PARTY FIRST, ASK QUESTIONS LATER: INTERROGATING THE PRIVACY IMPLICATIONS OF FIRST-PARTY DATA COLLECTION

Conor Kane [*]

CITE AS: 7 GEO. L. TECH. REV. 251 (2023)

TABLE OF CONTENTS

I. Introduction 251
II. How and Why Google is Changing Data Collection 253
III. From Third-Party to First-Party 257
A. Privacy concerns ignored and created by the shift to first-party data 260
B. Reconceptualizing the contours of the first-party relationship 264
1. Antitrust 264
2. Trust and Information Fiduciaries 264
3. Data Minimization (and Trust Minimization) 266
IV. Conclusion 271

I. INTRODUCTION

[*] J.D., Georgetown University Law Center 2023. Thanks to Professor Julie Cohen, the Nebraska Law Review, and the Nebraska Governance and Technology Center.

In January 2020, Google announced plans to discontinue support for the third-party cookie on their popular Chrome browser within the next two years.[1] The third-party cookie is a widely-relied-upon tool that enables marketers to gather information about consumer behavior across the Internet. Google gave privacy justifications for this move, claiming that it wanted to "make the web more private and secure."[2] Given the amount of tracking that third-party cookies enable, such justifications are not entirely off-base. Nevertheless, Google's announcement created great concern among marketers, who worried about how they would advertise in a post-cookie world. The announcement also garnered justified antitrust scrutiny from academics and some European regulators because of Google's ability to simultaneously set the rules to and play the AdTech game. But the potential privacy problems this move would create have yet to be as thoroughly examined.

By cutting off access to one of marketers' key sources of consumer information, Google has inspired an industry-wide shift in data collection practices. Marketers are reorienting their data-driven marketing from relying heavily on purchased third-party data to aggregating as much first-party data, data collected directly from users, as possible. Marketers gather this information by asking consumers to provide personal information or by observing consumer behavior via the company's digital properties. Data gathered through first-party relationships are often described as more pro-privacy and as a way to gain consumers' trust. From a business standpoint, marketers believe that "owning" more of their data, rather than continuing to rent or buy most of it from data brokers, will give them more control over their marketing and a better understanding of their customers in a post-cookie world.

This Note examines Google's stated and potentially unstated justifications for ending support for the third-party cookie in addition to the ripple effects that this move creates for data collection across the digital advertising ecosystem. It argues that Google's allegedly pro-privacy move and marketers' allegedly pro-privacy switch to first-party data both ignore and create privacy harms. The only way to protect privacy in the era of first-party data (and protect against future shifts in collection techniques) is to reconceptualize the corporation/consumer data relationship. On a broader level, this Note aims to provide a warning about the nature of shifting data collection practices. New privacy concerns and regulations beget new data collection practices. The data collection "party" never stays in one place for

---

[1] Justin Schuh, *Building a More Private Web: A Path Towards Making Third Party Cookies Obsolete*, CHROMIUM BLOG (Jan. 14, 2020), https://blog.chromium.org/2020/01/building-more-private-web-path-towards.html [https://perma.cc/KGZ6-DE83].

[2] *Id*.

too long, and the tactics discussed in this Note will likely soon be outdated themselves. Moving away from one invasive practice does not automatically mean that new practices will be inherently pro-privacy. As such, the roving nature of data collection techniques requires a larger reconfiguration rather than a tactic-by-tactic approach.

Part II examines how and why Google is changing data collection practices. Part III traces the shift from third-party cookie tracking to first-party data collection. Part IV identifies the various privacy harms ignored or created by the increase in first-party data collection. Part V looks to current conceptions of the consumer/company data relationship to identify ways to reconceive the proper structure of first-party data collection.

II. HOW AND WHY GOOGLE IS CHANGING DATA COLLECTION

The third-party cookie, "the common currency for the online ad industry,"[3] is a browser-based tool that allows marketers to track consumers across websites and target ads based on their behavior.[4] As of October 2020, roughly 80% of marketers relied on third-party cookies.[5] Citing concerns about cookie-based tracking's impact on privacy, Apple and Mozilla began taking steps to block third-party cookies in their browsers in 2017 and 2018, respectively.[6] However, Google, which operates Chrome, the world's most popular browser,[7] was slower to do the same.

In January 2020, Google announced that Chrome would stop supporting third-party cookies in two years.[8] This date has been pushed back multiple times, and as of July 2022, the change was slated to occur in the second half of 2024.[9] The January 2020 announcement, titled "Building a

---

[3] Damien Geradin, Dimitrios Katsifis & Theano Karanikioti, *Google as a de facto Privacy Regulator: Analyzing Chrome's Removal of Third-Party Cookies from an Antitrust Perspective* 38 (TILEC, Discussion Paper No. 2020-034, 2020).

[4] *What a World Without Third-Party Cookies Means*, EPSILON, https://www.epsilon.com/us/insights/trends/third-party-cookies [https://perma.cc/N3DG-9TGZ] (last visited Aug. 1, 2023).

[5] *Id*.

[6] Natasha Lomas, *Mozilla Beefs Up Anti-Cross-Site Tracking in Firefox, as Chrome Still Lags on Privacy*, TECHCRUNCH (Feb. 24, 2021, 7:14 AM), https://techcrunch.com/2021/02/24/mozilla-beefs-up-anti-cross-site-tracking-as-chrome-still-lags-on-privacy/ [https://perma.cc/PRG5-SR53].

[7] *Browser Market Share Worldwide*, STATCOUNTER, https://gs.statcounter.com/browser-market-share [https://perma.cc/8WLC-TLWP] (last visited Aug. 1, 2023[date]).

[8] Schuh, *supra* note 1.

