PERSPECTIVE
https://doi.org/10.1038/s42256-021-00358-3

nature machine intelligence

Epistemic fragmentation poses a threat to the governance of online targeting

Silvia Milano[1]✉, Brent Mittelstadt[2], Sandra Wachter[2] and Christopher Russell[3,4]

Online targeting isolates individual consumers, causing what we call epistemic fragmentation. This phenomenon amplifies the harms of advertising and inflicts structural damage to the public forum. The two natural strategies to tackle the problem of regulating online targeted advertising, increasing consumer awareness and extending proactive monitoring, fail because even sophisticated individual consumers are vulnerable in isolation, and the contextual knowledge needed for effective proactive monitoring remains largely inaccessible to platforms and external regulators. The limitations of both consumer awareness and of proactive monitoring strategies can be attributed to their failure to address epistemic fragmentation. We call attention to a third possibility that we call a civic model of governance for online targeted advertising, which overcomes this problem, and describe four possible pathways to implement this model.

Online targeted advertising (OTA) is the engine of the digital economy. Many of the services that we have come to rely on, from e-mail to search engines, entertainment and social media, are financed through advertising. Platforms collect data from consumers as a condition of access to services and personalized content, typically through processes marked by stark information and power differentials that call into question the possibility of meaningful consent[1–7]. In turn, advertisers pay these companies to target their campaigns to specific audiences built on inferences drawn from user data[8,9]. The technologies that draw these inferences, build audiences and personalize content create the fuel needed to drive the digital economy[10].

OTA is not a benign application of machine learning. It can lead to individual consumers losing contact with their peers, a problem we call epistemic fragmentation. Epistemic fragmentation increases consumers' vulnerability to the harms of advertising, while simultaneously damaging information consumption in the public forum and hindering efforts to institute effective regulation. Despite the rapid evolution of OTA, regulatory frameworks internationally remain anchored to traditional methods of advertising[11]. This is changing, as governments are moving to draw new regulations[12], mobilizing huge economic and political interests[13,14]. But effective regulation cannot be guaranteed if the underlying challenge is not fully conceptualized.

This Perspective makes several contributions to conceptualize and combat epistemic fragmentation. First, we introduce a taxonomy of the harms of advertising. We argue that targeting tends to normalize harmful content and makes it increasingly difficult to monitor compliance with existing codes of conduct in the advertising industry. The strategy pursued by regulators to address these issues at present is inadequate because it does not address their root cause, which we trace to epistemic fragmentation. By hiding each individual's personal context, targeting makes consumers more vulnerable and increases the overall costs of instituting protections via proactive monitoring. Instead, regulators should promote an active role for consumers in fighting epistemic fragmentation by adopting a civic model of governance for advertising. We describe four potential pathways to implement our proposal, and conclude by summarizing our argument and pointing to future work.

Harms of advertising

From the point of view of consumers, the increasing ubiquity of OTA may increase access to relevant information and products[15,16]. But personalization also carries notable risks of harm.

We can distinguish four categories of harms that advertising can cause to consumers, which are organized along two dimensions (Table 1). One dimension (the absolute row in Table 1) groups harms that originate in the nature of the content that is either included or excluded from what is shown to an individual consumer. Content that is genuinely bad, for example ads making false claims about a product or using racist stereotypes to promote a product, is harmful when it is included among what a consumer sees (absolute harms of inclusion). On the other hand, if a piece of content is genuinely good or even vital, for instance an important public health announcement, it may be harmful for an individual not to see it (absolute harms of exclusion).

Along the second dimension (the contextual row in Table 1), we find harms that do not stem from the nature of the content per se, but depend on the context in which the content is delivered. An exploitative context is one where, independently of whether the content is intrinsically harmful, the way in which it is delivered exploits a consumer's vulnerability. For example, ads for high-fat-content foods may not be generally bad, but they should not be targeted to children. Similarly, gambling ads can be exploitative. Finally, a deprived context is one where a consumer is not shown ads that would be relevant and beneficial[17]. For example, a deprived context might be one where a consumer who would be interested in finding a job is not shown ads for jobs in their area.

Although these types of harms are conceptually distinct, they are not mutually exclusive and can co-occur. For example, a racist message or a false claim (bad content) may be presented next to reputable brands or news items, where the context helps to normalize the harmful message (exploitative context). Together, absolute and contextual harms can contribute to the degradation of public

[1]Future of Humanity Institute, University of Oxford, Oxford, UK. [2]Oxford Internet Institute, University of Oxford, Oxford, UK. [3]The Alan Turing Institute, London, UK. [4]Present address: Amazon, Tübingen, Germany. ✉e-mail: silvia.milano@philosophy.ox.ac.uk

Table 1 | Taxonomy of the possible harms of advertising

	Inclusion	Exclusion
Absolute	Bad content	Omission of essential content
Contextual	Exploitative content	Deprived of content

discourse. Table 2 exemplifies each type of harm, how it is covered by current regulation and how it is monitored in practice.

