# MARKET DESIGN FOR PERSONAL DATA

By

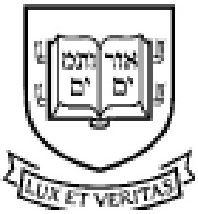
Dirk Bergemann, Jacques Crémer, David Dinielli,

Carl-Christian Groh, Paul Heidhues, Maximilian Schäfer,

Monika Schnitzer, Fiona M. Scott Morton, Katja Seim,

and Michael Sullivan

COWLES FOUNDATION PAPER NO. 1844




COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2023

http://cowles.yale.edu/

# Market Design for Personal Data

**Authors:**[†]

Dirk Bergemann, Yale University
Jacques Crémer, Toulouse School of Economics
David Dinielli, Yale University
Carl-Christian Groh, University of Bonn
Paul Heidhues, DICE, Heinrich-Heine University Düsseldorf
Maximilian Schäfer, Institut Mines-Télécom, Business School and Yale University
Monika Schnitzer, Ludwig-Maximilians-University Munich
Fiona M. Scott Morton, Yale University
Katja Seim, Yale University
Michael Sullivan, Harvard University

[†] The authors are a collection of economists and policy experts in the United States, the United Kingdom, and the European Union who have studied, and are committed to the improvement of, competition in digital markets.

Many thanks to Brian O'Kelley and Alissa Cooper for helpful discussions that contributed to the Article.

Many thanks also to Rosella Argenziano for her precise and trenchant comments on early drafts, and for sharing her substantial knowledge of the relevant legal and economic literature.

Authors' full titles and conflict disclosures can be found in *Disclosures Regarding Authors' Conflicts of Interest*, 40 YALE J. ON REGUL. 1121 (2023) [https://perma.cc/PD3D-PUQY]. Omidyar Network, the James S. and James L. Knight Foundation, and the Alfred P. Sloan Foundation have provided funding and other support for this Article and other articles relating to regulation of digital platforms.

## I. Introduction

### *A. The Problem*

It is now generally understood that personal data—that is, data that relate to individual consumers—drive digital markets. Personal data underlie targeted advertising, which draws billions of dollars into ad-supported markets. Personal data are useful for other purposes as well. Firms in digital markets rely on personal data to deliver their core products and services—we refer to these collectively as "web services"[1]—to hone and improve them, and to recommend related products and services. These data facilitate innovation, allowing yet more services and "smart" products with increasingly personalized functionalities. Personal data can allow governments to deliver better public services, such as transportation systems, or can help researchers better understand how humans interact with algorithms and which policies might best serve society. And data can also facilitate competition, by improving quality and providing insight into consumer conduct that encourages entry. In these various ways, the massive quantity of personal data currently collected undoubtedly contributes to consumer welfare.

But there also are downsides to the collection and use of personal data on such a grand scale. "Surveillance capitalism," as Professor Shoshana Zuboff has termed it,[2] has blurred the line between the personal and the public, and has commodified our habits, interests, and beliefs in ways that can feel distasteful and invasive. Massive data collection also has made information about us more accessible to government and commercial actors who often face little to no accountability for its misuse.

Many of the reactions and proposed responses to the current situation examine these concerns through the lens of privacy. Economists, however, look at this same set of facts—massive data collection from users of web services, a stranglehold over data by a handful of large firms facing weak competition, and monetization largely through the sale of targeted advertising—and see an *additional* set of problems.

- Personal data fuel digital markets, but the users, whose unique set of characteristics, actions, and experiences give rise to the data, receive no cash compensation for the personal data they generate. Users generate a resource of tremendous value—personal information—and yet firms extract this resource without payment (other than the provision of digital goods and services in exchange). This exchange stands in sharp contrast

1. We use the term *web service* to refer to all online services—whether they are websites or apps on a mobile device—with which web users interact or that seek to use web users' data for service provision, ad targeting, conducting analytics, product improvement, or any other reason.

2. *See generally* SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019).

to what we see in other markets, in which those who control resources are paid for their extraction or use.[3]

- We see a handful of firms that have significant market power controlling vast swathes of personal data. One way this market power manifests itself is in lower quality services, including the collection of personal data without effective user control, and the use of that data to extract a surplus from consumers.