[9] Anthony Chavez, *Expanding Testing for the Privacy Sandbox for the Web*, GOOGLE: THE KEYWORD (July 27, 2022), https://blog.google/products/chrome/update-testing-privacy-sandbox-web/[https://perma.cc/5923-5H7M].

more private web: A path towards making third-party cookies obsolete," stated that "[u]sers are demanding greater privacy."[10] It discussed Google's "Privacy Sandbox," an initiative meant to "develop a set of open standards to fundamentally enhance privacy on the web."[11] The Privacy Sandbox is Google's attempt to "make the web more private and secure for users, while also supporting publishers."[12] In order to do this, Google states that it is collaborating with "the web community" to avoid "undermining the business model of many ad-supported websites" in a post-cookie world.[13]

Google's move comes during a period of increasing privacy regulation. Europe's two major data protection laws, the General Data Protection Regulation (GDPR) and the ePrivacy Directive, regulate cookie usage. The GDPR covers data collection and processing, while the ePrivacy Directive covers information storage on a user's device.[14] Since third-party cookies are used by websites to store information on users' browsers, they are covered by the ePrivacy Directive, which requires user consent.[15] The information gathered by third-party cookie activity is covered by the GDPR, which also requires user consent.[16] In the U.S., the California Consumer Privacy Act (CCPA) regulates cookies as well, but does not require a company to obtain consent from a user before placing cookies on their browser.[17] However, the CCPA requires a business that sells personal information to make certain disclosures in its privacy policies.[18] The CCPA defines "sale" broadly enough to potentially cover the use of third-party cookies. As a result, corporations have been advised to ask for consent or provide disclosures and opt-outs to reduce the risk of liability.

Interestingly, neither of these provisions do not necessarily require Google to cut off support for third-party cookies in its browser. Rather, they increase procedural steps for websites to use third-party cookies, such as by requiring the acquisition of user consent or providing disclosures regarding sales of data. These laws may have increased public attention on the use of third-party cookies, but they do not ban their use outright. As Damien Geradin et al. put it, "Chrome's Privacy Sandbox is arguably motivated by privacy

---

[10] Schuh, *supra* note 1.
[11] *Id*.
[12] *Id*.
[13] *Id*.
[14] Geradin et al., *supra* note 3, at 28.
[15] *Id*. at 30.
[16] *Id*. at 31.
[17] DAVID ZETOONY ET AL., BRYAN CAVE LEIGHTON PAISNER LLC, CALIFORNIA CONSUMER PRIVACY ACT (CCPA): ANSWERS TO THE MOST FREQUENTLY ASKED QUESTIONS CONCERNING COOKIES AND ADTECH 2, 5 (2020), https://ccpa-info.com/wp-content/uploads/2019/08/Handbook-of-FAQs-Cookies.pdf [https://perma.cc/SL9K-PP2H].
[18] *Id*. at 12.

considerations, but we do not see how it is *necessary* to ensure compliance with the GDPR or other equivalent legislation."[19] Recently, Google has used the attention on privacy laws as cloud cover to create their own privacy-inspired "regulations" with far-reaching implications for the digital advertising industry.

Google's primary stated justification for ending support for the third-party cookie is to create a more private Internet, but the move leaves many other avenues for tracking users unaddressed. Brands can still individually identify users on their own sites and create first-party profiles based on visitors' behaviors.[20] Brands, however, will no longer be able to enrich these first-party profiles with the cross-site information that third-party cookies provide. This has created the scramble to reorient databases and tech stacks around first-party data, which is discussed further below, since this may soon be the main way for brands to "know" their customers. Given the depth of information that first-party data contains, the privacy benefits of removing third-party cookies are notable but limited. Although it is difficult to quantify, the change is inspiring a shift from a broad-but-shallow collection of consumer data to a narrow-but-deep collection. Although current privacy laws inspired this shift, they are not poised to address AdTech's response.

Although Google may claim that this move is driven by privacy concerns, the competitive advantages it would create make this claim suspect. Geradin et al. have carefully traced the ways that the proposal leaves many privacy harms unaddressed while benefitting Google. First, Google operates many consumer-facing properties on which it could continue to gather first-party data, such as Google Search, YouTube, and Gmail.[21] While other brands will no longer be able to conduct cross-site tracking, Google may continue to do so across some of the most valuable online real estate. Initial Privacy Sandbox proposals would even enable Chrome to continue tracking user behavior across the Internet. As Geradin et al. put it, "insofar as the user browses through Chrome, the open web becomes part of Google's logged-in environment."[22] In response to an investigation by the UK Competition and Markets Authority (CMA), Google committed to "not give preferential treatment or advantage to Google's advertising products or to Google's own sites," and stated that it would apply this commitment globally if CMA accepted it.[23] However, these commitments do not address Google's ability to

[19] Geradin et al., *supra* note 3, at 35 n.140.
[20] *Id*. at 41–42.
[21] *Id*. at 42.
[22] *Id*.
[23] Oliver Bethell, *Our Commitments for the Privacy Sandbox*, GOOGLE: THE KEYWORD (June 11, 2021), https://blog.google/around-the-globe/google-europe/our-commitments-privacy-sandbox/ [https://perma.cc/6F2E-NV5E].

"track users across the multiple services it owns and operates while denying others the same opportunity."[24] While cutting off the third-party cookie may seem like a way to limit consumer surveillance across the board, it would actually create a massive collection imbalance in Google's favor. As Jeremy Tillman of Ghostery, an anti-tracking browser extension development company, put it, "Google defines [privacy] as protecting against any data collection that it's not doing itself. [This] can be seen as a way to consolidate their own power."[25]

Google's ability to self-preference has drawn warranted scrutiny. The advertising industry has raised significant concerns about the move.[26] It has also gotten the attention of antitrust regulators. As noted above, the CMA investigated Google's decision to remove third-party cookies from Chrome to "assess whether the proposals could cause advertising [spending] to become even more concentrated on Google's ecosystem at the expense of its competitors."[27] In response, Google made commitments to the CMA to ensure that the "design, development, and implementation of the Privacy Sandbox proposals [would] not lead to a distortion of competition in digital advertising

---

[24] Mark MacCarthy, *Controversy Over Google's Privacy Sandbox Shows Need for an Industry Regulator*, BROOKINGS INST.: TECHTANK (June 23, 2021), https://www.brookings.edu/blog/techtank/2021/06/23/controversy-over-googles-privacy-sandbox-shows-need-for-an-industry-regulator/ [https://perma.cc/K4J8-7CZK].