Effective monitoring and enforcement mechanisms are essential components of regulation. It is here that OTA poses the greatest challenge. Current industry codes of conduct address most types of harms that arise from OTA, with the partial exception of contextual and omission harms that could nonetheless easily be incorporated going forwards. However, updated codes will remain inadequate without effective methods to monitor and enforce advertiser and platform compliance. Traditionally, regulatory agencies have relied heavily on reactive enforcement mechanisms including consumer complaints and post-publishing reporting[18,19]. The scale and method of distribution of OTA fundamentally challenge reactive monitoring regimes.

Human reviewers cannot manually inspect the huge volume of ads exchanged online on a daily basis[19]. Screening ad content and monitoring the contextual effects of ad delivery are increasingly automated[20,21], despite the machine learning capabilities needed to replace human judgement in content review not yet being available[22]. Even if platforms, consumers or regulators could reliably check every advertisement for harmful content, contextual harms and harms of exclusion are much more difficult to monitor because they can present in myriad combinations[23–27]. Regulatory bodies are nonetheless moving to fill this gap[14].

Two natural but insufficient responses

Regulators already recognize weaknesses in current monitoring and enforcement methods. A recent report by the UK Centre for Data Ethics and Innovation[18], for example, details regulatory proposals to enhance consumer protection through collaboration with social media platforms. In the United Kingdom, the ASA's More Online Presence strategy, launched in 2019, sees it—in the words of the ASA's Chairman David Currie—as "rebalancing away from reactive complaints casework and towards […] proactive tech-assisted intelligence gathering, complaint handling, monitoring and enforcement"[18].

The limits of consumer awareness. Part of the challenge created by OTA is that it hides from consumers when, how and why they are being targeted. In contrast, more traditional forms of advertising use more transparent delivery mechanisms. To give an example, consider newspaper advertising. In the printed version, advertisers buy spaces and their ads are seen by all readers of the same newspaper. Readers can be certain they are seeing the same ads as everyone else in the same context; the advertising environment is the same for all readers. In contrast, readers of online versions of the same newspaper will not be able to make the same inference, as digital ads are typically customized to the reader on the basis of targeting data (for example tracking cookies, inferred interests), meaning different readers see different ads when visiting the same web pages.

One common approach to create equivalent protections for consumers against OTA is to give consumers more control over how they are targeted, for example control over the categories of data they share with advertisers or platforms. This 'awareness strategy', as we will call it, is a constant among the recommendations made across recent reports on regulating online advertising[12–14,18,28]. Initiatives such as YourAdChoices aim to inform consumers about how they are targeted, giving them access to an explanation for why they see a certain ad, and options to control the types of ads they see[29,30]. Initiatives of this kind may be important to build consumer trust[31].

However, shifting the responsibility back to consumers to protect themselves from unwanted targeting is an ineffective strategy. Current methods adopted by the industry fall short of providing adequate explanations of actual targeting mechanisms[32]. They do not offer meaningful protection in cases where consumers are unfairly targeted or manipulated without their knowledge. Consumers cannot be expected to identify when they are affected by differential pricing, or manipulated through repeated exposure and remarketing practices, because they lack contextual information about how and when these processes occur at a platform level. The awareness strategy therefore risks victimizing vulnerable consumers while doing little to monitor advertiser compliance with relevant regulatory frameworks and codes.

Moreover, identifying harmful content and contexts requires an especially critical and empathetic eye. Take, for example, a case of an ad that uses sexist stereotypes to promote a car, and that is somehow perfectly targeted to consumers who agree with the stereotype. Is this an objectively harmful practice? From the view of the targeted consumers, possibly not. At a minimum, they would seemingly be less likely than an average consumer to raise a complaint or experience harm, either because they do not realize that the message uses a harmful stereotype, or because they agree with the stereotype. An explanation for why such targeted content is nonetheless harmful must consider the broader social effects of the stereotype.

Transparency can also backfire if it reveals that advertisers used information that is unacceptable to the user, and do so via a platform that the user does not fully trust[31]. Moreover, the suggestion by platforms that consumers benefit from OTA because it helps them find more relevant content[15,16] is at odds with the strategy of improving awareness. If consumers must be more vigilant about how they are targeted, then this takes away some of the supposed benefit of targeting, which should reduce information overload faced by consumers. Shifting responsibility onto consumers undermines the supposed efficacy of targeting; seeing more relevant ads is not worth being harmed in the process.

Contextual harms can also go undetected. For example, targeting a consumer based on their inferred interest in gambling[33] may be considered unethical, regardless of whether the recipients appreciate the ads. But of course, appreciative recipients are less likely to raise a complaint. Similarly, exclusion from opportunity ads, such as ads for high paying jobs, may go unnoticed without further investigation[27]. Giving consumers explanations for why they see specific ads may seem to be a natural solution to mitigate these concerns. However, transparency in practice fails to deliver in this regard.

Finally, raising consumer awareness is ineffective for discovering systemic issues, such as statistically differential exposure to opportunity ads or discrimination by association[27]. In these cases, where the omission of relevant information harms a consumer, informing individuals about why they are seeing certain ads (even if platforms could provide adequate explanations, something that is not yet being achieved[32]) would not be enough to raise awareness of being harmed, and would not provide a sufficient basis for recourse.

