Echo Chambers and Social Media: On the Possibility of a Tax Incentive Solution

Megan Fritts, University of Arkansas at Little Rock, mcabrera@ualr.edu

––––––––––––––––––––

In "Regulating Social Media as a Public Good: Limiting Epistemic Segregation" (2023), Toby Handfield tackles a well-known problematic aspect of widespread social media use: the formation of ideologically monotone and insulated social networks. Handfield argues that we can take some cues from economics to reduce the extent to which echo chambers grow up around individual users. Specifically, he suggests looking to the use of a "Pigouvian" tax scheme—a tax on market transactions that create negative externalities—to discourage the formation of ideologically "homophilous" social media networks. Likewise, he suggests encouraging "heterophilous" networks via tax breaks. He argues that these taxes may be levied at any of three different groups: individual social media users, social media sites/companies, or advertisers who use social media to promote their products and material.

Regarding Incentives and Penalties

In this response, I examine the plausibility of levying these incentives on each of the three groups Handfield suggests. I argue:

> First, that using tax incentives or disincentives on either (1) social media companies or (2) advertisers would be ineffective as these incentives could not feasibly be made strong enough to override the enormous financial gain of using the standard social media algorithms.

> Next, that levying the incentives/penalties on individual users would be a hazard, due to the risk of what is called the epistemic "backfire effect".

> Finally, that the problem lies in relying on incentives and disincentives— rather than direct regulation—to increase network heterophily.

"Echo chambers", a term initially coined by Cass Sunstein (2001; see also Nguyen 2020), refers to situations in which somebody primarily, or solely, encounters views and beliefs that match their own. Echo chambers may form in the "real world", so to speak, but they form much more easily on social media. This is due to how social media algorithms operate. When a social media user follows, likes, or reposts another account, the algorithms learn data about the sort of content the user prefers and engages with. The algorithms can then promote similar content to the user, predicting increased satisfaction with the site, and/or increased interaction with content. In other words, as I use social media more, I see more and more of the stuff I like—or, at least, stuff I tend to interact with through likes, posts, and shares.

A typical result of this process is that the average social media user's content feed is homogenous; they see content they tend to agree with and see very little content that expresses an opposing argument or viewpoint. The more closed-off my network is from other viewpoints or objections to my views, the fewer corrective influences I am likely to have on my beliefs. And beliefs, and belief systems, without any challengers, tend to become grotesques.

Handfield agrees, writing, "An echo chamber adds to the filter bubble a self-reinforcement mechanism, where a small amount of evidence shared within the network is amplified beyond its proper epistemic weight" (11-12). This does not always result in false beliefs, of course. Handfield writes, "it is not the case that in all instances, a more segregated network performs epistemically worse; and even when homophily is bad for the collective, it may be epistemically beneficial for some individuals" (7). Simply put, sometimes the homophilous network one forms will include people who all share the same true beliefs, and therefore their individual experience with a homophilous network will be beneficial. But such homophily is clearly bad in the collective. It is incontestable that social media has created a perfect environment for the spread of fake news, misinformation, and conspiracy theories.

As my co-author and I argue (Fritts and Cabrera 2022), this is largely due to the social media algorithms that give rise to echo chambers, and particular the way that these echo chambers allow for the function of a misinformation market. In our paper, we further argued that this market fits Satz's (2010) criteria for noxious markets and, given this, we have prima facie reason to close off this market from potential consumers. But Handfield contends that attacking the "goods" of this market itself—that is, the individual pieces of (mis)information—would be inefficient. Rather, we should try to prevent the homophilous networks from forming in the first place.

There are two primary dangers of homophilous social networks that Handfield draws our attention to. The first one is that such social networks may increase partisanship and, thereby, decrease matters on which the general public is able to attain consensus. Handfield argues that, while homophilous networks may not hinder an individual's ability to form true beliefs, truth is not always the only epistemic goal; rather, we have pragmatic reasons for wanting to come to agreement with our fellow citizens regarding matters of policy and governance, among other things. The second danger is that homophilous networks may screen off expert testimony from the very individuals who would most benefit from hearing it. Using tax incentives, or tax penalties, to diversify the social media user experience, Handfield argues, can allay these dangers to some extent, not eliminating echo chambers, but perhaps making their presence less common or less pernicious. He writes:

> [A] social network user may be asked to pay a tax – perhaps in the form of a subscription fee – that increases the more homophilous their immediate network is. So a user might initially subscribe to several highly partisan Republican sites and sources, but then realise they could reduce their subscription fee if they also followed some Democrat users (11).