[25] Allison Schiff, *Ghostery and Google: When your Destiny Depends on Another Platform's Whims*, ADEXCHANGER (Oct. 20, 2020, 12:35 AM), https://www.adexchanger.com/privacy/ghostery-and-google-when-your-destiny-depends-on-another-platforms-whims/ [https://perma.cc/EU8W-TJUB].

[26] *See, e.g., Statement From the 4A's and ANA responding to Google's announcement regarding third-party cookies*, 4A's (Jan. 17, 2020, 4:27 PM), https://www.aaaa.org/statement-from-the-4as-and-ana-responding-to-googles-announcement-regarding-third-party-cookies/ [https://perma.cc/ZH36-KPNF] ("Google's decision to block third-party cookies in Chrome could have major competitive impacts for digital businesses, consumer services, and technological innovation. It would threaten to substantially disrupt much of the infrastructure of today's internet without providing any viable alternative, and it may choke off the economic oxygen from advertising that startups and emerging companies need to survive."); Kendra Clark, *'Opportunity for Industry to Unite and Align': Marketers React to Google's Delayed Cookie Cull*, THE DRUM (June 24, 2021), https://www.thedrum.com/news/2021/06/24/google-postpones-the-death-the-cookie-until-2023 [https://perma.cc/AMH7-L7H3] (quoting the vice president of communications for search engine company DuckDuckGo in alleging that "Google's 'pro-privacy' commitment to reduce their reliance on cookies was a means to strengthen their already dominant position in the ad market").

[27] Press Release, Competition and Markets Authority, CMA to Investigate Google's 'Privacy Sandbox' Browser Changes (Jan. 8, 2021), https://www.gov.uk/government/news/cma-to-investigate-google-s-privacy-sandbox-browser-changes [https://perma.cc/XQK2-YXAS].

markets…and/or the imposition of unfair terms on Chrome's web users."[28] Dimitrios Katsifis, an international antitrust attorney who has criticized Google in the past for its potentially anticompetitive behavior,[29] claims that these commitments would "be a landmark," given Google's commitment to work closely with the CMA in the development of its Privacy Sandbox proposals.[30] However, Mark MacCarthy is skeptical that competition law on its own can provide an adequate fix for privacy harms.[31]

III. FROM THIRD-PARTY TO FIRST-PARTY

In anticipation of the loss of third-party cookie data, advertisers are building massive first-party datasets to maintain access to valuable consumer information. Google describes first-party data as "data that you own and collect with direct consent from consumers, through interactions on apps and websites, and in response to marketing initiatives, like email and loyalty programs."[32] This kind of data is not new, as it includes any information collected directly by a company from a consumer. What is new is the scale: the number of companies positioned to acquire this information and the volume of information available for capture.

"First-party data" covers a wider range of information than is apparent at first glance. Classic examples of first-party data include names, telephone numbers, and email addresses, which Geradin et al. refer to as "volunteered data."[33] This alone raises unique privacy issues, since many key components of first-party data are personally identifiable information. However, as these authors note, companies can also obtain "observed data (that is information recorded about the user and her activity, e.g., browsing history, time of log-in and log-out etc.)"[34] and "inferred data, i.e. additional information about the user, not directly provided by or observed from the user, but which is derived

---

[28] Dimitrios Katsifis, *CMA Publishes Commitments Offered by Google with Respect to its Privacy Sandbox Proposals, Seeks Comments*, THE PLATFORM L. BLOG (June 14, 2021), https://theplatformlaw.blog/2021/06/14/cma-publishes-commitments-offered-by-google-with-respect-to-its-privacy-sandbox-proposals-seeks-comments/ [https://perma.cc/V5BW-5468].

[29] *See* Geradin et al. *supra* note 3, at 64.

[30] Katsifis, *supra* note 27.

[31] *See* MacCarthy, *supra* note 23.

[32] Shannon Trainor Stark, *5 Keys to Creating Value with First-Party Data*, THINK WITH GOOGLE (Mar. 2021), https://www.thinkwithgoogle.com/future-of-marketing/digital-transformation/sustainable-first-party-data-strategy/ [https://perma.cc/GJE4-CBYZ].

[33] Geradin et al., *supra* note 3, at 14.

[34] *Id*.

from this information."[35] Google's inclusion of "interactions" in its first-party data definition likely covers observed and inferred data, revealing the depth of information collected within the first-party data category.

Despite the depth of this collection, consultants have enthusiastically endorsed the shift to first-party data as a response to consumer concerns about privacy. Deloitte and Boston Consulting Group (BCG), which both partnered with Google to research first-party data trends,[36] have published multiple reports detailing how and why companies should increase their collection of first-party data.[37] In reaction to a study finding that more than 40% of U.S. consumers do not trust online services to protect their data, Deloitte recommended that brands "[i]nvest more in first-party data. Take time to understand what customers are willing to share with you. With the demise of third-party cookies, executives should review their strategies and think about the value exchange with their customers and work to gain and improve their trust."[38]

Companies have followed this advice to an astonishing degree. Across industries, marketers are rushing to acquire data directly from their consumers. At the Consumer Electronics Show in 2019, P&G's Chief Brand Officer stated that the company had one billion consumer IDs, each presumably identifying

---

[35] *Id*. at n.36 (referencing *Online Platforms and Digital Advertising: Market Study Final Report Appendix F*, U.K. COMPETITION AND MARKETS AUTHORITY (July 1, 2020)), https://assets.publishing.service.gov.uk/media/5fe495438fa8f56af97b1e6c/Appendix_F_-_role_of_data_in_digital_advertising_v.4_WEB.pdf [**https://perma.cc/SN5A-RGY8].**

[36] *See Future-Proofing Ad Sales Growth Through First-Party Data*, DELOITTE, https://www2.deloitte.com/us/en/pages/technology-media-and-telecommunications/articles/future-proofing-ads-through-first-party-data.html [https://perma.cc/8Y9B-YSBC] ("In light of the value publishers, commerce players, and telecom can create from collecting and leveraging first-party data, Google commissioned Deloitte to investigate how they are making the most of this opportunity"); *see also* Derek Rodenhausen, Lauren Wiener, Kristi Rogers & Mary Katerman, *Consumers Want Privacy. Marketers Can Deliver.*, BOSTON CONSULTING GROUP (Jan. 21, 2022), https://www.bcg.com/publications/2022/consumers-want-data-privacy-and-marketers-can-deliver [https://perma.cc/523U-PXPS] ("To explore the perils inherent in this balancing act—and to learn how companies can adopt pro-privacy policies that create real value—BCG partnered with Google.").

