Online filters and social trust: why we should still be concerned about Filter Bubbles

Andreas Falck[1*] & Kurtis Boyer[2]
[1]Department of Special Needs Education, University of Oslo
[2]Johnson Shoyama Graduate School of Public Policy, University of Saskatchewan
* Corresponding author: andreas.falck@isp.uio.no

Abstract

Eli Pariser's (2011) notion of a "Filter Bubble" describes the effect of social media filters tuned to predict what types of online contents social media users are likely to interact with, and subsequently presenting "more of the same" in order to maximise clicks. The Filter Bubble concept originally fuelled worries that users will find themselves in positive feedback loops, becoming exposed mostly to content that they already agree with, subsequently missing out on news and information that would contradict their pre-existing views. This initial worry has subsequently been challenged, by research showing that the views and sources that social media users are exposed to are actually quite diverse. Here, we argue that the original "Filter Bubble" theory, as well as subsequent criticisms, rest on a too simplified model of human belief formation, in which information content is over-emphasised at the expense of social dynamics. We argue that filter bubbles are still problematic, as they moderate peer feedback in a way that distorts how we evaluate information together with others.

Introduction

In platform-mediated online interactions, the information flow is often restricted in a novel and worrisome way: algorithms, tailored around the platform's instrumental goal to maximize clicks and sell ads, controls both the information that reaches the user, and whom do the user's output on social media (i.e. "posts") reach. This is a source of concern, not because the information flow is restricted per se (all media constrains the flow information in some way), but because it is designed to be undetected by the user. This sets it aside from most forms of online moderation, and many instances of censorship, where the user would at least be aware of how their information channels are tampered with. While we do not know the exact weights and criteria of the content-curating algorithms of platforms like Facebook and Youtube, Eli Pariser (2011) formulates the gist of their operation:

"Internet filters looks at the things you seem to like – actual things you've done, or the things people like you 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."

According to this model, the algorithms predict what kind of information the user is likely to interact with, and shows this kind of information more often. On this backdrop, Pariser goes on to define the *Filter Bubble* as an information-based phenomenon:

"Together, these engines create a unique universe of information for each of us – what I've come to call a filter bubble."

In short, the Filter Bubble according to Pariser (2011) is the information landscape resulting from the operation of social media algorithms whose goal is to maximise user interaction with the network ("engagement"), in order to sell clicks and ad space. The seeming consequence is that individuals are presented with a too restricted selection of perspectives and information, so that their pre-existing ideas are reinforced in a positive feedback loop. Thus, the main problem with the Filter Bubble is supposed to be the selection of information that the social

media user encounters: the users will simply not be exposed to potentially "good" ideas to a sufficient extent. This concern follows the tradition of previous research on misinformation in online and offline settings. A large empirical study by Flaxman and colleagues (2016) calls into question the very essence of the argument, as they showed that social media users are exposed to more diverse views and news than non-users. Following results like this, many authors have thus suggested that the initial worries about filter bubbles are unfounded (Zuiderween Borgesius, 2016; Bruns, 2019; Dahlgren, 2021). However, few commenters take into account the social context of belief formation, which is not so much about which information is available, but about which information to trust. Here, we will argue that it is the social dynamic of the online environment, and not the information landscape per se, that is affected negatively by click-maximizing social media algorithms. To do so, we must first discuss the social context of knowledge formation outside of social media contexts.

The Adaptive Features of Unmoderated Social Interaction

No human being would get along in their world without the aid of others. Knowledge is no exception: we rely on others to form knowledge, and culture implies that we build upon the knowledge of previous generations (Boyd & Richerson, 2009). Moreover, there exists a body of evidence that beliefs and sentiments that we share with others are privileged in human cognition. We encode information more strongly if we believe it to be attended to by others (Shteynberg, 2010). In addition, the valence of the information itself is inflated when it is shared with others: funny videos are judged as funnier (Fridlund, 1994), and persuasive political speeches are judged as more persuasive (Shteynberg et al., 2016). Research on groupthink (Turner & Pratkanis, 1998) and conformity (Baron et al. 1996) suggest that we tend to accept the beliefs that are salient in our social group.