Failures of proactive monitoring. Given that promoting consumer awareness is an insufficient monitoring strategy, regulators need additional checks to ensure that advertisers comply with codes of conduct. In this context, a natural strategy is to augment the regulators' online presence through a proactive monitoring strategy. Regulatory bodies in the United Kingdom, Europe and other Western countries are moving in this direction, working in partnership with online platforms[13,14,18,28,34,35]. For example, in the United Kingdom, the ASA is implementing this strategy in various ways,

Table 2 | Examples of harms of advertising, current regulation and monitoring

Type of harm	Example	Why it is harmful	Current regulation	Current monitoring
Bad content	Ad using sexist stereotypes to promote a car.	Reinforces prejudices, exposes consumers to harmful messages, may be detrimental to consumers' well-being.	Covered by existing codes, which set out categories of harmful content (for example sexist, racist, violent).	Consumer reports[18,78]; proactive monitoring by self-regulatory bodies[18,34]; manual and automated screening conducted by online platforms prior to approval[79].
Omission of essential content	Vital information not received (for example health emergency communications).	Missed opportunities; exposure to risk due to the lack of essential information.	Covered by health and safety standards and public service authorities.	Health and safety standards authorities; public communication services[80,81].
Exploitative context	High-fat-content food ads shown to children; ads for gambling shown to consumers from deprived demographics.	Exploits consumers' vulnerabilities; is detrimental to well-being; impairs consumers' ability to make autonomous and well-informed choices.	Partly covered by codes, which prohibit exploitative advertising practices to vulnerable groups[11,28]. Grey areas remain as behavioural targeting exposes consumers to more subtle forms of exploitation.	Consumer reports; platform monitoring[18,28,34].
Deprived context	No or few opportunity ads; higher differential pricing.	Discrimination; discrimination by association; broader systemic effects such as harming competition.	Partially covered in the codes, which ban discrimination on the basis of protected categories[11].	Platform monitoring, investigative journalism[82].

including the use of avatars posing as children to monitor the ads that a child would see online[34].

Switching to a proactive oversight model implies that the criteria for what counts as harmful content, and which categories of consumers are protected, become controlled by either the platform, the regulator or some form of cooperation between the two. However, 'harmful content' is a malleable category that evolves together with public awareness of social issues[36–38]. It is likely that some of the messages that are acceptable today will be seen as problematic in the future. While benevolent paternalistic monitoring may appear effective in the short term, it could preclude social progress and freedom of speech. The ongoing shift to automated monitoring using machine learning tools, which intensified during the coronavirus pandemic[21], adds to this worry given well-established cases of applications of machine learning to other domains perpetuating discrimination and injustice[27,39–41].

Contextual harms of exclusion, which by definition are produced by the absence of something, are particularly difficult to identify. Individual consumers struggle to know what they are not being shown without outside help. Agencies filling this role would need to know which vulnerable groups or audiences to monitor. Consistent identification would require a dataset containing all (or a representative sample) of the advertisements served on a given platform categorized by audience, raising privacy concerns[42]. Regulators may also lack access to sensitive audience demographics data.

Active oversight programmes are also expensive. To effectively monitor ad quality and the fairness of their distribution, platforms need to create monitoring systems, often using machine learning[21,43,44]. However, such safeguards carry substantial costs, both in computation[45] and human labour (for example, to label harmful content to train and maintain automated monitoring systems[46]). These costs of monitoring, as well as the human and economic cost of maintaining current labelling processes, should be considered in future regulatory strategy.

A challenge from epistemic fragmentation

OTA thus poses new challenges to regulators, who are responding by introducing more proactive monitoring and consumer awareness campaigns, but both of these strategies face serious limitations. We trace the source of the issue to a systemic feature of OTA that we call epistemic fragmentation, and argue in favour of a civic form of monitoring to address this.

Epistemic fragmentation. The predominant self-regulatory framework used by the advertising industry has been reactive and complaints-based. Complaints have historically not been made by the most vulnerable groups or harmed consumers in isolation, but rather by concerned third parties or organizations acting from civic concern. Complaints about potentially exploitative ads are predominantly raised by consumers with relevant expertise or interests, who are likewise unlikely to be misled. For example, a financial advisor could report an ad for fraudulent tax reduction services[18], or, a teacher could report an ad that exploits body image issues to promote a fashion product out of concern that it could harm a teenage student who may not realize how it exploits common insecurities[35].

In these examples, the consumers most likely to have the relevant information and motivation to raise a complaint are not themselves vulnerable, but are aware of those more vulnerable to harm. Parties raising a complaint have sufficient contextual information to recognize how the ad may harm others. Instances of contextual harms can also often only be flagged by third parties who have access to contextual information about targeting. For example, identifying that certain opportunities are only advertised in newspapers read by specific demographics, which may be discriminatory, requires that someone with access to both contexts raises the complaint[17,47]. Access to contextual information, as well as motivation to speak out in favour of an affected community, are thus necessary to identify and flag certain harms.

With OTA, this shared context is effectively destroyed. The teacher and financial advisor may never see ads directed at teenagers or people facing financial difficulties, and therefore never raise a complaint. Without shared advertising space, the extent to which opportunity ads are shown differentially to members of different demographics cannot be reliably measured by the public[27,40,48]. Each consumer's 'personal context' is hidden from others, meaning

nobody knows exactly what others see and cannot raise a complaint on their behalf.