[37] *See, e.g.*, Brooke Auxier, David Jarvis & Ivana Bartoletti, *The Consumer Data Privacy Paradox: Real or Not?*, DELOITTE (Jun. 29, 2021), https://www2.deloitte.com/us/en/insights/industry/technology/consumer-data-privacy-paradox.htm [https://perma.cc/EAW5-VWRT].

[38] *Id*.; *see also* Rodenhausen et al., *supra* note 35 (In their recent "Consumers Want Privacy, Marketers Can Deliver" report, Boston Consulting Group included "[a]ccelerate first-party data collection" as one of "Three actions to Win at Privacy-First Marketing.").

a single individual.[39] Unilever's CMO stated that the company had a goal to reach the same number in 2019.[40] Clorox plans to acquire information on roughly 100 million people by 2025.[41] As BCG recently put it, "[n]ow, in order to decrease reliance on third-party data and adapt to shifting consumer preferences, many marketers are looking to expand their identifiable first-party data, some by 100% year-over-year."[42] Publishers are also building massive first-party datasets to enhance their offerings to marketers. NBCUniversal's first-party data platform boasts "150 million individual deterministic consumer IDs, as well as 50 million household IDs" and plans to reach over 200 million by 2023.[43] In the growing retail media space, companies like Walmart and Walgreens are doing the same, building massive first-party datasets that allow marketers to target advertisements on the retailers' websites and across the Internet.[44]

Established brands are also feeling pressure to compete with direct-to-consumer (DTC) brands, who sell products online without retail intermediaries.[45] The now seemingly ubiquitous DTC business model is built around first-party data exchanges because the consumer must provide their information directly to the brand when making a purchase. These brands have "a treasure trove of first-party data available at their fingertips."[46] Further,

---

[39] Peter Adams, *'A World with No Ads': P&G, Unilever's Top Marketers Envision Different Paths Forward*, MARKETING DIVE (Jan. 10, 2019), https://www.marketingdive.com/news/a-world-with-no-ads-pg-unilevers-top-marketers-envision-different-pat/545733/ [https://perma.cc/WRA7-4CU9].

[40] *Id*.

[41] Alexandra Bruell, *Google's Ad Changes Prompt Big Brands to Revamp Data Strategies*, WALL ST. J. (April 1, 2021, 5:30 AM), https://www.wsj.com/articles/googles-ad-changes-prompt-big-brands-to-revamp-data-strategies-11617269400 [https://perma.cc/N9QK-HLVC].

[42] Rodenhausen et al., *supra* note 35.

[43] Max Willens, *The Rundown: NBCUniversal's First-Party Data Platform Keeps Pace*, DIGIDAY (Jan. 6, 2022), https://digiday.com/media/the-rundown-nbcuniversals-first-party-data-platform-keeps-pace/ [https://perma.cc/DX2D-ZAUX].

[44] *See About Us*, WALMART CONNECT, https://walmartconnect.com/content/wmg/home/about-us.html [https://perma.cc/85SC-W35Z] (last visited Oct. 4, 2022) (boasting that Walmart is "the nation's largest omnichannel retailer" with "a comprehensive picture of all 150 million weekly Walmart customers."); *You Know Your Brand. We know your shoppers.*, WALGREENS, https://www.walgreens.com/topic/marketing/walgreens-advertising-group.jsp [https://perma.cc/9BRB-JZY7] (last visited Oct. 4, 2022) (stating that Walgreens allows marketers to reach nearly 100 million Walgreens customers across their website, app, email database, and across platforms, such as Facebook, Pinterest, and YouTube).

[45] Jenni Baker, *How Advertisers Can Unlock the Power of First-Party Data*, THE DRUM (Nov. 15, 2021), https://www.thedrum.com/news/2021/11/15/how-advertisers-can-unlock-the-power-first-party-data [https://perma.cc/UME3-9RC2].

[46] *Id*.

direct sales generally create better margins by cutting out retail intermediaries.[47] Seeing the benefits that this model provides, legacy brands are adopting similar practices. Nike announced a "consumer direct offense" strategy in 2017 to increase direct sales.[48] In the 2021 fiscal year, Nike's direct sales accounted for roughly 39 percent of total sales, and the company expects that number to reach 60 percent by 2025.[49] Nike has even acquired three companies to assist with predictive analytics, demand sensing, and machine learning.[50] Companies are not just seeking new ways to extract information online, but are also reorienting their sales and operations to capitalize on direct data collection.

However, first-party data is not just a survival tactic. Consultants pitch first-party data solutions as an opportunity for marketers. They claim that first-party data improves a marketer's ability to understand and target consumers.[51] Incentivizing customers to "reveal" themselves in various contexts enables marketers to connect each consumer's first-party data across multiple data sources.[52] This "360-degree" view of consumers allows "hyper-personalization" in marketing.[53] The effectiveness of these practices are outside the scope of this paper, but they illuminate the way that marketers are reconceptualizing customers. Customers are creatures to be followed, tricked, and scrutinized in as many parts of their lives as possible

A. Privacy concerns ignored and created by the shift to first-party data

Although Google's plan to phase out the third-party cookie has justifiably garnered much attention for its anticompetitive effects, less attention has been paid to how this industry-reconfiguring move would impact

---

[47] Marc Bain, *The Balance in Nike's Business is Shifting Dramatically*, QUARTZ (June 28, 2021), https://qz.com/2025862/nikes-direct-to-consumer-sales-are-taking-off/ [https://perma.cc/JEK2-BET2].

[48] *Id*.

[49] *Id*.

[50] Dan Bodley, Andreas Liedtke & Pinar Tekin, *Even Big Brands Need a Direct-to-Consumer Strategy*, BOSTON CONSULTING GROUP (Nov. 16, 2021), https://www.bcg.com/publications/2021/direct-to-consumer-strategy-business-benefits [https://perma.cc/NM8C-ZU32].

