July 8, 2021

To: Angela Nissyrios and Simon Deeble
Antitrust Division, CMA

From: Professor Helen Nissenbaum, Cornell Tech
Professor Lee McGuigan, University of North Carolina

Subj: Comments on proposed Privacy Sandbox commitments regarding user welfare
Case reference 50972

I. Introduction and overview

We offer these comments in response to the CMA's call for public consultation, specifically addressing the potential "imposition of unfair terms on Chrome's web users," as a result of Google's proposed Privacy Sandbox.

Exploitative data practices establish an acutely uneven playing field between consumers and the companies that execute them. Google's Privacy Sandbox, which represents a significant change to Chrome, is a case in point. Despite being presented as "privacy first," the so-called "Privacy Sandbox" exposes Chrome's users to significant privacy risks. Proposed design safeguards do not mitigate the most severe forms of these privacy risks.

A. About us[1]

Helen Nissenbaum and Lee McGuigan are researchers focusing on the impact of technology and data usage on consumer welfare, with special attention to privacy harms.

- Helen Nissenbaum is a Professor of Information Science and the Director of the Digital Life Initiative at Cornell Tech in New York City. For more than three decades she has been a leading authority on the social implications of technology. She developed a privacy theory "contextual integrity," in books and articles that has been cited thousands of times by academic and industry researchers across the disciplinary spectrum, and has influenced lawmakers around the world, including in US, Europe, India, and China.
- Lee McGuigan is an assistant professor in the Hussman School of Journal and Media, and a Faculty Research Fellow in the Center for Information, Technology, and Public Life, at the University of North

[1] In collaboration with Tyler Zhu who provided careful reading and invaluable suggestions. The authors also thank Isra Hussain, Arzu Mammadova, and Megan Wilkins for exceptional research assistance.

Carolina at Chapel Hill. He researches and writes about the impact of advertising technology and its impact on consumer behavior and choices.

B. Summary of views

It goes without saying that change is welcome. The status quo of the web is a privacy nightmare; personal information circulates among a baffling assortment of unfamiliar companies who use this data to build up detailed consumer profiles, which not only may be inaccurate but may enable unfair discrimination.[2] [3] [4] We would be happy to see the demise of third-party tracking and individual-level identification in web advertising.

But Google deserves no credit for leadership on privacy, despite its efforts to present itself as privacy forward. Other major web browsers, such as Firefox and Safari, already block third-party cookies by default.[5] Further, analysts question whether Google's methods of anonymization would effectively prevent motivated actors from tracking and profiling individuals.[6] There is the risk that Google is using the veneer of privacy to disguise an otherwise naked power play, as the CMA has rightly identified. These new plans may make the digital advertising industry even more reliant on Google's data assets, worsening this already anti-competitive environment.[7] Through FLoC, Google has anointed itself a first-party to all browsing on Chrome. "Don't worry," says the fox to the farmer, "I will guard your hens."

We will focus on FLoC because of its promise as an alternative to online privacy's bogeyman: third-party cookies. Even if we ignore these concerns *and* imagine that FLoC "works" as claimed, the question remains: does FLoC measure up to a rigorous definition of privacy? Are Google's plans, and the notions of privacy supporting them, adequate for confronting how digital surveillance operates today? We believe they are not. They fail to address the normative, democratic goals at the heart of privacy, which go beyond issues of secrecy and consent. Google's Privacy Sandbox therefore fails in its basic premise of achieving a more meaningful privacy outcome