On Handfield's picture, these incentives (or penalties) may be aimed at any of three targets: social media companies, advertisers who advertise on social media, or individual social media account owners. Aiming at the first target would involve incentivizing sites to make their user experience more heterophilous or using tax disincentives to monetarily punish them for certain degrees of ideological homophily among their users, incentivizing advertisers to advertise to a more diverse audience. Handfield himself quickly dismisses the latter option: even entertaining the idea of incentivizing advertisers to stop advertising in the most high-

tech, targeted, and efficient way possible is a non-starter, because the incentives would simply not be enticing or onerous enough to make up for the massive loss in earnings (12). On the other hand, the former option does not look much more workable. Social media sites typically function in the exact opposite way from what Handfield suggests would be optimal, using machine-learning algorithms to compose a perfectly-curated experience of posts and profiles that the individual users want to see.

These algorithms are typically created to maximize user engagement through likes, comments, and shares. Sometimes this can result in a social media user being shown something they explicitly disagree with—after all, disagreement can yield engagement as well—but this is less likely.[1] Maximized user engagement is how social media sites make money—companies profit both by (1) collecting user data (gathered via post engagement) as well as (2) through advertisers who pay to benefit from the machine-learning algorithms, for maximized targeted advertising. The creation of curated user experiences is almost entirely ubiquitous—all major social media sites function this way. To alter this feature would surely be a death-sentence to any platform, and it is virtually certain that no amount of tax incentives could make up for it. Sufficiently burdensome tax disincentives may have some kind of effect, but one could reasonably predict that this would more likely be the end of particular social sites rather than diversified social media user experiences.

The Value of Data

In "Fake News and Epistemic Vice: Combatting a Uniquely Noxious Market", my co-author and I discuss the intricacies of the relationships among social media users, site owners, advertisers, and political figures. These relationships, perhaps equal parts symbiotic and destructive, give rise to a bevy of opportunities for financial gain and exploitation. Markets arise—in particular, as we discuss, the market for conspiracy theory media and misinformation—from the massive amounts of information that is up for grabs. Much of the information driving these markets use personal user data. This information is highly valuable.

As of 2017, data surpassed oil as the most valuable asset in the world. Social media sites like Facebook and Instagram (now existing under the same company, Meta) accumulate unfathomable amounts of user data, due to the number of active account owners. Facebook's 2016 scandal, involving the sale of personal data to Cambridge Analytica, resulted in the company's pledge to no longer sell user data to outside companies. However, great profit is still made using this data. For example, advertisers can pay Facebook to ensure that the advertisements are targeted to specific audiences in the most efficient ways.[2] The more active the account owners are—the more they engage with posts, advertisements, discussion boards, etc.—the more personal data companies are able to gather.

---

[1] https://www.forbes.com/sites/forbesagencycouncil/2022/10/14/a-guide-to-social-media-algorithms-and-seo/?sh=2accffcb52a0.
[2] https://www.facebook.com/help/152637448140583.

The individual social media users themselves constitute consumers in the information market, as well. Around 62% of Americans get their news either primarily or exclusively from social media.[3] Being able to follow the activity of acquaintances whose judgment you trust—or even journalists, celebrities, and public figures—can give nearly everyone instant access to all kinds of news. The present age of information, made possible largely by social media, is unprecedented; consumers in the market are, or can be, more informed than anyone else at any other time in history. Yet, even though social media sites are free, the consumers still "pay". By willingly forfeiting their personal data, agreeing to terms and conditions that allow this data to be used for purposes of advertising and "enhanced" user experiences, social media users give up a degree of autonomy over how they navigate the glut of media content and information. And this is a cost nearly all of us who voluntarily use social media are willing to pay.

It is clear from the function of the social media-based markets that the existence and functioning of these sites relies on the experience-curating algorithms that show individual social media users the content that they like, agree with, and/or want to see. These algorithms result in optimized user experiences that spur platform growth. These companies are, therefore, incredibly successful, to the tune of (for Meta) $117.346 billion dollars for the 2022-2023 fiscal year. It seems highly unlikely, therefore, that any realistic incentives/disincentives could dissuade social media sites from using these algorithms. To do so would, one can predict, result in their obsolescence.

So, we have good reason to think that Handfield's incentive/disincentive proposal would be ineffective if levied on social media conglomerates and/or advertisers. The remaining option is to target these incentives to individual social media account owners, as a way to nudge them toward intentionally building a heterophilous network.