[51] Shilpa Patel et al., *Responsible Marketing with First-Party Data*, BOSTON CONSULTING GROUP 3-4 (May 18, 2020), https://www.bcg.com/publications/2020/responsible-marketing-with-first-party-data [https://perma.cc/XSS8-9SV9].

[52] *Id* at 10.

[53] *Optimising First-Party Data and Personalisation in Media*, DELOITTE: ARTIFICIAL INTELLIGENCE BLOG 5 (Apr. 13, 2021), https://www2.deloitte.com/uk/en/blog/experience-analytics/2021/optimising-first-party-data-and-personalisation-in-media.html [https://perma.cc/25P9-FQG4].

consumer privacy. This Section considers some of the less-apparent privacy concerns that Google's allegedly pro-privacy move both creates and ignores.

First, as discussed above, Google's Privacy Sandbox is not a privacy panacea. Currently, Google stands poised to track users across its many online properties while cutting off cross-site tracking by other entities. To combat this, Dimitrios Katsifis has urged Google to extend the Privacy Sandbox's prohibition on cross-site tracking to its own properties.[54] Even if Google adopted this approach, first-party data collection would remain untouched. A siloed approach to Google's data collection still leaves YouTube, Google Search, and Gmail free to mine as much data directly from consumers as they please. Further, despite Google's massive influence across the web, it cannot use its sway to change the practices conducted within other web properties, including individual brands and other major surveilling platforms like Facebook. Although applying cross-site tracking limitations to Google as well as other online entities would arguably be fairer, the amount of information collected within those properties leaves many privacy concerns unaddressed.

Given that Google's move has inspired and will continue to enable a remarkable increase in first-party data collection, it is important to question the idea that first-party data better addresses online consumer privacy concerns. The idea that first-party data is friendlier to privacy concerns is often grounded in two assumptions: 1) that the user consents to the data collection, and 2) that the user is more aware of the data being collected.[55]

Consent as a justification for data collection has been widely discredited.[56] Woodrow Hartzog and Neil Richards have written about the "Control Principle," the mistaken belief that "people can adequately make choices to protect their information."[57] Much of self-managed privacy control relies on lengthy, rarely read privacy policies. As Julie Cohen notes, "[t]he issues that users must navigate to understand the significance of consent are

---

[54] Katsifis, *supra* note 27.

[55] *See, e.g.*, Alannah Sheerin & James Cooper, *Delivering on the Promise of First-Party Data*, BOSTON CONSULTING GROUP (Mar. 29, 2021) https://www.bcg.com/publications/2021/the-value-of-first-party-data [https://perma.cc/U9DQ-XSKN]; Noah Samuels & Shilpa Patel, *How to Unlock the Power of First-Party Data*, THINK WITH GOOGLE (June 2020), https://www.thinkwithgoogle.com/intl/en-cee/marketing-strategies/data-and-measurement/first-party-data-bcg-report/ [https://perma.cc/Q6U5-GQUP].

[56] *See, e.g.*, Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV 1880 (2013).

[57] Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431, 444 (2016).

too complex and the conditions surrounding consent too easy to manipulate."[58] Nevertheless, "the Control Principle is the key element of American data regulation, but it is false."[59] Despite the falsity of such control, the Principle shapes how advertisers talk about privacy and permits them to legitimize first-party data collection through the language of consent. Although users may formally consent to disclosing volunteered data, access to the digital services is often conditioned on such disclosure. Further, as Cohen notes, "[m]ost formulations of user control rights don't clearly include information derived from user behavior."[60] Given the problems of user consent, observed and inferred data should not benefit from any aura of legitimacy that consent provides. This is particularly true with inferred data since it is impossible for a user to consent to unknown inferences.

The idea that users are more aware of the data collected in first-party relationships is similarly illusory. Again, this is especially an issue with observed and inferred data. Although users may have knowingly shared their email address with a company, they may not know the company is observing all of their on-site interactions. Further, it is impossible for the user to know what information is being inferred about them. For example, Celect, Nike's predictive analytics company, helps "better predict what styles of sneakers and apparel customers want, when they want it and where they want to buy it from."[61] Although data "derived from user behavior . . . lie[s] at the core of advertising-based business models,"[62] none of that information is gathered with the user's awareness. From the perspective of the user, first-party data has an iceberg-like quality: there is an unknowable amount of data lurking beneath the surface.

Also troubling are the hidden third-party effects that exist within first-party relationships. Information provided by one user can create external effects on third parties without the third parties' knowledge or consent. Information provided by users to genetic testing services has been used to identify suspects who themselves had never provided genetic information.[63]

---

[58] Julie E. Cohen, *How (Not) to Write a Privacy Law*, KNIGHT FIRST AMEND. INST. AT COLUM. UNIV. 5 (March 23, 2021), https://knightcolumbia.org/content/how-not-to-write-a-privacy-law [perma.cc/VHT9-C93M].

[59] Richards & Hartzog, *supra* note 56, at 445.

[60] Cohen, *supra* note 57.

[61] Lauren Thomas, *Nike Acquires A.I. Platform Celect, Hoping to Better Predict Shopping Behavior*, CNBC (Aug. 6, 2019, 4:33 PM), https://www.cnbc.com/2019/08/06/nike-acquires-ai-platform-celect-hoping-to-predict-shopping-behavior.html [perma.cc/MU7R-CMZ6].

[62] Cohen, *supra* note 57.

[63] Samual A. Garner & Jiyeon Kim, *The Privacy Risks of Direct-to-Consumer Genetic Testing: A Case Study of 23andMe and Ancestry*, 96 WASH. UNIV. L. REV 1219, 1243 (2019).