---

[2] Brave, "Selected evidence submitted to data protection authorities to demonstrate RTB's GDPR problems," available at https://brave.com/rtb-evidence/.
[3] Cox, Joseph, "Tech Giants Won't Name Foreign Companies They Give US 'Bidstream' Data To," Vice, 9 April 2021, available at https://www.vice.com/en/article/k78ewv/bidstream-data-google-twitter-att-verizon-foreign.
[4] Singer, Natasha, "Mapping, and Sharing, the Consumer Genome," New York Times, 16 June 2012, available at https://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html.
[5] Wood, Marissa, "Today's Firefox Blocks Third-Party Tracking Cookies and Cryptomining by Default," Mozilla, 3 September 2019, available at https://blog.mozilla.org/blog/2019/09/03/todays-firefox-blocks-third-party-tracking-cookies-and-cryptomining-by-default/; Wilander, John, "Full Third-Party Cookie Blocking and More," WebKit, 24 March 2020, available at https://webkit.org/blog/10218/full-third-party-cookie-blocking-and-more/.
[6] Rescorla, Eric, "Privacy analysis of FLoC," Mozilla, 10 June 2021, available at https://blog.mozilla.org/en/mozilla/privacy-analysis-of-floc/; Kaye, Kate, "As ad tech firms test ways to connect Google's FLoC to other data, privacy watchers see fears coming true," Digiday, 10 June 2021, available at https://digiday.com/marketing/as-ad-tech-firms-test-ways-to-connect-googles-floc-to-other-data-privacy-watchers-see-fears-coming-true/; Cyphers, Bennett, "Google's FLoC is a terrible idea," Electronic Frontier Foundation, 3 March 2021, available at https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea.
[7] Srinivasan, Dina, "Why Google Dominates Advertising Markets: Competition Policy Should Lean on the Principles of Financial Market Regulation," *Stanford Technology Law Review 24*(1) (2020): 55-175, available at https://law.stanford.edu/publications/why-google-dominates-advertising-markets/.

for its users than third-party cookies. We will also show how Google's proposed commitments to "transparency and control" are unlikely to meaningfully resolve any of the risks we highlight.

II. Privacy Sandbox Imposes Unfair Terms on Chrome Users

A. What does Privacy Sandbox mean by privacy?

What does privacy mean to Google? It is hard to tell. Announcements about the company's "privacy-first" future don't express a clear definition of the term, and the Privacy Sandbox includes multiple plans and proposals.[8] We can only infer how Google views privacy by analyzing properties of the new technologies it proposes to replace third-party cookies, including "FLoC" (typically, Federated Learning of Cohorts). By examining the design of FLoC, we infer that Google's notion of privacy is characterized by the following three components:[9]

1) The first is getting rid of "third-party" tracking of user behavior. Under the Privacy Sandbox, Chrome would disallow third party tracking cookies while continuing to allow "first parties," granted formal consent by the user, to pursue respective data practices. Instead of countless websites, advertisers, and surveillance companies following consumers and building dossiers of behavior, Chrome would perform tracking and profiling on the user's device.

2) The second is that no individualized ID would travel with Chrome users across the Web. Under FLoC, Google's Chrome browser would monitor browsing histories, while machine learning algorithms sort users into cohorts with others who share similar browsing behaviors. Chrome then assigns a FLoC cohort ID to each user, which it exposes to advertisers and ad tech vendors for targeting purposes. Browsing histories ("raw data") never leave users' browsers. Google makes a big deal of this, arguing that individuals, spared exposure, can "hide in the crowd." Importantly, a FLoC ID label does not, at least on its face, contain semantic meaning.

3) The third is that "sensitive" information about the user would not be leaked to third parties. Under the Privacy Sandbox, if a significant proportion of a particular FLoC is associated with sensitive topics and websites, that FLoC would be disabled and its members dispersed.

B. The Privacy Sandbox, even with the proposed commitments, ignores the critical role of context on data collection, usage, and transfer

We suggest that Google's maneuvers should be measured against a definition of privacy that centers around the *appropriate flow of information*.[10] That means ensuring that data about us is collected and used only for purposes that

[8] Temkin, David, "Charting a course towards a more privacy-first web," Google Ads & Commerce Blog, 3 March 2021, available at https://blog.google/products/ads-commerce/a-more-privacy-first-web/.
[9] "Federated Learning of Cohorts," GitHub, available at https://github.com/WICG/floc; Dutton, Sam, "What is FLoC?" Web.Dev, 14 May 2021, available at https://web.dev/floc/.
[10] Nissenbaum, Helen, Privacy in Context: Technology, Policy, and the Integrity of Social Life. (Stanford, California: Stanford University Press, 2009).

align with our expectations and the norms and goals associated with the social contexts that make up our lives. This is a principle that is embodied in various pieces of privacy legislation and regulatory bills, including the GDPR.[11] [12]