The advantage of forming beliefs by drawing on those around us becomes apparent if we think about belief formation through the lens of evolutionary psychology. Humans have throughout history depended strongly on others within their social group for survival. While groups in pre-industrial settings were often formed incidentally around variables such as kinship or proximity, they were often kept together by instrumental goals of profit or survival, which in turn provide an external metric to judge information by. For example, if an agrarian community neglects harvesting the crops in time, the consequences could be devastating for both the group and for the individual.

In order to attain such critical goals, it is in the common interest of the group to find the best common understanding of any situation. Hugo Mercier and Dan Sperber have recently (2017) shown how social interaction promotes truth-seeking beyond what any individual can achieve. They point out that the so-called confirmation bias (Nickerson, 1998), which is counted among the heuristics that leads to choice error, apply selectively to views held by oneself. Therefore it supports correctness in social settings: arguing for one's position is optimized by selecting positive evidence, whereas others are better suited to question one's argument (Mercier & Sperber, 2017). Likewise, when peers fail to find flaws in one's arguments, then the corresponding beliefs are likely to spread in the group. Open discussion is therefore a regulatory system: the group uses positive and negative feedback to support good arguments, and pruning bad arguments, as to (paraphrasing Whitehead and Popper) let mistaken beliefs die instead of their carriers. Importantly, this happens in public discourse, not in individual minds.

When this system works as expected, the consensus of the group has a heuristic value as a guide to truth, or at least, to collective action that is effective for attaining the goal at hand. This explains why humans are more inclined to accept beliefs that are perceived as shared with their social group. As social interaction facilitates truth-seeking on the group level, then the perceived consensus among the group becomes a useful cue to truth for the individual. Negotiating beliefs in social interaction has been an adaptive strategy throughout human history. However, for this to be adaptive two conditions must be in place: First, there needs to be a free exchange of ideas, where negative feedback is allowed. Second, the individual needs to have an accurate perception of who takes part in the discourse, without which they will have a false impression of to what extent beliefs are being shared. Next, we will argue that many social media platforms pose problems for both these conditions.

How Social Media Distort the Context of Belief Evaluation

The platform's selection of information not only affects the kinds of information the user consumes, but also distorts the selection of peers that the user interacts with. Hence, the user does not only get to see more information that they tend to agree with, they will interact more with the people they tend to agree with, and less with those they disagree with. This may lead to the impression that more people share one's views than is actually the case. Not only will users see fewer social media posts contradicting their pre-existing views, they would also get less negative feedback on the views which they advertise through posts: simply because fewer of the peers that would disagree would actually see the message. Facebook's "friends list" makes a case in point here. Even if only a small subset of a particular user's Facebook friends have views similar to their own, on a specific topic, their views on this topic would be a larger part of the user's information flow. While users in principle could assess the number of friends whose views they typically hear (e.g. by comparing the posts visible on individual friends' pages with the posts appearing in the user's feed), it is unlikely to be done regularly by most users. Similarly, my posts as a Facebook user overtly appear to be broadcast to "my (Facebook) friends", while in reality they would reach these friends differentially. If human rationality relies on having our views tested against people we trust, but those whose reactions would be most valuable to this end never sees the content we post, the virtuous social feedback mentioned above is diminished. Confirmation bias is still active, but it has lost its adaptive quality suggested by Mercier and Sperber (2017). Since the actual social network is tampered with by the filters, peer feedback becomes less effective in helping users assess their own convictions. The result is a false sense of consensus, in which many of the user's pre-existing beliefs and convictions will appear as shared with the user's group, when they are actually not.

Contradicting information within our bubbles: how is it perceived by the user?