For our present purposes, we define a personal context as the sum of two components: the personal information about an individual consumer (that is, who they are, what they are interested in and value, where they live and so on) and the content that they see (that is, the specific ads that are served to them by the platform). In offline advertisement, elements of consumers' personal contexts are available to others. Full details of personal contexts, for example extensive personal (profile) data and ads encountered, are of course not routinely available, but sufficient contextual information exists to form an initial judgement and raise a complaint for further investigation. Contrast this to what happens online: personal contexts are entirely opaque. Major platforms maintain ad repositories[49,50], but do not grant any specific tools to external organizations who wish to report on how ad campaigns are targeted at different demographics. As each individual consumer is served different (personalized) ad content, and as consumers do not know what ads others are seeing when visiting the same websites, each consumer's personal context is hidden from others. We refer to this lack of shared context in relation to a given practice of content personalization as epistemic fragmentation.

Epistemic fragmentation shares some similarities with other phenomena connected to content personalization systems, such as so-called filter bubbles, which can arise when a content personalization system limits individual users' exposure to information and viewpoints that differ from their own[51,52]. While the extent to which filter bubbles polarize may be less than previously thought[53,54], the reason they seem to matter is that they could reduce individuals' capacity to access relevant information and form opinions. This suggests that the way to address filter bubbles would be to grant individuals access to more varied information sources, limiting the extent to which discordant opinions can be filtered out by the system.

Epistemic fragmentation could be a factor determining the formation of filter bubbles, but the latter are not a necessary component of the former. For example, a consumer could be exposed to a diversified range of ads online, and so not live in a filter bubble, but at the same time have no information about what ads other consumers are seeing. The reason why epistemic fragmentation matters, moreover, is not just because it limits individuals' ability to access information that is relevant to them, but also because it limits their ability to assess the quality of content that is accessed by others. In doing so, it limits their ability to care for those others, and to take a role in the governance of the system.

A different model of governance. Regulating OTA is a pressing issue. We call the current approach, stressing consumer awareness together with increased proactive monitoring[13,55], the monitoring model of governance (Fig. 1a). In this model, individual consumers interact exclusively with the central monitoring agent, without access to others' personal contexts. But isolated consumers, as we have argued, remain vulnerable.

In what we call the civic model, by contrast (Fig. 1b), consumers (and civil society more generally) are given the power, epistemic resources and procedural means to contribute to the monitoring of the system and affect change. In this model, epistemic fragmentation is reduced by design, ensuring that individual consumers have access to aspects of others' personal contexts and to the operations of the monitoring agent.

The civic model of governance is consistent with recent proposals put forwards by regulators in the EU and elsewhere. The European Commission's Digital Services Act proposal[56], for instance, contains recommendations to increase transparency and institute ad repositories that are publicly available and searchable, including information about the groups that receive an ad (see article 30 of ref. [56]).

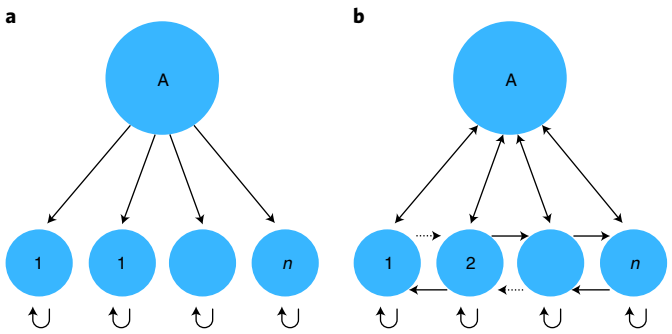

Fig. 1 | Models of governance for online targeted advertising. a, The monitoridng model. Individual consumers (1, 2…*n*) and the monitoring agent (A) have access to information about consumers (arrows represent access to personal context). **b**, The civic model. Consumers have access to elements of others' personal contexts and of A. (Levels of access can vary; dashed arrows represent relatively limited access.)

OTA has long been identified as a threat to privacy[57], but epistemic fragmentation is a systemic feature distinct from privacy. Improving the privacy of ad delivery systems, such as Google's recent changes concerning third-party cookies, will not address the issues we have identified[58]. Two things are worth noting in this context. First, based on our discussion, epistemic fragmentation emerges as a characteristic feature of online targeted advertising, but generalizes to other domains where information is shared online via systems that filter and personalize what each user sees. Insofar as epistemic fragmentation hinders effective monitoring, it gives rise to systemic harms: by fragmenting individuals' access to each other's personal contexts, we lose vital access to the necessary information to identify and address harmful content, practices and omissions. Second, the harm caused by epistemic fragmentation is not captured by legislation such as the EU General Data Protection Regulation, which focuses on the personal data and rights of individuals, but largely ignores collective harms and protections[5,8,27,59–62].