The "Backfire Effect"

Here, I want to focus on the second danger of homophilous networks that Handfield discusses: the danger that they screen off expert testimony from being seen by those who most need to see it. *Prima facie,* it seems as though exposure to expert testimony, especially on knowledge-intensive and complicated topics in fields such as medicine, biology, physics, economics, and so on, would be epistemically beneficial to anyone. "Given scientific knowledge in many policy-relevant areas is highly advanced and inaccessible to non-experts, many policy issues are likely affected by this doubly problematic variety of homophily" (14). Increasing general exposure to expert testimony, especially increasing exposure to those whose "echo chamber" may have excluded those particular experts, would naturally be assumed to benefit the collective epistemic good. However, much empirical data undercuts the seemingly-obvious idea that more exposure to expert testimony on difficult/controversial subjects leads to more veridical beliefs. In fact, in some cases, such exposure may have exactly the opposite effect.

---

[3] Gottfried and Shearer (2016).

There is phenomenon known in psychology as "belief perseverance", where subjects maintain their belief in a certain proposition, despite being presented with a case of disconfirming evidence. A sub-set of cases of belief perseverance go even further. In these cases, subjects display something known as the "backfire effect". The backfire effect "turns humans reasoning ironic; people in its grip believe *more firmly* their original opinions in the face of strong countervailing evidence" (Aikin 2018, emphasis added). This is especially true when the topic in question is one at the center of people's self-conception, such as politics or religion (Mandelbaum 2018). For example, a new study has shown that "fact-checks were more likely to backfire when they came from a political outgroup member", and that "corrections from political outgroup members were 52% more likely to backfire — leaving people with more entrenched beliefs in misinformation".[4]

One of the most famous original examples of the backfire effect comes from Batson (1975), which details a study in which participants were provided with an article that claimed that evidence had been found that showed the religious movement of Jesus Christ and early Christianity to be fraudulent. The evidence in question was a set of scrolls found in Jordan, describing the efforts of early Christians to cover up Jesus' death with a made-up resurrection story. The article claimed that this information had been suppressed from the world due to the catastrophic effect it would have on the religious people around the world.

After reading the article, participants were asked two questions: 1) did they believe in the divinity of Jesus prior to reading the article, and 2) did they accept the veracity of the article and the weight of the evidence against Jesus' divinity. Unsurprisingly, those who did not previously accept the divinity of Jesus took the article at face value, and their credence in the divinity of Jesus was lowered by the disconfirming evidence. However, those who did accept the divinity of Jesus prior to reading the article did not see their credence in this proposition decreased by the evidence; to the contrary, their credence in his divinity actually increased. This remained true between Christians who accepted the evidence as factual and those who rejected the evidence.

Instances of belief perseverance are not restricted to the religious domain. Consider the following recent occurrence. In 2017, Facebook attempted to combat the rampant misinformation making its way across users' timelines by "flagging" misinformation with a warning, alongside expert explanation about the topic in question. If a piece of misinformation of fake news made it onto your Facebook newsfeed, you would be able to see the response of experts in the field, letting you know why, in their educated opinion, this was bad or misleading information. Unfortunately, this plan backfired, and "flagging" misinformation was stopped later that year. Facebook product manager Tessa Lyons wrote: 'Academic research on correcting misinformation has shown that putting a strong image, like a red flag, next to an article may actually entrench deeply held beliefs—the opposite effect to what we intended'.[5] And, in fact, this is precisely what Facebook saw happening.

---

[4] Reinero et al (2023), preprint.
[5] https://www.bbc.com/news/technology-42438750.

While this backfire effect seems manifestly irrational, this is not necessarily the case. Imagine the following scenario: Tim is a high school band teacher, and John is the parent of one of his best students, Myra. Based on previous experiences with John, Tim has come to believe that John is attempting to sabotage his work so he can take Tim's job. At the beginning of the school year, John complains to the principal that Tim's carefully-crafted summer practice regimen is not working—John claims that Myra is not any better at the cello than she was at the start of summer break, and that Tim should consider a different regimen altogether. Tim's confidence that his practice regimen is good and effective is not reduced due to the evidence of testimony supplied by John; indeed, Tim is now *more* confident than he was before that the summer practice routine is sharpening Myra's skills.

If Tim's credence in John's saboteurial schemes is sufficiently high, then it may be rational for John's testimony *against* Tim's practice regimen to actually constitute evidence in *favor* of it. Of course, such conditions of rationality may be rare—it is unlikely that most, or even many, of those experiencing belief perseverance meet such conditions.