For example, if we disclose something about romantic engagements, we would expect this disclosure to help us find a wedding venue, or maybe a marriage counselor, and we would not expect that it could affect the job opportunities available to us. This approach to privacy also demands that the parties interacting with personal data handle it in ways that balance the interests of affected parties *and* promote societal objectives. The end goals of privacy are not secrecy or anonymity, or even control by data subjects alone; instead, privacy is important also because it helps to secure fundamental political, moral, and human values. These values help us decide as a society what sorts of relationships we want personal data to materialize and maintain, and what justifications are valid for classifying people and groups. In other words, privacy is about defining the legitimacy of various regimes for using data to organize both private and public affairs. This is a theory of privacy as *contextual integrity*, where information flows respect the norms of social contexts and contribute to the integrity of social life.[13]

Contextual integrity demands an understanding of privacy that focuses on how personal information, and the technologies that generate and exploit it, are used to assemble and classify social groups and mediate social life. The legitimacy of these uses is based on how they organize and sustain social spheres, not on how effectively they organize populations for targeted advertising, or for the profit of other vested interests. There is no denying that information flows affect values, motivations, and power relationships that define how people come together, form identities, and participate in communities. These are the stakes of a positive definition of privacy, if Google plans to make a good-faith commitment.

If we examine Google's proposal under the lens of contextual integrity, it becomes clear that the Privacy Sandbox perpetuates a system of behavioral classification that continues to violate the privacy norms of various social contexts. We will show below how the three components that make up Google's notion of privacy do not stand up to critical scrutiny.

---

[11] For example, see GDPR, Recital 47 (emphasis added): "At any rate the existence of a legitimate interest would need careful assessment including whether a data subject can reasonably expect at the time and in the context of the collection of the personal data that processing for that purpose may take place." Available at https://www.privacy-regulation.eu/en/recital-47-GDPR.htm.

[12] See also "Respect for Context" principle outlined in the Consumer Privacy Bill of Rights proposal published by the Obama White House in 2012: "Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data." Available at https://obamawhitehouse.archives.gov/sites/default/files/privacy-final.pdf.

[13] Nissenbaum, Helen, Privacy in Context: Technology, Policy, and the Integrity of Social Life. (Stanford, California: Stanford University Press, 2009).

1) The first and third-party distinction is a privacy red herring

Declaring itself a "first-party" to all browsing activity, Google banks on the idea that data practices deemed illegitimate if performed by disparate third-parties become appropriate simply because they are performed by one, viz. Google. It is a step in the right direction to deny myriad unknown trackers' access to personal data, but contextual integrity demands more. If the sin of third-party trackers is that they enable cross-site surveillance, a single first party that does the same and also federates your email, work-flow, online search, location and other physical traces, and much more, is arguably even more dangerous. With first parties like Google – the joke may go – who needs third-parties?

According to contextual integrity, the first and third party distinction is not necessarily a relevant distinction for privacy – a point with which the CMA and ICO clearly concur, according to a recently published joint statement.[14] What is relevant for privacy are the roles played by those parties and their standing in relation to data subjects, as defined within respective social contexts – for example, marriage counsellor to clients, employer to employee, physician to patient etc. By monitoring and classifying people across all the websites participating in the FLoC program, Google is collapsing the diverse social contexts on the Web (health, work, family, commerce, finance, education, politics, etc.) into one, all-encompassing universe, anointing itself the master of the whole.

2) Aggregation and the usage of cohorts does not remove the risk of privacy harm

The crux of privacy is the appropriate flow of information. That means personal information – including details we volunteer, traces of behavior collected by the devices and software we use, and the inferences that technical systems draw about us (and others like or unlike us) – is collected and circulated in ways that align with the expectations and norms people associate with particular social domains. Even if the information in question is inferred from "raw data," which never leaves the users device, contextual integrity still requires the inferred information flow to be appropriate.

The assignment of users to cohorts involves inferences drawn from observations of browsing histories, and these inferences are then communicated to websites and advertisers in the form of a cohort ID. This is information about your similarity to or difference from other users, and it is shared and used to define you in contexts other than ones you may have expected when you visited websites looking for a restaurant, an education, or an insurance policy.

