Algorithm Connects the U.S. Far-Right.”
<https://medium.com/@MediaManipulation/unite-the-right-how-youtubes-recommendation-algorithm-connects-the-u-s-far-right-9f1387ccfabd>.

Kant, Immanuel. 1998. *Groundwork of the metaphysics of morals*. Cambridge: Cambridge University Press.

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

Kelly, Thomas. 2003. "Epistemic rationality as instrumental rationality: A critique." *Philosophy and Phenomenological Research* 66 (3):612–640.

Kerstein, Samuel. 2013. *How to Treat Persons*. Oxford: Oxford University Press.
doi:10.1093/acprof:oso/9780199692033.001.0001

Kerstein, Samuel. "Treating Persons as Means." *The Stanford Encyclopedia of Philosophy* Summer 2019. <https://plato.stanford.edu/archives/sum2019/entries/persons-means/>.

Korsgaard, Christine M., 1996, *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press. doi:10.1017/CBO9781139174503.

Kramer, Adam D.I., Jamie E. Guillory, and Jeffrey T. Hancock. "Emotional contagion through social networks." *Proceedings of the National Academy of Sciences* Jun 2014, 111 (24) 8788-8790. DOI:10.1073/pnas.1320040111.

Mosseri, Adam. 2017. "Working to Stop Fake News." *Facebook Newsroom*.
<https://newsroom.fb.com/news/2017/04/working-to-stop-misinformation-and-false-news/>

Negroponte, Nicholas. 1995. *Being Digital*. New York: Vintage.

Nguyen, C. Thi. (*forthcoming*). "Echo Chambers and Epistemic Bubbles." *Episteme*:1-21.

Obar, Jonathan A. and Oeldorf-Hirsch, Anne. 2016. "The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services."

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

TPRC 44: The 44th Research Conference on Communication, Information and Internet Policy.

<http://dx.doi.org/10.2139/ssrn.2757465>.

Olsen, Parmy. 2016. Germany Wants Facebook to Pay for Fake News. *Forbes*.

<https://www.forbes.com/sites/parmyolson/2016/12/19/germany-wants-facebook-to-pay-for-fake-news/>.

O'Neill, Onora. 1989. *Constructions of Reason: Explorations of Kant's Practical Philosophy*, Cambridge: Cambridge University Press. doi:10.1017/CBO9781139173773.

Pariser, Eli. 2012. *The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think*. New York: Penguin Books.

Pew Research Center. 2018a. "Social Media Use in 2018."

<https://www.pewinternet.org/2018/03/01/social-media-use-in-2018/>.

Pew Research Center. 2018b. "News Use Across Social Media Platforms 2018."

<https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/>.

Pew Research Center. 2019. "Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018." <https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/>.

Roose, Kevin. 2019. "The Making of a YouTube Radical." *New York Time*.

<https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html>.

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

Rosen, Guy. 2019. "Remove, Reduce, Inform: New Steps to Manage Problematic Content." *Facebook Newsroom*. <https://newsroom.fb.com/news/2019/04/remove-reduce-inform-new-steps/#reduce>.

Ryan, Shane. 2016. "Paternalism: An Analysis." *Utilitas*, 28(2), 123-135.
doi:10.1017/S0953820815000254.

Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.

Scanlon, Thomas M. 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge: Harvard University Press.

Schroepfer, Mike. 2014. "Research at Facebook." *Facebook Newsroom*.
<https://newsroom.fb.com/news/2014/10/research-at-facebook/>.

Schroepfer, Mike. 2018. "An Update on Our Plans to Restrict Data Access on Facebook." *Facebook Newsroom*. <https://newsroom.fb.com/news/2018/04/restricting-data-access/>.

Sosa, Ernest. 2000. "For the Love of Truth?" In Linda Zagzebski & Abrol Fairweather, eds. *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*. Oxford: Oxford University Press, 49-62.

Stewart, Emily. 2019. "8chan, a Nexus of Radicalization, Explained." *Vox*.
<https://www.vox.com/recode/2019/5/3/18527214/8chan-walmart-el-paso-shooting-cloudflare-white-nationalism>.

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

Sunstein, Cass R. 2017. *#republic: Divided Democracy in the Age of Social Media*.

Princeton: Princeton University Press.

Twenge, Jean M. 2017. *iGen: why today's super-connected kids are growing up less rebellious, more tolerant, less happy—and completely unprepared for adulthood (and what this means for the rest of us)*. New York: Atria Paperback.

United States Department of Justice. 2019. Special Counsel Report on the Investigation into Russian Interference in the 2016 Presidential Election. Washington, DC. March 2019.

https://www.justice.gov/storage/report_volume1.pdf;

https://www.justice.gov/storage/report_volume2.pdf.

Verma, Inder M. 2014. “Editorial Expression of Concern: Experimental evidence of massive scale emotional contagion through social networks.” *Proceedings of the National Academy of Sciences*. Jul 2014, 111 (29) 10779; DOI:10.1073/pnas.1412469111

Wingfield, Nick, Mike Isaac, and Katie Benner. “Google and Facebook Take Aim at Fake News Sites.” <https://www.nytimes.com/2016/11/15/technology/google-will-ban-websites-that-host-fake-news-from-using-its-ad-service.html>.

World Health Organization. 2019. “Ten Threats to Global Health.”

<https://www.who.int/emergencies/ten-threats-to-global-health-in-2019>.

This is a preprint of a chapter that appears in revised form in *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, edited by Amiel Bernal and Guy Axtell (London: Roman & Littlefield, 2020): pp. 29-44. This version is free to view and download for private research and study only. Please cite to the final version.

associated with the radicalization of many mass shooters (Stewart 2019), being a potent example—foster echo chambers that light-touch policies will likely be ineffective against. This raises questions about which interventions can dismantle echo chambers and which (if any) of them may be enacted.