Finally, we see three sets of reasons why civic governance is worth pursuing: epistemic, ethical and political. First, from an epistemic standpoint, as we have argued, distributed monitoring is more effective at surfacing cases of harmful content and/or context, which would be difficult or impossible to notice without the contextual information that is only available to uniquely situated consumers. Addressing epistemic fragmentation removes the barrier to accessing this contextual information. Second, the current system of targeted advertising limits people's ability to care for their communities. This gives ethical reasons in favour of civic governance. Third, from a political standpoint, epistemic fragmentation hinders informal 'deliberation in the wild'[63–66] that is necessary to provide the proper input to the more formal aspects of democratic governance.

Towards civic governance

We propose four possible pathways to implement civic governance. With these proposals, our aim is to spark debate to identify new strategies for regulating OTA. None of these proposals would be sufficient in isolation. An independent regulatory agency responsible to set the codes, respond to complaints and enforce decisions is still essential. Our discussion highlights that the debate on regulating OTA should move beyond privacy concerns, engaging instead with solutions to limit epistemic fragmentation, promote civic governance and ultimately improve the quality of public discourse.

Noisy targeting. Regulators could allow OTA, but only on the basis of coarse-grained categories that can be easily communicated to consumers, or introduce some noise in targeting so that out-group

members are included. Regulations should define the level of granularity that is allowed for different categories of targeting, which groups of consumers can be targeted and how. Google and Facebook already implement this strategy to some extent, as they do not allow advertisers to target their campaigns at a level of specificity that would be technically possible given their algorithms[67]. However, this is not yet done under any kind of robust oversight. A clear limitation of this approach is that it primarily addresses privacy (and filter bubbles) more than epistemic fragmentation. To be feasible, there must be transparency about the context in which an ad is seen and access to other perspectives.

Targeting quotas. As a second option, a quota system could be introduced by which regulators limit the proportion of targeted ads per customer per platform, while the rest of the advertisements on display must be non-targeted. Alternatively, regulators could allow an ad to be targeted only a fraction of the times it is displayed, with the rest of the impressions via contextual placements. This would limit the possibility to exploit consumers' individual vulnerabilities, reducing both contextual harms of inclusion and exclusion.

Ban targeting. Targeting hinders the ability of consumers to learn and care for each other by obscuring what ads others are seeing. Accessing this distributed knowledge is especially valuable. A radical but obvious response to these observations would be to ban targeted advertising entirely. This suggestion has recently been advocated, for different reasons than those we present here, by others who point to the limited economic effects of targeting on consumer markets[68,69]. A problem and possible limitation for this approach is that it does not address epistemic fragmentation if the content that provides the contextual anchoring for the ads is itself targeted.

Reconstruct the public forum. While the previous approaches are low-tech, another possible solution is to use technology to reconstruct a digital public forum. Instead of trying to reproduce the conditions for the traditional, offline monitoring regime, it may be possible to create new forms of digital public spaces that sustain civic governance.

As an example, the Citizen Browser Project, recently launched by the investigative journalism organization The Markup[47,70–72], proposed a tool to improve the accountability of OTA. As part of the project, a representative sample of internet users were to be paid to voluntarily install and use a custom browser, allowing researchers to gather data on what ads online platforms serve to individuals in different demographics. Generalizing the idea of this project, Wachter suggested that a way to address epistemic fragmentation would be to support independent research, 'white hat hacking'[27] and collective and group rights[27,59,61,62,73–75] to keep platforms and advertisers accountable. While this approach would not be sufficient to solve epistemic fragmentation, it is nonetheless a step in the right direction in recreating the public forum. However, Facebook's move to shut down this project, on privacy grounds[76,77], illustrates the difficulty of aligning the interests of regulators, tech platforms and citizens with respect to monitoring online harms.

Other technological means could help users contextualize the content they encounter and understand what other users are seeing. For example, Wachter et al. recently proposed that AI system controllers should be required to routinely produce summary statistics based on conditional demographic disparity to help fight algorithmic discrimination[40]. This type of disclosure can help reconstruct the public forum by showing how outcomes are distributed across groups affected by a given AI system, which helps identify disadvantaged groups that might otherwise go unnoticed. This approach could be adapted to facilitate conversations about the distribution of advertisements and outcomes across relevant groups to identify contextual harms of exclusion.

The four proposals above are merely starting points for further discussion to create a workable regulatory strategy to combat epistemic fragmentation. One may worry that, to operationalize any of these proposals, we must define a demarcation between problematic targeting and socially acceptable audience segmentation. A wide-ranging and inclusive political debate over appropriate thresholds and levels of granularity in targeting is precisely what our proposals are meant to foster. This type of debate is an essential complement to ensure technological advances in flagging and reporting problematic personalized content serve the needs of the public. At the moment, these decisions are predominantly taken by advertisers, platforms and industrial bodies with little external oversight. Civic monitoring and technological innovation must be complementary to uphold the democratic governance of online spaces.

Conclusion

OTA creates a disconnection between individual consumer's experiences and the experiences of their social circles. We refer to this as epistemic fragmentation. This phenomenon amplifies the harms of advertising and deteriorates the forum of public discourse around what, as a society, we consider harmful. Absolute harms, especially harms of inclusion, can be normalized by presenting them to consumers in personalized settings where they are placed next to trusted sources, or showing them only to consumers who already agree with the harmful messages, and who thus fail to question them. Contextual harms, both of inclusion and exclusion, are especially likely to arise in the presence of targeting, and their co-occurrence can give rise to summative effects, amplifying patterns of discrimination and disadvantage.