[14] Paragraph 82 in *Competition and data protection in digital markets: a joint statement between the CMA and the ICO*: "It is important to note, therefore, that neither competition nor data protection regulation allows for a 'rule of thumb' approach, where intra-group transfers of personal data are permitted while extra-group transfers are not. Under both data protection law and competition law, a careful case-by-case assessment is needed, regardless of the size of a company, the business model adopted, or the nature of any processing activity." Available at https://ico.org.uk/media/about-the-ico/documents/2619797/cma-ico-public-statement-20210518.pdf.

FLoC is therefore *still* a system of behavioral classification. Whether people are identified by cookies or cohorts, it remains the case that their treatment and the opportunities made available to them are shaped by their actions anywhere and everywhere on the Web and the conclusions various entities draw from those actions. A case in point is the health insurance industry determining health status, and subsequently insurance premiums, based on non-health information such as plus-sized clothing purchases or TV viewing habits.[15] In another example, researchers found that Google's advertising algorithm discriminated against female visitors when displaying job advertisements, demonstrating that users do not avoid risks of unfair treatment simply because they are hiding "in a crowd."[16] Studies on disinformation also observe that the "strategic communication services offered by digital ad platforms like Google" has enabled an "increasingly sophisticated toolkit for digital influence peddling." Such systems can be weaponized to "identify and target weak points where groups … are most vulnerable to strategic influence" via targeted disinformation.[17]

3) Disabling 'sensitive' cohorts does not prevent privacy harms

Privacy is not about hiding or keeping secrets. Privacy is not merely about "private" matters or a promise to avoid "sensitive" categories. All data holds potential to cause harm, if distributed inappropriately.

Consider Google's approach: FLoC cohorts that visit pages (presumably above a frequency threshold of a given level) associated with "sensitive topics, perhaps medical websites or websites with political or religious content" would be disabled.[18] For the approach to work, Google will need to first define and maintain a list of categories it judges to be "sensitive." However, if we take sensitivity to mean likelihood of causing hurt or harm, what is sensitive or not is highly contextual. What is considered a sensitive category in one context, disclosed to some parties and under certain conditions, may not be so in another context. For example, disclosing one's income level to listing platforms for rental properties is generally unproblematic under conditions of confidentiality, but may be risky when this information is shared with payday loan advertisers.[19] A user may not mind a retailer targeting them with ads for plus-sized clothing based on prior purchases of such products, but would likely

[15] Allen, Marshall, "Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates," ProPublica, 17 July 2018, available at https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates.
[16] Datta, Amit, Michael Carl Tschantz, and Anupam Datta, "Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination," *Proceedings on Privacy Enhancing Technologies 2015* (1) (2015): 92, 112, available at https://www.andrew.cmu.edu/user/danupam/dtd-pets15.pdf.
[17] Nadler, Anthony, Matthew Crain, and Joan Donovan, "Weaponizing the Digital Influence Machine: The Political Perils of Online Ad Tech," Data & Society Research Institute, October 2018, pp. 6-7, available at https://datasociety.net/wp-content/uploads/2018/10/DS_Digital_Influence_Machine.pdf.
[18] Vale, Marshall, "Privacy, sustainability and the importance of 'and'," Google Blog, 30 March 2021, available at https://blog.google/products/chrome/privacy-sustainability-and-the-importance-of-and/.
[19] Dayen, David, "Google said it would ban all payday loan ads. It didn't," The Intercept, 7 October 2016, available at https://theintercept.com/2016/10/07/google-said-it-would-ban-all-payday-loan-ads-it-didnt/.

object if the same information is used by health insurers to raise their premiums.[20] Since there are no contextual boundaries built into the FLoC setup in Google's Privacy Sandbox, any fixed list of sensitive categories will not protect consumers against harms of disclosure to the wrong types of actors.