- Personal data now are collected in a huge variety of settings, and yet there is no basis to believe these data are put to their highest use. The private firms that control data have no incentive or mechanism to share data for valuable research or public benefit (transportation planning or prevention of technology addiction, for example), or with other private firms that could use the data to offer better services to consumers.

We would be less concerned with the fact that users are paid in barter rather than in cash for the extraction and use of their personal information if it appeared that the trade were a fair one. But the evidence strongly indicates that it is not. Data-driven markets in recent years have consistently generated tens of billions of dollars in annual profits for the largest digital platforms. These profits are significantly larger than would be expected in competitive markets and suggest the exercise of market or even monopoly power.[4]

We also would be less concerned if there were evidence that the data are being used for purposes other than simply advancing each firm's independent financial interest. We are aware of no evidence that this happens on a large scale, however. And in a classic example of the exception proving the rule, the few instances in which large firms *have* allowed data generated through use of their products to be used in the public interest, seem to have failed or backfired *specifically because* web users distrust the large firms and suspect they will use

3. For a description of externalities arising from the social use of personal data, see generally Dirk Bergemann, Alessandro Bonatti & Tan Gan, *The Economics of Social Data*, 53 RAND J. ECON. 263 (2022).

4. For a more detailed discussion of the economic forces driving market power in digital markets, see Fiona M. Scott Morton & David C. Dinielli, *Roadmap for a Digital Advertising Monopolization Case Against Google*, OMIDYAR NETWORK (May 2020), https://omidyar.com/wp-content/uploads/2020/09/Roadmap-for-a-Case-Against-Google.pdf [https://perma.cc/PY6H-L4F9]; and Fiona M. Scott Morton & David C. Dinielli, *Roadmap for an Antitrust Case Against Facebook*, 27 STAN. J.L. BUS. & FIN. 267 (2022).

Government enforcers in various jurisdictions around the globe have accused Google, Facebook, Apple, and Amazon of monopolistic practices. European antitrust proceedings against Google resulted in a record €4.3 billion fine. *See* Summary of Commission Decision of 18 July 2018 Relating to a Proceeding Under Article 102 of the Treaty on the Functioning of the European Union and Article 54 of the EEA Agreement (Case AT.40099 – Google Android), 2019 O.J. (C 402) 19. The Digital Markets Act and Digital Services Act are explicitly targeted towards dominant firms defined as digital gatekeepers. *See* Digital Markets Act, Council Regulation 2022/1925, 2022 O.J. (L 265) 1, 2 (focusing on the "small number of large undertakings providing core platform services"); Digital Services Act, Council Regulation 2022/2065, 2022 O.J. (L 277) 1, 11 (providing for additional obligations for "very large online platforms").

the data to benefit themselves rather than to advance the users' interests. Covid exposure tracking apps, which faced significant headwinds in the United States especially, are a key example.[5] Also, new entrants and competitors cannot access the biggest trove of data even if they would use them in ways that are socially beneficial.

Natural forces in a well-functioning market should correct for the significant economic profits enjoyed by the largest platforms, quality or control rights below competitive levels, and the inefficiencies that keep data out of the hands of those who would put it to good use. Market forces would require large platforms and other firms that rely on personal data to share some portion of the billions in annual surplus with web users.[6] The firms might do this through a combination of cash payments in exchange for the right to use personal data and/or additional product improvements that lower quality-adjusted prices. If the market were functioning as it should, we also would expect the platforms that facilitate digital advertising to share the surplus with publishers that supply digital ad space (through higher pass-through rates of total ad spend) and with advertisers (through lower ad prices). Firms that currently hoard data that could benefit other suppliers of services, public or private, would have incentives to share data at reasonable prices.