How is this related to Handfield's individual incentive proposal? In cases like Batson's (1975) experiment, it is not clear that something like the Tim and John case is occurring. But in other cases, especially cases regarding current political events, it is not hard to imagine this type of reasoning going on. My concern is the following: if people follow the incentives and diversify their social media experience, and if they begin this process with high enough credences in their core political, moral, or religious beliefs, rather than seeing the diminishing of echo chambers and epistemic bubbles, we may instead see a backfire effect. This may be especially true of the specific kinds of epistemic bubbles that Handfield is interested in eliminating. Those in politically homogenous environments may be especially inclined to have very high credences in their political positions, and may also be deeply convinced in conspiratorial or nefarious activity from the other side of the political aisle. In this situation, while Handfield's proposal may indeed diversify the kinds of perspectives that these groups are exposed to, the diversity may fail to yield the benefits of heterophilous social networks that Handfield describes.

Advantages of Direct Regulation

I have argued that there is good reason to doubt that tax incentives or disincentives can be effectively used—either with social media companies, advertisers, or individual account owners—to diversify social media networks in ways that will yield epistemic benefits. The crux of the problem lies, I believe, in using incentives and/or penalties, rather than directly regulating minimum levels of social network heterophily. I have argued that the extreme wealth accumulated by both social media companies and those who advertise on social media would render any realistic amount of tax incentives or penalties causally inert. This is because the social networks, and the advertisements that exist therein, are successful and profitable *in virtue of* the very user-experience-curating algorithms that give rise to echo chambers.

Targeting individual social media users with tax incentives or penalties could avoid this problem; but, I have argued, it may give rise to something worse. The epistemic "backfire" effect" is a phenomenon in which people maintain and strengthen beliefs in particular propositions (especially political or religious propositions) after encountering some disconfirming evidence for those propositions. Incentivizing users to diversify their social media experience—especially along the ideological lines that Handfield suggests—would risk "backfiring", more deeply engraining the beliefs that people already have. Rather than providing an opportunity to gain new perspectives on various issues, social media sites could become places that people use merely to seek out ideological enemies for tax advantages.

By relying on tax incentives and disincentives to drive network individual users to choose network heterophily, these users encounter diverse viewpoints *explicitly as* content from political outgroup members, priming them for the backfire effect. But if user networks were made heterophilous automatically—due to legislation that social media sites must comply with—users would not encounter diverse viewpoints as, first and foremost, enemy material. Rather, different perspectives would be a normalized aspect of all social media use, plausibly tempering the potential for epistemic backfire.

References

Aikin, Scott. 2018 "Empirical Assumptions and Philosophical Ethics: On Mark Alfano's Moral Psychology." *Syndicate* https://syndicate.network/symposia/philosophy/moral-psychology/.

Batson, C. Daniel. 1975. "Rational Processing or Rationalization? The Effect of Disconfirming Information on a Stated Religious Belief." *Journal of Personality and Social Psychology* 32 (1): 176–184. https://doi.org/10.1037/h0076771.

Gottfried, Jeffrey and Elisa Shearer. 2016. "News Use across Social Media Platforms 2016." *Pew Research Center* https://www.pewresearch.org/journalism/2016/05/26/news-use-across-social-media-platforms-2016/.

Fritts, Megan and Frank Cabrera. 2022. "Fake News and Epistemic Vice: Combating a Uniquely Noxious Market." *Journal of the American Philosophical Association* (3): 1-22.

Handfield, Toby. 2023. "Regulating Social Media as a Public Good: Limiting Epistemic Segregation." *Social Epistemology* doi: 10.1080/02691728.2022.2156825.

Mandelbaum, Eric. 2018. "Troubles with Bayesianism: An Introduction to the Psychological Immune System." *Mind and Language* 34 (2): 141-157.

Nguyen, C. Thi. 2020. "Echo Chambers and Epistemic Bubbles." *Episteme* 17 (2): 141-161.

Reinero, Diego A., Elizabeth A. Harris, Steve Rathje, Annie Duke, Jay J. Van Bavel. 2023. "Partisans are More Likely to Entrench Their Beliefs in Misinformation When Political Outgroup Members Fact-Check Claims." *PsyArXiv Preprints* doi: 10.31234/osf.io/z4df3.

Sunstein, Cass. 2001. *Echo Chambers*. Princeton University Press.