But epistemic fragmentation causes serious damage also in a second, more indirect way, by limiting our ability to act as empowered citizens, which is the foundation of civic governance. The current way in which discussions around regulating OTA are framed portrays consumers in two ways: either as individual agents who must be educated about privacy and given more individual control over the types of content that they want to see (what we called the strategy of improving consumer awareness); or as passive subjects in need of increased protections, which should be provided by monitoring agencies in partnership with online platforms, who are the only entities that have access to enough data and resources (the proactive monitoring strategy). We have argued that this binary framing is too restrictive. Under epistemic fragmentation, even educated individual consumers remain vulnerable to exploitation, and the epistemic resources necessary to implement effective proactive monitoring remain largely inaccessible, in addition to generating social and economic costs.

If epistemic fragmentation is a root source of these problems, then any successful approach at regulating OTA should address it. Restoring a shared public forum should be the priority if we want OTA, and personalized content more broadly, to be safer, more accountable and ultimately better for all of us. We put forward four possible ways to do so: (1) blunting the precision with which ads can be targeted to individual consumers to increase the likelihood that harmful content or contexts are identified; (2) instituting targeting quotas to limit the ability of advertisers and platforms to decide who should be included or excluded from certain messages; (3) applying a blanket ban on targeted advertising; or finally (4) reconstructing a digital public forum via targeted technological interventions. All of these suggestions raise technical and political questions that should urgently be the object of debate.

Received: 27 November 2020; Accepted: 12 May 2021;
Published online: 17 June 2021