Rather than a competitive data market, what we see is a market failure. The status quo regarding personal data collection and use presents concerns about competition, efficiency, innovation, and the distribution of the surplus generated in digital markets, *in addition* to the various privacy concerns that others have identified. Our proposal offers a more comprehensive set of potential solutions than do other proposals we've examined that address the collection and use of personal data. We attempt to solve for *three principal market failures* with respect to personal data within one policy framework:

1) the failure to provide users the ability to control how their personal data are collected and used—which contributes to a status quo that threatens user privacy, lowers the effective quality of online services, and facilitates market power;

2) the failure to provide users a way to derive financial benefit from the data they generate—which enforces a status quo that distributes

5. *See, e.g.*, Jessica Rich, *How Our Outdated Privacy Laws Doomed Contact Tracing Apps*, BROOKINGS INST. (Jan. 28, 2021), https://www.brookings.edu/blog/techtank/2021/01/28/how-our-outdated-privacy-laws-doomed-contact-tracing-apps [https://perma.cc/QPQ9-GEYJ].

6. Digital advertising is not the only setting in which firms with access to substantial amounts of data can use the data to divert surplus to themselves and away from consumers. For example, in the context of third-degree price discrimination (i.e., charging different prices to different categories of consumers), access to large amounts of personal data can help a firm segment consumers into groups whose members are charged close to the maximum they would pay. Personal data help the firm to identify the appropriate amount to charge each group. *See generally* Dirk Bergemann, Ben Brooks & Stephen Morris, *The Limits of Price Discrimination,*105 AM. ECON. REV. 921 (2015).

surplus from digital markets to platforms rather than consumers and facilitates market power; and

3) the failure to ensure that the data that are collected can be put to their highest use, including by firms other than the big digital platforms, as well as the nonprofit sector and governments—which generates a status quo that implicates or even impedes innovation.

Moreover, these three problems appear to be related. The lack of effective privacy regulation or other restrictions on data collection allows large platforms to collect and use data nearly unfettered, giving them *higher monetization rates*; these advantages promote and protect market power directly. Because platform market power derives so directly from the platforms' data advantages, the platforms have a strong incentive to prevent others from accessing or benefiting from the data they perceive to be "theirs." The platforms' exclusionary approach puts their data out of reach of rivals who might use the data to train their own algorithms, design competing products, or prepare for seamless interoperability, for example. The exclusionary approach also frustrates legitimate requests from the government or from academic researchers,[7] allowing platforms to forestall or delay the sort of complete understanding of their business practices necessary for effective regulation. Forestalling regulation, in turn, preserves market power. That market power, in turn, insulates the platforms from the constraints on data collection and use that vigorous competition would impose.

### *B. A Proposed Solution*

This Article explores a possible intervention that—unlike antitrust enforcement alone or enhanced privacy regulation alone—would address competition, efficiency, *and* privacy concerns, directly and simultaneously.[8] We

7. Large platforms sometimes claim the data themselves constitute trade secrets and cannot be disclosed, or they assert that disclosure would undermine user privacy. *See A Preliminary Opinion on Data Protection and Scientific Research*, EURO. DATA PROT. SUPERVISOR 9 (Jan. 6, 2020) (discussing "[c]orporate secrecy as a barrier to research"); Jef Ausloos, Paddy Leerssen & Pim ten Thije, *Operationalizing Research Access in Platform Governance*, ALGORITHM WATCH 25 (June 25, 2020).

8. In this way, our proposal may serve as a counterexample to the popular notion that efforts to enhance competition in digital markets necessarily undermine privacy interests and efforts to protect privacy necessarily undermine competition interests. We perceive no unavoidable tension between competition and privacy. The notion seems to reflect misconceptions about interoperability, a tool that these authors and others have proposed to encourage entry and facilitate consumer choice in markets including the social network market. *See Online Platforms and Digital Advertising*, COMPETITION & MKTS. AUTH. 34 (July 1, 2020), https://assets.publishing.service.gov.uk/media/5efc57ed3a6f4023d242ed56/Final_report_1_July_2020_.pdf [https://perma.cc/64LE-D9LQ]; Fiona M. Scott Morton, Gregory S. Crawfor, Jacques Crémer, David C. Dinielli, Amelia Fletcher, Paul Heidhues, Monika Schnitzer & Katja Seim, *Equitable Interoperability: the "Super Tool" of Digital Platform Governance*, 40 YALE J. ON REGUL. 1013 (2023). Interoperability certainly could make a user's personal data accessible to additional firms, but there is no reason to think personal data is more safe and secure with a large platform than with a smaller firm that must be licensed to interoperate with it. Further, interoperability need not allow the interoperating firm to do whatever it wants with personal data to which it gains access. The interoperating