Even if we suppose that a global list of sensitive topics or categories is coherent, sensitive information may nevertheless be inferred about cohorts via non-intuitive, auxiliary signals instead of obvious cues such as visits to sensitive sites. Predictive analytics has been shown to uncover surprising correlations. For example, researchers found that Internet usage patterns alone, such as frequency of email checking, can be used to identify students suffering from depression without needing to know what sites or content they visited.[21] Since there is no inherent semantic labeling to FLoC cohort IDs, it may become even more difficult for users and regulators to notice that such inferences are happening.

C. Google's promise of "transparency and control" is empty

To assuage concerns of unfair treatment of Chrome users, Google promises "transparency and control." As for transparency, it's the same promise we've been hearing since at least 2011[22], yielding a 4000-words long privacy notice that far surpasses anyone's ability to engage meaningfully with it.[23] [24] Google answers questions no-one asks, but demurs when we ask for information that really matters. It is difficult to imagine that explaining the Privacy Sandbox will be different. On a personal note: the authors of this submission have spent countless hours reading Google's blogs, imposing on our technology-savvy colleagues, and picking the brains of online advertising experts to arrive at some semblance of an understanding of Google's Privacy Sandbox. We still would not stake our lives on it. It is unclear how the degree of transparency would be achieved that might "level the playing field" for Chrome users so they might grasp the workings, meaningfully deliberate, and make proper choices (i.e., not take-it-or-leave it).

As for choice, the outlook is no better. Despite its previous promise to improve user choice, Google combined more than 60 of its privacy policies across different consumer-facing products into a single umbrella policy in

---

[20] Allen, Marshall, "Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates," ProPublica, 17 July, 2018, available at https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates.

[21] Katikalapudi, Raghavendra et al., "Associating Internet Usage with Depressive Behavior among College Students," *IEEE Technology and Society Magazine* 31(4) (2012): 73–80, doi:10.1109/MTS.2012.2225462.

[22] For example, see: Wakefield, Jane,"Google chairman Schmidt promises privacy controls," BBC, 18 May 2011, available at https://www.bbc.com/news/technology-13439963; Melanson, Don, "Google promises 'greater transparency' for targeted ads, gives users more control over them," Engadget, 1 November 2011, available at https://www.engadget.com/2011-11-01-google-promises-greater-transparency-for-targeted-ads-gives-u.html.

[23] Warzel, Charlie and Ash Ngu, "Google's 4,000-Word Privacy Policy Is a Secret History of the Internet," The Privacy Project, New York Times, 10 July 2019, available at https://www.nytimes.com/interactive/2019/07/10/opinion/google-privacy-policy.html.

[24] Navarro, Kevin L, "We Read 150 Privacy Policies. They Were an Incomprehensible Disaster," The Privacy Project, New York Times, 12 June 2019, available at https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html.

2012 to offer what regulators called "a fairly binary and somewhat brutal choice" for users.[25] Google also manipulated users by offering the illusion of control. Recent unsealed evidence from a US court case shows that Google designed its privacy settings in such a way that it is nearly impossible for a user to fully switch off location data tracking with Google's products.[26] Google engineers remarked that "[t]he current UI feels like it is designed to make things possible, yet difficult enough that people won't figure it out."[27]

It is well-known among privacy activists, advocates, researchers, and regulators that privacy protection through transparency and choice has been an abject failure; admittedly, Google is not alone in perpetuating this illusion. Research from multiple disciplines has repeatedly demonstrated that transparency and control do not lead to informed privacy choices from users. This is a fundamental limitation with a "self-management" approach to privacy, which imposes impossible burdens on the individual.[28] Privacy scholars have long noted that it is difficult for users to meaningfully engage with privacy notices to make rational trade-offs.[29] Behavioral economists note that individuals are often unable to process and act optimally on privacy decisions, even if they have access to complete information.[30] Users will often be "overwhelmed with the task of identifying possible outcomes related to privacy threats and means of protection" and "face difficulties to assign accurate likelihoods to those states."[31] Consequently, users tend to defer to simplified mental models and shortcuts to make these decisions. For example, empirical research shows users frequently misinterpret the content of privacy notices, projecting their privacy expectations onto notices regardless of their actual content.[32] Even when information is