References

1. Solove, D. J. Introduction: privacy self-management and the consent dilemma. *Harv. Law Rev.* **126**, 1880–1903 (2013).
2. Cohen, J. E. What privacy is for. *Harv. Law Rev.* **126**, 1904–1933 (2013).
3. Waldman, A. E. Privacy law's false promise. *Wash. Univ. Law Rev.* **97**, 0773 (2020).
4. Cate, F. H. & Mayer-Schönberger, V. Notice and consent in a world of Big Data. *Int. Data Priv. Law* **3**, 67–73 (2013).
5. Mantelero, A. Personal data for decisional purposes in the age of analytics: from an individual to a collective dimension of data protection. *Comput. Law Secur. Rev.* **32**, 238–255 (2016).
6. Zuboff, S. Big other: surveillance capitalism and the prospects of an information civilization. *J. Inf. Technol.* **30**, 75–89 (2015).
7. Véliz, C. Privacy and digital ethics after the pandemic. *Nat. Electron.* **4**, 10–11 (2021).
8. Wachter, S. & Mittelstadt, B. D. A right to reasonable inferences: re-thinking data protection law in the age of Big Data and AI. *Columbia Bus. Law Rev.* **2019**, 494–620 (2019).
9. Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (PublicAffairs, 2019).
10. Choi, H., Mela, C. F., Balseiro, S. R. & Leary, A. Online display advertising markets: a literature review and future directions. *Inf. Syst. Res.* **31**, 556–575 (2020).
11. *Advertising and Marketing Communications Code* (ICC, 2018); https://cms.iccwbo.org/content/uploads/sites/3/2018/09/icc-advertising-and-marketing-communications-code-int.pdf
12. *Update Report into Adtech and Real Time Didding* (ICO, 2019); https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf
13. *AI Barometer Report* (CDEI, 2020); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/894170/CDEI_AI_Barometer.pdf
14. *Online Harms White Paper* (UK Department for Culture, Media and Sport & Home Office, 2019).
15. Sahni, N. S. & Nair, H. *Does Advertising Serve as a Signal? Evidence from Field Experiments in Mobile Search* (SSRN, 2018); https://doi.org/10.2139/ssrn.2721468
16. Sahni, N. S. & Zhang, C. *Search Advertising and Information Discovery: Are Consumers Averse to Sponsored Messages?* (SSRN, 2020) https://doi.org/10.2139/ssrn.3441786
17. Kingsley, S., Wang, C., Mikhalenko, A., Sinha, P. & Kulkarni, C. Auditing digital platforms for discrimination in economic opportunity advertising. Preprint at https://arxiv.org/abs/2008.09656 (2020).
18. *Using Technology for Good: Annual Report 2019* (ASA/CAP, 2020); https://www.asa.org.uk/uploads/assets/68dd32b5-ae6a-4993-820a3ff8f1163b8e/ASA-CAP-2019-Annaul-Report-Full-Version-Singles.pdf
19. *Internet Advertising Revenue Report: Full Year 2019 Results & Q1 2020 Revenues* (IAB/PWC, 2020); https://www.iab.com/wp-content/uploads/2020/05/FY19-IAB-Internet-Ad-Revenue-Report_Final.pdf
20. Dave, P. Social media giants warn of AI moderation errors as coronavirus empties offices. *Reuters* (16 March 2020).
21. Newton, C. The coronavirus is forcing tech giants to make a risky bet on AI. *The Verge* https://www.theverge.com/interface/2020/3/18/21183549/coronavirus-content-moderators-facebook-google-twitter (2020).
22. Heilweil, R. Facebook is flagging some coronavirus news posts as spam. *Vox* https://www.vox.com/recode/2020/3/17/21183557/coronavirus-youtube-facebook-twitter-social-media (2020).
23. Ali, M. et al. Discrimination through optimization: how Facebook's ad delivery can lead to skewed outcomes. *Proc. ACM Hum. Comput. Interact.* **3**, 1–30 (2019).
24. Datta, A., Datta, A., Makagon, J., Mulligan, D. K. & Tschantz, M. C. Discrimination in online advertising: a multidisciplinary inquiry. *Proc. Mach. Learn. Res.* **81**, 20–34 (2018).
25. Kim, P. T. & Scott, S. Discrimination in online employment recruiting symposium: law, technology, and the organization of work. *St. Louis Univ. Law J* **63**, 93–118 (2018).
26. Sweeney, L. Discrimination in online ad delivery. *ACM Queue* **11**, 36 (2013).
27. Wachter, S. *Affinity Profiling and Discrimination by Association in Online Behavioural Advertising* (SSRN, 2019); https://papers.ssrn.com/abstract=3388639
28. *EASA Best Practice Recommendation on Online Behavioural Advertising* (EASA, 2016).
29. *YourAdChoices* (Digital Advertising Alliance); https://youradchoices.com/about
30. *YourOnlineChoices* (EDAA); https://www.youronlinechoices.com/uk/about-behavioural-advertising
31. Kim, T., Barasz, K. & John, L. K. Why am I seeing this ad? The effect of ad transparency on ad effectiveness. *J. Consum. Res* **45**, 906–932 (2019).
32. Andreou, A. et al. Investigating ad transparency mechanisms in social media: a case study of facebook's explanations. In *Proc. 2018 Network and Distributed System Security Symposium* (Internet Society, 2018); https://doi.org/10.14722/ndss.2018.23191
33. Hern, A. & Ledegaard, F. H. Children 'interested in' gambling and alcohol, according to Facebook. *The Guardian* (9 October 2019).
34. *Harnessing New Technology to Tackle Irresponsible Gambling Ads Targeted at Children* (ASA, 2019); https://www.asa.org.uk/news/harnessing-new-technology-gambling-ads-children.html
35. *ASA System Submission to the Women and Equalities Committee Inquiry on Body Image* (ASA, 2020).
36. Medina, D. J. Hermeneutical injustice and polyphonic contextualism: social silences and shared hermeneutical responsibilities. *Soc. Epistemol.* **26**, 201–220 (2012).
37. Medina, J. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination* (Oxford Univ. Press, 2013).
38. Fricker, M. in *The Epistemic Dimensions of Ignorance* (eds Peels, R. & Blaauw, M.) 160–177 (Cambridge Univ. Press, 2016); https://doi.org/10.1017/9780511820076.010
39. Angwin, J., Larson, J., Mattu, S. & Kirchner, L. Machine bias. *ProPublica* https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=LrqwtD3z1Jth8ag9cay6c0yzKoghtu9C (2016).
40. Wachter, S., Mittelstadt, B. & Russell, C. *Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI* (SSRN, 2020); https://doi.org/10.2139/ssrn.3547922
41. Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. On the dangers of stochastic parrots: can language models be too big? In *Proc. 2021 ACM Conference on Fairness, Accountability, and Transparency* 610–623 (ACM, 2021); https://doi.org/10.1145/3442188.3445922
42. Mittelstadt, B. Auditing for transparency in content personalization systems. *Int. J. Commun.* **10**, 12 (2016).
43. Chotiner, I. The underworld of online content moderation. *The New Yorker* (5 July 2019); https://www.newyorker.com/news/q-and-a/the-underworld-of-online-content-moderation
44. Gillespie, T. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale Univ. Press, 2018).
45. Strubell, E., Ganesh, A. & McCallum, A. Energy and policy considerations for deep learning in NLP. In *Proc. 57th Annual Meeting of the Association for Computational Linguistics* 3645–3650 (ACL, 2019).
46. Gray, M. L. & Suri, S. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass* (Houghton Mifflin Harcourt, 2019).
47. Keegan, J. Introducing 'split screen'. *The Markup* https://themarkup.org/citizen-browser/2021/03/11/introducing-split-screen (2020).
48. Split screen: how different are Americans' Facebook feeds? *The Markup* https://themarkup.org/citizen-browser/2021/03/11/split-screen?feed=biden_trump (2021).
49. *Ad Library* (Facebook); https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=GB (accessed May 2021).
50. *Ads Transparency* (Twitter, Inc.); https://business.twitter.com/en/help/ads-policies/product-policies/ads-transparency.html (accessed May 2021).
51. Nguyen, C. T. Echo chambers and epistemic bubbles. *Episteme* **17**, 141–161 (2020).
52. Pariser, E. *The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think* (Penguin, 2012).
53. Bruns, A. Filter bubble. *Internet Policy Rev.* **8**, 1–14 (2019).
54. Borgesius, F. J. Z. et al. Should we worry about filter bubbles? *Internet Policy Rev.* **5**, 1–16 (2016).
55. *Online Platforms and Digital Advertising: Market Study Final Report* (CMA, 2020); https://assets.publishing.service.gov.uk/media/5efc57ed3a6f4023d242ed56/Final_report_1_July_2020_.pdf
56. *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and Amending Directive 2000/31/EC* COM/2020/825 final (European Commission, 2020); https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1608117147218&uri=COM%3A2020%3A825%3AFIN
57. Toubiana, V., Narayanan, A., Boneh, D., Nissenbaum, H. & Barocas, S. Adnostic: privacy preserving targeted advertising. In *Proc. Network and Distributed System Symposium* (NDSS, 2010).
58. Charting a course towards a more privacy-first web. *Google* https://blog.google/products/ads-commerce/a-more-privacy-first-web/ (2021).
59. Taylor, L. in *Group Privacy: New Challenges of Data Technologies* (eds Taylor, L. et al.) 13–36 (Springer International, 2017); https://doi.org/10.1007/978-3-319-46608-8_2
60. Mittelstadt, B. From individual to group privacy in big data analytics. *Phil. Technol.* **30**, 475–494 (2017).
61. Bygrave, L. A. *Data Protection Law: Approaching its Rationale, Logic and Limits* (Kluwer Law International, 2002).