We will make a final point about the encounters social media users do experience with information that contradicts their pre-existing views. As pointed out by Flaxman and colleagues (2016), social media users are by no means isolated from views they do not agree with. However, this fact is more compatible with the Filter Bubble concept than the original information-centred view suggests. Recall that the content-curating algorithms predict what information we are likely to interact with, rather than simply what we are likely to endorse or like. Therefore it may make sense for the algorithms to select more radical and extreme views regardless of leanings, as these are expected to generate more interest from users[1]. Extreme

[1] We thank an anonymous reviewer for this insight.

views are however less likely to change someone's mind across political boundaries, so it would not help nuancing the discourse. One may also ask how people engage with views they don't agree with. We conjecture that among the arguments and views that one do not endorse, the main mode of engagement beyond consuming these posts, is arguing against them. However, this presupposes that one can formulate the counter-argument, i.e. one finds the opposing view weakly argued in the first place. The risk is thus, that the algorithm over time learns to present us the counter-arguments (against our views) that we already perceive as unconvincing, while becoming less likely to present those counter-arguments that have potential to change our minds. On the larger scale, users may end up with the impression that the opposing side has worse arguments than they actually have. Whether this may be the case has to our knowledge not been investigated, and ought to be targeted by future studies. Because of these considerations, the fact that contradicting views are encountered inside our bubbles, is not by itself evidence against adverse effects of social media algorithms.

Conclusion

In sum, click-maximizing social media platforms such as Facebook curate not only users' access to views and opinions, but also the social context in which views and opinions evolve. The epistemic virtues of social interaction are attenuated, and the user's own beliefs become inflated by how they appear to be shared with the user's group. This warrants further caution regarding filter bubbles and related phenomena, despite the accessible (and accessed) media landscape being more diverse than ever before. More importantly, research about social media would benefit from widening its scope, to take into account the social context of human cultural evolution to a larger extent.

Acknowledgments

We thank the two anonymous reviewers for their valuable comments and insights. An earlier incarnation of this paper was presented as Boyer & Falck (2021) at the 2021 PERITIA conference "Trust in Expertise in a Changing Media Landscape". The authors wish to thank the attendants of said conference, for many valuable comments. A.F. acknowledges support from the Swedish Research Council (grant no. 2016-06783).

References

Baron, R. S.; Vandello, J. A.; Brunsman, B. (1996). "The forgotten variable in conformity research: Impact of task importance on social influence". *Journal of Personality and Social Psychology*. **71** (5): 915–927. doi:10.1037/0022-3514.71.5.915.

Boyd, R. and Richerson, P. J. (2009) 'Culture and the evolution of human cooperation', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533), pp. 3281–3288. doi: 10.1098/rstb.2009.0134.

Bruns, A. (2019). Filter bubble. *Internet Policy Review*, *8*(4). https://doi.org/10.14763/2019.4.1426

Dahlgren, P. M. (2021). A critical review of filter bubbles and a comparison with selective exposure. *Nordicom Review*, *42*(1), 15–33. https://doi.org/10.2478/nor-2021-0002

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly*, *80*(S1), 298–320. https://doi.org/10.1093/poq/nfw006

Fridlund, A. J. (1991) 'Sociality of solitary smiling: Potentiation by an implicit audience', *Journal of Personality and Social Psychology*, 60(2), pp. 229–240. doi: 10.1037/0022-3514.60.2.229.

Mercier H, Sperber D (2017) *The enigma of reason*. Cambridge, MA: Harvard University Press.

Nickerson, R. S. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* 2, 175 (1998).

Pariser, Eli. (2011). *The Filter Bubble: What the Internet is Hiding from You.* London: Penguin UK.

Shteynberg, G. (2010) 'A silent emergence of culture: The social tuning effect.', *Journal of Personality and Social Psychology*, 99(4), pp. 683–689. doi: 10.1037/a0019573.

Shteynberg, G. *et al.* (2016) 'The broadcast of shared attention and its impact on political persuasion.', *Journal of Personality and Social Psychology*, 111(5), pp. 665–673. doi: 10.1037/pspa0000065.

Turner, M. E.; Pratkanis, A. R. (1998). "Twenty-five years of groupthink theory and research: lessons from the evaluation of a theory". *Organizational Behavior and Human Decision Processes*. **73** (2–3): 105–115.

Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, *5*(1). https://doi.org/10.14763/2016.1.401