offer the idea of a "data intermediary," bound by fiduciary duties to users, empowered to monetize users' personal data—a category we define above and delineate further below—and permit other uses in accordance with user instructions. This proposal is similar to other proposals that rely on some form of intermediary that sits between consumers and firms that wish to use their personal information. Prior proposals generally would empower the intermediary to prevent misuse of personal data or increase user control, and some might facilitate innovative uses of data, but they are not designed to monetize the data on the users' behalf.[9]

Our proposal would encourage the development of a market for personal data in which users who generate personal data are assumed to *control* their data in the first instance. In the United States, laws creating the right for consumers to control their personal data would need to be adopted, or courts would need to acknowledge that existing statutory schemes or common law principles already provide such rights.[10] In Europe, the General Data Protection Act (GDPR), already confirms that "data subjects"—the individuals we call users—have a right to restrict the "processing" their data.[11] For that reason, "processing" of personal information is lawful under GDPR *only* when certain conditions are met, the most important being consent of the data subject to processing for a specific purpose.[12] The presumption we adopt here—that users control the data they generate—would imply a similar corollary: that a web service may use personal data, including data that would not exist but for the user's interactions

firm only gets to do what the large platform gets to do; there are no new or additional data uses that could raise "privacy" concerns. A recent theoretical study highlights the potential consumer welfare benefits of *data linkages*—i.e., data-sharing relationships between firms. Rossella Argenziano & Alessandro Bonatti, *Data Linkages and Privacy Regulation*, 13-18 (Mar. 6, 2021), https://www.mit.edu/~bonatti/protection.pdf [https://perma.cc/H5LK-QTM5]. Thus, interoperability requirements could lead to welfare benefits.

9. A notable exception is that offered by Eric Posner and E. Glen Weyl, who posit that the large platforms exercise monopsony power over users, whose personal data has marginal value and who therefore should be compensated for it. *See* ERIC A. POSNER & E. GLEN WEYL, RADICAL MARKETS: UPROOTING CAPITALISM AND DEMOCRACY FOR A JUST SOCIETY 177-194 (2018).

10. Creative lawyers, however, are beginning to pursue lawsuits that include statutory and common-law claims that presume users have various rights in the data they generate. Notable among these is a proposed class action filed in a California federal court against OpenAI (the developer of DALL-E and ChatGPT) and Microsoft for collecting data about plaintiffs and using the data to train its AI products without permission from the plaintiffs and without compensating them. *See* Class Action Complaint, P.M. v. OpenAI LP, No. 23-cv-3199 (N.D. Cal. June 28, 2023). The data at issue include data users generate through their interactions with web services such as location information, keystrokes, search queries, image data, health information, etc.—all of which constitute what we in this Article term "personal data." Plaintiffs allege that defendants collect the relevant data via apps such as Spotify and MyChart that incorporate OpenAI's products. *See id.* ¶ 16. Based on these alleged facts, plaintiffs assert several causes of action that lie only if plaintiffs have a proprietary or property interest in the personal data they generate through their interactions with web services. The causes of action include "larceny/receipt of stolen property" (*id.* ¶¶ 575-92), "conversion" (*id.* ¶¶ 593-97), and "unjust enrichment" (*id.* ¶¶ 598-606). Plainly, this case (and the fate of these three claims in particular) merits monitoring. So too do the small number of cases in which courts already have acknowledged users' proprietary or property interests in the data they generate. *See id.* ¶ 576 (collecting cases).

11. *See* Council Regulation 2016/679, art. 4(2), 2016 O.J. (L 119) 1, 33 ("'[P]rocessing' means any operation or set of operations which is performed on personal data or on sets of personal data . . . such as . . . use . . . .").

12. *See id.* art. 6(1)(a).

with the web service, *only if* authorized by specific statutory or regulatory permission. In all other instances, the legal regime we propose would require that web services buy the right to use personal data from a data intermediary acting on users' behalf.