Gmail monitors users' correspondence including "who you are talking to, and topics you choose to email about," [64] meaning that Google can gather information from non-Gmail-user third parties via first-party relationships. In such circumstances, seemingly exclusive data sharing relationships can have unforeseen and indirect external consequences. On the other hand, software development kits (SDKs) allow external parties to peer into seemingly exclusive data relationships. SDKs are off-the-shelf tools that allow developers to simplify app development.[65] Apps built with SDKs can then share information, such as device IP address, time of use, and advertising IDs with the SDK's original developers.[66] Google provides some of the most popular SDKs.[67] As part of its Privacy Sandbox proposals, Google announced an update called the SDK Runtime, which is intended to create "stronger safeguards and guarantees around user data collection and sharing." [68] However, the proposal currently only focuses on ad-related SDKs, and Google says that updated safeguards are "likely unsuitable for SDKs that need real-time or high throughput communications with the hosting app."[69] SDKs that track user location data, for example, appear to fall outside of the restrictions, meaning that apps using SDKs to track user location would continue to leak data. Further, Google has not yet said if the SDK Runtime would apply to its own SDKs in the same way that it applies to SDKs created by other entities, raising additional self-preferential antitrust concerns. From a privacy standpoint, it could enable Google to maintain a wide-reaching covert data collection operation that intrudes into seemingly first-party relationships between other companies and their customers.

In conclusion, Google's Privacy Sandbox restrictions leave major privacy concerns unaddressed, particularly those that create a competitive imbalance in Google's favor. The post-third-party cookie/Privacy Sandbox world would still enable, if not increase, massive amounts of first-party data collection. First-party data is not necessarily a more private form of data for any consent, knowledge, or exclusiveness reasons. Instead of allowing Google to set the rules, regulators need to reconfigure the consumer/company data relationship

---

[64] Kate O'Flaherty, *How Private is Your Gmail, and Should You Switch?*, THE GUARDIAN (May 9, 2021), https://www.theguardian.com/technology/2021/may/09/how-private-is-your-gmail-and-should-you-switch [perma.cc/D3ZF-KSRF].

[65] Charlie Warzel, *The Loophole That Turns Your Apps Into Spies*, N.Y. TIMES (Sept. 24, 2019), https://www.nytimes.com/2019/09/24/opinion/facebook-google-apps-data.html [perma.cc/JWH5-NRQD].

[66] *Id*.

[67] *Id*.

[68] *SDK Runtime*, ANDROID: DEVELOPERS (Feb. 16, 2022), https://developer.android.com/design-for-safety/ads/sdk-runtime [perma.cc/W384-MLE9].

[69] *Id*.

B. Reconceptualizing the contours of the first-party relationship

1. Antitrust

The antitrust concerns raised by regulators and scholars, discussed above, are warranted and should be addressed. Google's ability to set the rules and play the AdTech game allows them to reorient a multi-billion-dollar industry in its favor. However, antitrust law has limited reach when it comes to data collection practices. It would not be able to change the business practices of companies like Nike, NBCUniversal, and other major first-party data collectors. As Mark MacCarthy notes, "[a]ntitrust is generally powerless to mandate or forbid specific business practices unless they harm competition."[70] Further, weakening monopolies may change who sets the rules, but companies will continue to play the game with the same ruthless adaptability as before. AdTech has developed in a hydra-like fashion: each time one tactic is cut off, the industry grows a new one. This behavior requires an enduring reconfiguration of the consumer/company data relationship that effectively restricts companies' ability to collect data, regardless of whatever tactic is in vogue.

2. Trust and Information Fiduciaries

One of the foremost proposals for reconfiguring the consumer/company data relationship is the idea of the information fiduciary. Richards and Hartzog have written about the need to infuse greater trust into information relationships, saying that "modern privacy law is incomplete because from its inception it has failed to account for the importance of trust."[71] They advocate for privacy rules that increase trust, saying that "privacy rules are necessary to build the trust our digital society needs not merely to function sustainably over the long term, but also to flourish."[72] They acknowledge that this approaches fiduciary duties but advocate for a more flexible approach than other scholars, believing that not every information relationship requires the duties that a fiduciary role imposes.[73] Jack Balkin takes a bolder approach, advocating for wide adoption of an information fiduciary model. He argues for the application of common law fiduciary duties to online service providers in hopes of driving companies to "act in ways that do not harm the interests of the people whose information they collect,

[70] MacCarthy, *supra* note 23.
[71] Richards & Hartzog, *supra* note 56, at 435.
[72] *Id*. at 449.
[73] *Id*.

analyze, use, sell, and distribute."[74] The two basic duties are those of care, the idea that "[t]he fiduciary must take care to act competently and diligently so as not to harm the interests of the principal, beneficiary or client," and of loyalty, meaning "[f]iduciaries must keep their clients' interests in mind and act in their clients' interests."[75]

Although applying fiduciary duties in this context may be an appealing idea in theory, it is a poor match for concerns arising from ever-shifting data collection practices. These proposals primarily focus on data use, not collection. Although Balkin repeatedly states that the fiduciary duty would apply to online service providers across "collection, analysis, use, disclosure, and sale,"[76] the bulk of his analysis focuses on use. Balkin even concedes that the fiduciary should hold limited power over collection, saying "we should not assume that online service providers have a positive obligation to stop asking people to reveal more of themselves in social media."[77] Richard and Hartzog take a similar *ex post* view, treating disclosure as a given and asking how to best structure the relationship between data "truster" and "entrustee."[78] They discuss information disclosure as a means of generating trust: "Because disclosure of personal data leaves people vulnerable, trust is the glue that holds together virtually every information relationship."[79] The problem for first-party data collection is, however, that consumers are vulnerable to unknowable disclosures from the outset. Because first-party data includes observed and inferred data, consumers cannot know the true contours of any data relationship before entering it. For this model to curtail first-party data collection, it would have to not only cover how data entrustees use the data that they collect, but also limit what data they can collect in the first place. Further, Richards and Hartzog note that "people disclose more when they trust."[80] This begs the question of whether it is desirable for consumers to place greater trust in companies. Consumers have good reason to be skeptical of companies looking to gather their personal information. The question should not be "How should an information fiduciary behave to protect my interests?" but instead "What shape should this relationship take to protect my interests?" This may bend the idea of the information fiduciary to a breaking point because a firm would need to position itself responsibly towards a user before it establishes a relationship. Ironically, although first-party data

---

[74] Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1186 (2016).
[75] *Id*. at 1207–08.
[76] *Id*. at 1997-99.
[77] *Id*. at 1229.
[78] Richards & Hartzog, *supra* note 56, at 450–51.
[79] *Id*. at 451.
[80] *Id*. at 454.

collection is often characterized as a way to build a consumer's trust,[81] the information fiduciary is an imperfect fit.