[25] Blagdon, Jeff, "Google's controversial new privacy policy now in effect," The Verge, 1 March 2012, available at https://www.theverge.com/2012/3/1/2835250/google-unified-privacy-policy-change-take-effect.
[26] Paragraph 45 in State of Arizona v. Google LLC, *Complaint for Injunctive and Other Relief,* available at https://www.azag.gov/sites/default/files/2021-05/Complaint%20%28redacted%29.pdf.
[27] Lopatto, Elizabeth, "Even Google engineers are confused about Google's privacy settings," The Verge, 26 August 2020, available at https://www.theverge.com/2020/8/26/21403202/google-engineers-privacy-settings-lawsuit-arizona-doubleclick.
[28] For example, see Solove, Daniel J, "Privacy Self-Management and the Consent Dilemma," *Harvard Law Review* 1880 (2013), available at https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=2093&context=faculty_publications.
[29] Studies since as far back as 2006 shows that only a minority of Internet users actually read a priva cy notice. For example, see TRUSTe, "TRUSTe/TNS Survey Shows A Majority of Internet Users Believe They Know How to Protect Their Privacy Online, Most Fail to Take Needed Actions," 6 December 2006, archived at https://web.archive.org/web/20070303111329/http://www.truste.org/about/press_release/12_06_06.php.
[30] Acquisiti, Alessandro and Jens Grossklags, "What Can Behavioral Economics Teach Us about Privacy?" *Digital Privacy: Theory, Technologies and Practices*, 2007, pp. 6-7, available at https://www.heinz.cmu.edu/~acquisti/papers/Acquisti-Grossklags-Chapter-Etrics.pdf.
[31] Acquisiti, Alessandro and Jens Grossklags, "What Can Behavioral Economics Teach Us about Privacy?" *Digital Privacy: Theory, Technologies and Practices*, 2007, pp. 6-7, available at https://www.heinz.cmu.edu/~acquisti/papers/Acquisti-Grossklags-Chapter-Etrics.pdf.
[32] Martin, Kirsten, "Privacy Notices as Tabula Rasa: An Empirical Investigation into how Complying with a Privacy Notice is Related to Meeting Privacy Expectations Online," *Journal of Public Policy & Marketing 34* (2) (2015): 210-227, available at https://journals.sagepub.com/doi/abs/10.1509/jppm.14.139?journalCode=ppoa.

presented in a streamlined manner to aid information processing, consumers often still react to such information in unexpected and irrational ways.[33]

III. Concluding Remarks

Google has presented its Privacy Sandbox as pro-consumer. Without explicitly committing to a definition of privacy, it claims to be "privacy first" on the basis of three promises: eradicating third party trackers, advertising to cohorts not individuals, and preventing disclosure of sensitive activities to advertisers. It promises also to maintain transparency and choice. We have demonstrated that none of these stands up to scrutiny.

As long as Google's business model promises to advertisers that it can make browser-users intelligible and profitable, it makes little difference that users are identified individually, or as belonging to a particular FLoC. As an advertising firm that simultaneously controls the browser and captures users' activity data through it, Google will inevitably set policies that prioritize interests of its paying customers, i.e., advertisers, even if this means distinguishing good prospects from poor prospects results in some individuals getting the proverbial "short end of the stick."

Protecting people from the harms of third-party tracking is a start toward reform, but it reckons with a mere fraction of the promise, pitfalls, and power dynamics of data governance. Google's continued investment in totalizing technologies and business models that sort people according to their inferred value to advertisers is antithetical to a meaningful conception of privacy. *That Google has chosen the browser as the vehicle for delivering this totalizing technology is particularly poignant, considering that the browser is one of the most intimate systems that people depend on for online transactions, communications, and interactions.* To make good on its promise of *privacy-first*, Google would, *first*, respect the privacy norms of disparate social domains such as labor, health, finance etc., and it would come to grips with *why* these norms are important for the good of individuals and the integrity of social life.

[33] For example, empirical research shows that consumers may not abort software installations, despite well-presented information warning them of the dangerous malware. See: Good, Nathaniel, et al., "Noticing Notice: A largescale experiment on the timing of software license agreements," *Proceedings of SIGCHI conference on Human factors in Computing Systems* (CHI'07), April 2007, pp. 607-616.