1. Our Proposal Governs the Collection and Use of All Personal Data

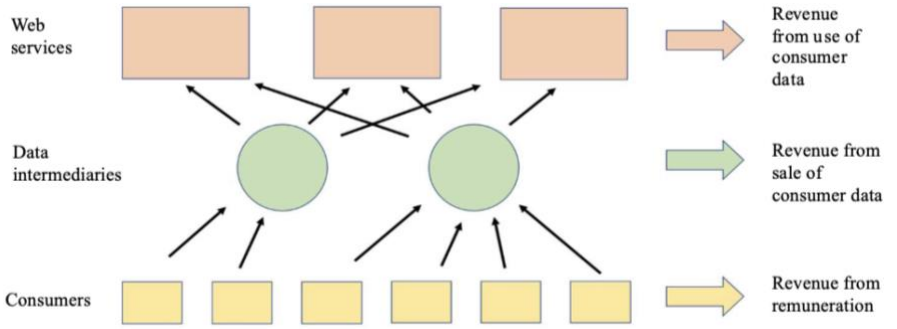
We propose a market for all personal data, without regard to whether the law deems the data or the information the data reveal to be "private" in a formal sense. Our proposal therefore offers something additional and complementary to proposals that focus narrowly on enhancing protections only for private information. We take this broader approach with the aim of capturing the full set of data extraction and use practices that contribute to the market failures described above.

The narrower approach would not offer a complete solution because many extraction and use practices that cause consternation and/or generate firm profit (and should therefore give users a right to compensation and to control) rely on personal data that, depending on circumstances and jurisdiction, might not be considered private. Private data can become nonprivate through disclosure, for example, and yet disclosure might not justify depriving the user permanently of any ability to control, or seek compensation for, the use of the data. (Consider someone who informs friends and family of a cancer diagnosis via a public Facebook post.) And some nonprivate data can be used in ways that feel intrusive or exploitative and therefore ought to support a right to compensation and control. (Consider someone who has parked her car in a spot reserved for patients at an abortion clinic and who is served "ads" produced by anti-abortion advocacy groups.) More generally, mobile users can be tracked across the web using seemingly innocuous, public data about their phone settings that is provided automatically when they access web services, such as keyboard layout and operating system version.[13] It is clear, therefore, that expansive collection and use of various forms of nonprivate data underlie some of the most urgent concerns with data-driven products and markets. These data can be highly valuable and should not be excluded from the market for personal data we propose herein.

We offer the diagram below to demonstrate the fundamental architecture of our proposal, which shows how it would govern the collection and use of *all* personal data:

13. *See* Matt Burgess, *The Quiet Way Advertisers Are Tracking Your Browsing*, WIRED (Feb. 26, 2022), https://www.wired.com/story/browser-fingerprinting-tracking-explained [https://perma.cc/8S9Q-L4PL] (describing how device fingerprinting based on basic phone settings can be combined with sensitive personal data).

Figure 1. Schematic Identifying Principal Categories of Participants in Data Intermediary Markets




2. We Propose a Control Right that Includes a Right To Be Paid for Data Use

This is why our proposal includes a broad *control* right that is more akin to the right to restrict processing of all personal data, rather than relying principally on any effort to enhance privacy *per se*. The control right envisioned would encompass the ability to decide what sort of data could be collected *and* the purposes for which the data could be used. The control right also would imply a right to be paid. Because users would not be required as a general matter to permit any collection or use of personal data, users could demand payment from web services who want to use it.

Recognizing a control right with these two principal features—(1) the ability to limit collection and use of personal data; and (2) the ability to bargain for payment—would constitute a significant change in the legal and economic landscape in all jurisdictions of which we are aware. Europe's GDPR, for example, acknowledges that individuals have a continuing right to exercise some control over the use of data that relate to them. But the GDPR does not envision that firms should pay to use such data or provide a mechanism for any such payments. Recognizing the two-part control right would be even more transformational in the United States, where individuals have few rights with respect to their personal data other than the right to prevent disclosure of that which is deemed private and to seek damages if the disclosure of private data causes harm.