3. Data Minimization (and Trust Minimization)

Instead of relying on a fiduciary relationship to create more responsible first-party data collection, privacy law should fully embrace principles of data minimization. Data minimization has multiple components. Article 5 of the GDPR defines data minimization as requiring that personal data be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed." [82] Asia Biega and Michèle Finck state that data minimization "requires that no more personal data than necessary to achieve the purpose is processed."[83] Seda Gürses, Carmela Troncoso, and Claudia Dias have expanded on the concept, stating that the term contains "a number of design strategies that experts apply intuitively when developing privacy preserving systems."[84] To Gürses et al., data minimization covers both "not collecting certain data inputs," but also "a number of other design strategies that make it possible to constrain the flow of data from the user-controlled domain to the domains controlled by other parties."[85] These strategies include minimizing collection, disclosure, replication, centralization, and linkability,[86] creating a system-wide approach to reducing the kind and amount of personal information collected. This approach starts with minimized data collection but also constrains downstream data flows.

Data minimization provides a more cautious perspective on data relationships than the fiduciary model. While Richards and Hartzog want data relationships to foster trust, data minimization looks to reduce the need for consumers to trust companies at all. Gürses et al. describe data minimization as a way to minimize "the need for trust," saying the practice should "whenever possible limit the need to rely on other entities to behave as expected with respect to sensitive data."[87] The authors Are careful to note that

---

[81] *See, e.g.*, Patel et al., *supra* note 50.

[82] Regulation (EU) 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), art. 5, 2016 O.J. (L 119). [hereinafter GDPR].

[83] Asia J. Biega & Michèle Finck, *Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems*, 2021 TECH. AND REGUL. 44, 55 (2021).

[84] Seda Gürses, Carmela Troncoso, and Claudia Diaz, *Engineering Privacy by Design Reloaded*, AMSTERDAM PRIV. CONF. 2 (2015), http://carmelatroncoso.com/papers/Gurses-APC15.pdf [perma.cc/7ZTQ-XRSY].

[85] *Id*. at 4.

[86] *Id*. at 5–6.

[87] *Id*. at 5.

"[m]inimizing trust is not about an emotional distrust towards any entity other than the user. Rather, it is about relying on entities to fulfill the functionality of the system, without this reliance being conditioned upon them collecting and handling large amounts of sensitive data that may later lead to privacy breaches."[88]

Notably, Gürses et al. use the term "trust" in a different way than Richards and Hartzog do. The former refer to trust as an act or an exchange that can create vulnerabilities, while the latter describe trust as an emotional bond. However, for both, "trust" is foundational, creating the contours of the data relationship. For Gürses et al., the need for trust should be whittled down to those exchanges that are necessary for the system's functioning. For Richards and Hartzog, "trust is the glue that holds together virtually every information relationship"[89] and should be cultivated because "privacy rules are necessary to build the trust our digital society needs not merely to function sustainably over the long term, but also to flourish."[90] In other words, Richards and Hartzog want better privacy rules to foster consumers trust. This is not to say that a data relationship with minimized collection (i.e., one built around minimizing Gürses et al.'s version of trust) could not produce the kind of trust that Richards and Hartzog desire. However, comparing these two conceptions of trust reveals a key distinction in how the data relationship should be considered. Gürses et al. want trust and collection only where necessary, while Richards and Hartzog want trustworthy relationships that help "make sustainable digital businesses possible."[91] Indeed, "people disclose more when they trust."[92] Given the rapid and massive changes in first-party data collection, a kind of collection often described as a way of earning consumers' trust,[93] Gürses et al. present a stronger corrective to first-party data collection overreach.

Some scholars have criticized data minimization and its closely related use-oriented regulatory partner "purpose limitation" as insufficient solutions to the privacy issues raised by collection in the era of Big Data. Regarding purpose limitation, the GDPR states that personal data must be "collected for specific, explicit, and legitimate purposes and not further processed in a manner that is incompatible with those purposes."[94] Although data minimization limits what can be collected and purpose limitation limits re-use, critics often conflate the two. In 2016, before the adoption of the GDPR,

---

[88] *Id*. at 5.
[89] Richards & Hartzog, *supra* note 56, at 451.
[90] *Id*. at 449.
[91] *Id*. at 454.
[92] *Id*.
[93] *See, e.g.*, Auxier et al., *supra* note 36.
[94] GDPR, art. 5.

Viktor Mayer-Schönberger and Yann Padova criticized the proposed regulation's reliance on data minimization and purpose limitation, noting that "with Big Data the latent value of data is unclear at the time of collection and can only be fully reaped as the data is being reused over and over again for different purposes."[95] Given the massive economic incentives behind Big Data practices, Mayer-Schönberger and Padova predicted that firms would find ways to exploit other openings in the GDPR to carry out "Big Data-esque" practices.[96] That same year, Lokke Moerel and Corein Prins similarly recognized Big Data's push to collect data for the sake of collection and claimed that "if data collection and analysis is in itself the purpose, purpose limitation is no longer meaningful and will no longer limit the types of data that can be collected."[97]

These predictions appear to have placed too little faith in data minimization. In a 2021 article titled "Reviving Purpose Limitation and Data Minimization in Data-Driven Systems," Biega and Finck argue that "the two legal principles [trust and minimization] continue to play an important role in managing the risks of personal data processing and that they may even increase the robustness of AI systems by reducing noise in the data."[98] The article recognized certain shortcomings in the approach, such as the fact that it is primarily a procedural requirement[99] and the difficulty of verifying compliance in complex and opaque data processes.[100] However, in response to Big Data's drive for massive collection, Biega and Finck note that "[e]mpirical evidence suggests that, in many data-driven settings, using increasingly larger amounts of data leads to diminishing returns in model performance."[101] Mireille Hildebrandt has spoken about the technical benefits of data minimization, stating that it "is not just a matter of data protection law, but a precondition for the methodological integrity of a robust, reliable output of machine learning."[102] Further, concerns like Moerel and Corein's about corporations rhetorically dodging the purpose requirement do not withstand

---

[95] Victor Mayer-Schönberger & Yann Padova, *Regime Change? Enabling Big Data through Europe's New Data Protection Regulation*, 27 COLUM. SCI. & TECH. L. REV. 315, 319–20 (2016).