62. Mantelero, A. in *Group Privacy: New Challenges of Data Technologies* (eds Taylor, L. et al.) 139–158 (Springer International, 2017); https://doi.org/10.1007/978-3-319-46608-8_8
63. Cohen, J. & Fung, A. in *Digital Technology and Democratic Theory* (eds Landemore, H. et al.) 23–61 (Univ. Chicago Press, 2021).
64. Landemore, H. Beyond the fact of disagreement? The epistemic turn in deliberative. *Democracy Soc. Epistemol.* **31**, 277–295 (2017).
65. Estlund, D. & Landemore, H. in *The Oxford Handbook of Deliberative Democracy* (eds Bächtiger, A. et al.) 112–131 (Oxford Univ. Press, 2018); https://doi.org/10.1093/oxfordhb/9780198747369.013.26
66. Dryzek, J. S. et al. The crisis of democracy and the science of deliberation. *Science* **363**, 1144–1146 (2019).
67. Statt, N. Facebook will remove 5,000 ad targeting categories to prevent discrimination. *The Verge* https://www.theverge.com/2018/8/21/17764480/facebook-ad-targeting-options-removal-housing-racial-discrimination (2018).
68. Dayen, D. Ban targeted advertising. *The New Republic* (10 April 2018).
69. Lewis, R. A. & Rao, J. M. The unfavorable economics of measuring the returns to advertising. *Q. J. Econ.* **130**, 1941–1973 (2015).
70. Angwin, J. Auditing the algorithms of disinformation. *The Markup* https://www.getrevue.co/profile/themarkup/issues/auditing-the-algorithms-of-disinformation-284735 (2020).
71. The Citizen Browser Project—auditing the algorithms of disinformation. *The Markup* https://themarkup.org/citizen-browser (2020).
72. How We Built a facebook inspector. *The Markup* https://themarkup.org/citizen-browser/2021/01/05/how-we-built-a-facebook-inspector (2021).
73. Bloustein, E. J. Group privacy: the right to huddle. *Rutgers Camden Law J.* **8**, 219–283 (1976).
74. Bloustein, E. J. & Pallone, N. J. *Individual and Group Privacy* (Routledge, 2018); https://doi.org/10.4324/9781351319966
75. van der Sloot, B. in *Group Privacy: New Challenges of Data Technologies* (eds Taylor, L. et al.) 197–224 (Springer International, 2017); https://doi.org/10.1007/978-3-319-46608-8_11
76. Horwitz, J. Facebook seeks shutdown of NYU research project into political ad targeting. *Wall Street Journal* (23 October 2020).
77. Naughton, J. Facebook has good reasons for blocking research into political ad targeting. *The Guardian* (31 October 2020).
78. *2018 European Trends in Advertising Complaints, Copy Advice and Pre-clearance* (EASA, 2018); https://www.easa-alliance.org/sites/default/files/2018%20European%20Trends%20in%20Advertising%20Complaints%2C%20Copy%20Advice%20and%20Pre-clearance.pdf
79. How Facebook ads get approved. *Facebook for Business* https://www.facebook.com/business/a/ad-review-process (accessed 9 June 2021).
80. *HM Government Communication Service* (GCS, 2020); https://gcs.civilservice.gov.uk/about-us/what-we-do/
81. *HM Government Communication Plan 2019/20* (GCS, 2019); https://communication-plan.gcs.civilservice.gov.uk/
82. Merrill, J. B. Does facebook still sell discriminatory ads? *The Markup* https://themarkup.org/ask-the-markup/2020/08/25/Does-Facebook-Still-Sell-Discriminatory-Ads (2020).

Acknowledgements

This work of the Governance of Emerging Technologies research programme at the Oxford Internet Institute has been supported by British Academy Postdoctoral Fellowship grant number PF2\180114 and grant number PF\170151, the Luminate/Omidyar Group and the Miami Foundation.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to S.M.

Peer review information *Nature Machine Intelligence* thanks Jathan Sadowski and the other, anonymous, reviewer(s) for their contribution to peer review of this work.