3. Designing a Market for Personal Data Is Complex, Perhaps Prohibitively So

Although it is straightforward to explain why establishing this new right with respect to personal data *should* have the beneficial consequences identified above, it is a complex undertaking to design a market that will facilitate that

outcome. The bulk of the Article below traces through the economic issues that would arise in creating such a system.

Creating and maintaining the institutions, procedures, and oversight necessary to permit users to exercise the newly proposed control right in a manner that would yield the desired competition and privacy benefits is complex and expensive, and some might worry whether the benefits justify the effort. Other proposals addressing the use of personal information could achieve some similar results, with less effort. A tax on digital advertising, for example, could redirect platform profits to public uses that benefit users indirectly, such as schools, public green spaces, or internet access subsidies. A ban on digital advertising that relies on personal data for targeting could significantly reduce privacy concerns relating to data collection and use. It would limit advertisers to placing ads based exclusively on context—a running shoe company might pay to have its ads appear alongside an article about the New York Marathon, whereas a high-end women's shoemaker might pay to have its ads appear alongside an article about Milan Fashion Week.

In this way, such a ban could reduce the competitive advantages that large platforms enjoy due to their access to large and rich data sets. But such a ban might also decrease welfare in that consumer "search costs"—the time and effort required to find products and services that match the consumer's need and ability to pay—would increase.

Targeted advertising's effect on search costs provides one example of the obvious fact that *some* data collection, *some* data uses, and *some* targeted advertising benefit consumers. Too much, or the wrong kind, of any of these activities may harm consumers. Our policy proposals reflect our goal of creating a market for personal data that is sufficiently efficient such that competition compels firms to collect, use, and share data, including for ad targeting, in amounts and ways that increase total welfare. None of the authors of this Article can guarantee this result. We nonetheless maintain that the thought experiment we engage in here is decidedly worth the effort, if for no reason other than to understand exactly how hard it might be to generate a market for personal data.

The remainder of this Article proposes minimum policies we consider to be necessary to allow a market for personal data to develop and exist over time and to operate in a manner that permits users to exercise both elements of the new control right described above: (1) the ability to limit collection and use of data; and (2) and the ability to be paid in a manner that solves for the three interrelated market failures highlighted above. We do not purport to offer a blueprint. Nor should our proposals be read by any government agency or official as a set of instructions on how to create a perfectly functioning market for personal data. Rather, we have applied our knowledge of economic theory, behavioral economics, strategic behavior of firms, and the current operation of data markets to identify critical features of such a market. An immediate conclusion is that a successful data market will not function without affirmative policy interventions by a regulator.

We noted above that in the market we envision, users' control right confers the ability to demand payment in exchange for the right to use their personal data. But it is apparent that no individual could be expected to negotiate for payment every time a web service uses their personal data, or even to negotiate payment schedules with web services that might use their data repeatedly or consistently. Nor could we expect web services to contract for each use of personal data, or even enter long-term or omnibus agreements with each individual whose data it might use.[14] Nor could researchers or other market participants who would study the data be able to obtain individual consent from the thousands or millions of individuals to whom data might relate. Individuals and firms would be overwhelmed, commerce would grind to a halt, and every human with an internet connection would tear their hair out.

4. Data Monetization Requires New Entities: Intermediaries Guided by User Instructions and Fiduciary Duties

Our central policy proposal provides a potential solution to this problem: regulations should establish a new kind of entity called a data intermediary. We propose that data intermediaries serve as the users' exclusive agent for permitting use of data consistent with user instructions, as well as the users' exclusive agent for purposes of monetizing that data and remitting a portion of the money received as payment for use of the data to the users. Each data intermediary would act as a one-stop shop for its users, who would exercise their control right through that intermediary. Because of this feature, users need not have any direct contact with web services about the services' use of their personal data. Each data intermediary also would serve as a one-stop shop for web services with respect to the personal data of the intermediary's set of customers. And the intermediary would serve as a similar one-stop shop for web services, researchers, and others who have no direct link to users but who seek access to personal data for market or product or other forms of research.

The design of the intermediary suggested in our proposal aims at providing users the highest value possible for the use of their personal data. Because ours is a market solution, we want the value of the payments to be determined through competition among the intermediaries. It therefore is crucial that the market design enhances competition. We discuss the way a regulator could enable consumers to choose intermediaries offering the highest payment and best service and switch easily in response to both service and payments.