[96] *Id*. at 329.

[97] Lokke Moerel & Corein Prins, *Privacy for the Homo Digitalis: Proposal for a New Regulatory Framework for Data Protection in the Light of Big Data and the Internet of Things* 44 (2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2784123 [perma.cc/LZ78-KL3Z].

[98] Biega & Finck, *supra* note 81, at 44.

[99] *Id*. at 55.

[100] *Id*. at 60.

[101] *Id*. at 45.

[102] Mireille Hildebrandt, *Primitives of Legal Protection in the Era of Data-Driven Platforms*, 2 GEO L. TECH. REV. 252, 268 (2018).

scrutiny: "the Article 29 Working Party considers that general statements such as 'improving user experience,' 'for commercial purposes' or 'for advertising' are generally not specific enough."[103] Biega and Finck identify steps that firms can take toward data minimization, but caveat that "the implementation of purpose limitation and data minimization in the context of data-driven systems bears a considerable research agenda."[104]

Data minimization has made limited inroads in American law. In state law, the Colorado Privacy Act[105] and the recently-passed California Privacy Rights Act[106] both contain data minimization provisions. Data minimization appears rarely in federal law. As Cohen notes, collection restrictions primarily only exist within Fourth Amendment law and some parts of national security law.[107] In a 2020 review of privacy legislation proposed in the 116th Congress, the Electronic Privacy Information Center (EPIC), a non-profit privacy research center, found that four out of eleven proposed bills included data minimization.[108]

Wider adoption of data minimization would provide helpful solutions to many of the privacy problems created by first-party data collection. Data minimization addresses the collection issue untouched by the data fiduciary model, since it would limit collection from the outset to that which is "adequate, relevant, and limited to what is necessary."[109] By limiting the need for trust and forcing companies to only collect that which is absolutely necessary, data minimization would bar collection for the sake of collection. Data minimization could be applied to all data collectors, reaching players that antitrust law could not. It would also be tactic-agnostic, limiting collection regardless of whether companies collect data through first-party relationships or some yet-to-be developed technique.

The increased procedural requirements for cookie tracking imposed by the CCPA and GDPR helped increase awareness of the tracking practice. By

---

[103] Biega & Finck, *supra* note 81, at 48.

[104] *Id*. at 60.

[105] S.B. 21-190, 2021 Leg., Reg. Sess. (Colo. 2021) (to be codified at COLO. REV. STAT. § 6-1-1308(3)) ("Duty of data minimization. A controller's collection of personal data must be adequate, relevant, and limited to what is reasonably necessary in relation to the specified purposes for which the data are processed.").

[106] CAL. CIV. CODE § 1798.100(c) (2020) (amending CAL. CIV. CODE § 1798.100(c) (2018)) ("A business' collection, use, retention, and sharing of a consumer's personal information shall be reasonably necessary and proportionate to achieve the purposes for which the personal information was collected or processed…").

[107] *See* Cohen, *supra* note 57.

[108] ELEC. PRIV. INFO. CTR., GRADING ON A CURVE: PRIVACY LEGISLATION IN THE 116TH CONGRESS (2019–2020) – UPDATED 4 (2020), https://epic.org/wp-content/uploads/2022/01/EPIC-GradingOnACurve-Apr2020.pdf [perma.cc/S2SG-BB92].

[109] GDPR, art. 5, §1(c).

drawing attention to specific tactics, legislation runs the risk of both over- and under-regulating, limiting one practice while fostering another. Rather than stymieing one specific collection tactic, data minimization would constrict all collection and change the contour of the company/consumer data relationship in a more enduring way. Companies would likely still find ways to develop new collection techniques attempting to work around minimization requirements, but the invasiveness of these techniques would be greatly curtailed.

Data minimization would also address many of the privacy vulnerabilities left open by Google's Privacy Sandbox proposals. It would limit the first-party data collection to only that which is necessary. Google would still be able to conduct cross-site monitoring, albeit less intrusively. SDK data collection would be limited to only the data necessary for the SDK's functionality. Some hidden third-party effects, like one person's DNA revealing information about another, would be addressed, since the data would have to be purpose-bound. However, Google's ability to gain information on a third-party by scanning text in Gmail exchanges may not be affected if Google can provide a purpose for such collection. Google's ability to self-preference and conduct cross-site tracking are likely better candidates for antitrust scrutiny.

The lens of data minimization reveals key opportunities for privacy-enhancing interventions across the three kinds of data accumulated in first-party relationships. Volunteered and observed data would need to be limited to the minimum amount necessary for the intended purpose. Inferred data would be greatly, if not entirely, restricted, since it is information derived from volunteered and observed data, and thus created by reprocessing volunteered and observed data for a new purpose. Such reprocessing would violate the principle that data be "limited to what is necessary in relation to the purposes for which they are processed."[110] Reprocessing data for a new purpose would fall outside of the initial requirement that the collected data be purpose-bound. Companies could provide a separate purpose for this collection, but it would need to be more specific than "to improve user experience," as indicated by the Article 29 Working Party. Of course, any version of data minimization would require high standards of specificity and robust enforcement in order to truly address collection overreach. However, effective incorporation and enforcement of these principles could significantly reshape first-party data relationships by forbidding corporations from collecting for the sake of collection.

---

[110] *Id.*

IV. CONCLUSION

As marketers and publishers race to build out first-party datasets in reaction to Google's allegedly pro-privacy move and under the guise of their own pro-privacy principles, new versions of intrusive data collection will proliferate. Google's dominant position in the AdTech ecosystem allows them to rewrite the rules that other corporations follow. This dominance has, justifiably, drawn great attention from antitrust law. Less examined are the potential privacy harms caused by Google's allegedly pro-privacy move. This Note has aimed to surface those harms and evaluate the best way to protect against them in a first-party data world. Since companies large and small engage in increased first-party data collection, reworking the corporation/consumer relationship to protect the consumer's privacy from the outset is critical. Data minimization is the best way to reshape and restrict corporate data collection practices but would require vigorous monitoring and enforcement to meaningfully change the corporation/consumer relationship.