14. Individual negotiations of this sort would be unlikely to shift significant surplus from the large platforms to users in any event. The marginal value to the platform of an individual user's personal data is small. If a platform can use the personal information of 100 million users to sell targeted advertising, for example, adding one more person to the group of potential targets does not change the price it can charge for advertising. This is the case even if average advertising spend per user is large, $500 for example. The individual user may want a portion of that $500, but the platform has no incentive to pay them anything close to that amount, or indeed anything at all.

Our proposal also addresses the danger that web services exploit consumers' behavioral biases to encourage them to share more data than they would if choice were more transparent and understandable. Real-world consumers require careful design of the choice architecture surrounding data sharing to protect them from poor choices and exploitation. Our proposal envisions a set of standard data sharing tiers from which a web user can select a level that most closely reflects their degree of comfort or discomfort with the collection and use of their data. The "sharing tier" determines what kind of personal data the intermediary can monetize on the user's behalf, and on behalf of all other users who have selected that sharing tier. We share the concern of many that a focus on remuneration would steer web users toward excessive sharing of their data despite the risks of data sharing, which are less salient than monetary rewards, and despite the possible social or societal harms. We take this possibility seriously, and we consequently develop our proposal to mitigate risks from sharing data.

The surplus at stake is large. The profits generated by the data-driven businesses of tech companies suggest that the economic value derived from consumer data is substantial. Consumers are likely not the only parties to benefit financially from a competitive data market. A competitive market for data would allow smaller entrants and innovators to enter and compete for the large revenues this sector generates. Today in the United States, advertisers spend hundreds if not thousands of dollars per year per person on digital advertising that uses personal data for targeting.[15] An important task for economists is to develop mechanisms to return control and a portion of that value to households so that in future years all consumers will share in the thousands of dollars in value they generate.

### 5. Ours Is One Idea Among Many; It May Not Be the Best

We are not the first to consider regulatory solutions to the problems of digital markets as enumerated above. Academics and think tanks around the world have put forward ideas for possible solutions. The motivation of almost all of them is to empower users to share and control personal data. Relatively few are focused on the economics of data markets—the efficient selling of information—and competitive remuneration for consumers. But because of the significant sums at stake, and the ability of competitive data payments to reduce deadweight loss and redistribute income to consumers, economic solutions could be very valuable.

The establishment of data intermediaries has also been suggested by several distinct groups, including the European Union, the UK Centre for Data Ethics and Innovation, and RadicalxChange. Moreover, there are private initiatives such as the web browser Brave and the startup Solid that aim to endow consumers

15. *See infra* app. 2.

with greater control over their data within the existing regulatory framework.[16] Brave is a web browser designed to minimize data collection at its source, by blocking all trackers and preventing data storage by first- and third-party cookies, thereby reducing the amount of data collected in the first instance.[17]

It is worth pausing to consider Solid as a particularly creative approach to solving many of the same problems we try to address with intermediaries.[18] Solid is a specification aimed at giving individual users control over collection and use by empowering individuals to store personal data in a virtual pod. Users choose what to put in and what to let out (and to whom and for what purpose), thereby decentralizing the web by providing its users ownership and control over their data.[19]

Tim Burners-Lee, the inventor of the World Wide Web, created Solid in 2016.[20] Solid provides its users with data ownership and privacy by, first, providing personal data storage units called "Pods" and assigning each Solid user a unique identifier.[21] The sort of data, including encrypted data, that a user can store in a Pod includes information about the user's preferences and data related to the user's behavior on the web. Solid is not solely a platform for storing encrypted data, however; Solid also facilitates sharing data with third parties in a controlled manner. Users can dictate how they share their data with third parties using Solid's settings. Solid is an open standard and a platform that changes certain features of the Web by changing how web services access consumer data.[22]

Our data intermediaries are similar to Solid in that both consolidate and process consumer data. Unlike our intermediaries, which would act as fiduciaries for consumers in managing and sharing their data, Solid provides consumers with direct control over their data. While both our proposal and Solid aim to protect consumer privacy by providing consumers control over their data, only our proposal leverages the value of consumer data to web services to benefit consumers in the form of cash and other forms of payment in addition to the simple barter of the service in exchange for data and attention.

16. *See Opinion 9/2016: Opinion of the European Data Protection Supervisor on Personal Information Management Systems*, EUR. DATA PROT. SUPERVISOR (Oct. 20, 2016) https://edps.europa.eu/sites/default/files/publication/16-10-20_pims_opinion_en.pdf [https://perma.cc/EQM7-8LUP] (European Union approach); *EDPS TechDispatch: Personal Information Management Systems*, EUR. DATA PROT. SUPERVISOR (2020), https://data.europa.eu/doi/10.2804/11274 [https://perma.cc/5QCX-KEJH] (same); *Unlocking the value of data: Exploring the role of data intermediaries*, UK DEP'T FOR DIGIT., CULTURE, MEDIA AND SPORTS (July 22, 2021), https://www.gov.uk/government/publications/unlocking-the-value-of-data-exploring-the-role-of-data-intermediaries/unlocking-the-value-of-data-exploring-the-role-of-data-intermediaries [https://perma.cc/YH6U-7BD6] (United Kingdom approach); *The Data Freedom Act*, RADICALXCHANGE (Feb. 18, 2020), https://www.radicalxchange.org/media/papers/data-freedom-act.pdf [https://perma.cc/7PR8-DVSA] (RadicalxChange proposal).

17. *See Advanced Privacy*, BRAVE, https://brave.com/privacy-features [https://perma.cc/XE52-C82T].

18. *About Solid*, SOLID, solidproject.org/about [https://perma.cc/9WST-Z5D7].

19. *Id.*

20. *Origin*, SOLID, https://solidproject.org/origin [https://perma.cc/63LQ-G3ZD].

21. SOLID, *supra* note 18.

22. *Id.*

Further, Solid does not appear to solve the collective action problems that currently stand in the way of consumers' ability to demand remuneration for their personal data. Remember that our proposed data intermediaries would manage many users' data—millions of users, in fact. This would allow them to bargain on their users' behalf with web services in a way that a single Solid user simply could not. Web services will not value a single Solid user's data that highly. They would, however, value the massive amounts of data that a data intermediary stands to offer. Intermediaries' relatively significant bargaining power as compared to the minimal power called on by any individual Solid user means that our proposal would result in compensation to consumers who choose to share their data, which owes to the fact that data intermediaries would be able to bargain with web services. Solid does not currently feature any scheme allowing consumers to be remunerated for their data.

Solid does share with intermediaries the potential benefit, if widely adopted, of putting pressure on app and platform developers that could result in increased innovation and creativity. This is because, absent uninhibited and aggressive data collection, web services would need to compensate users for data or improve their infrastructures to retain consumers.

We are not optimistic Solid ever will be widely adopted, however. Some advanced web users may desire to personally manage their data in the ways Solid makes possible, but the historical reluctance of web users to fine tune their privacy settings online suggests that most web users would prefer to simply set a general privacy level and then allow specialized intermediaries to handle data management on their behalf subject to the selected privacy level's constraints.

Shifting topics, we note that our proposal also adds to the budding literature on data intermediaries by analyzing the implications of economic theory for an advantageous design of intermediaries.

We note that setting up working data markets is difficult and policymakers may determine it is not worth the effort. The difficulties inherent in implementing our policy recommendations are multiplied by the need for authorities to coordinate establishing and then regulating the market across jurisdictions.[23]

23. The difficulties also are multiplied by the fact that our proposal assumes that all users are legally and functionally competent to participate in the market, even though that clearly is not the case. Minors are an important example. In our view, firms ought to pay *all* users for the use of their personal data, including minors—a key demographic targeted by advertisers. Our proposal, however, does not address personal data of minors, which web services collect, use, and monetize much in the same way as they do the personal data of adults. Determining the age at which minors should be presumed sufficiently mature to make decisions about their personal data is beyond the scope of this Article. So too are the laws governing who can act on behalf of minors and under what circumstances. Such questions are important; minors seem to us especially vulnerable to exploitation in this market. The potential cash payouts may seem especially large to minors, causing them to undervalue their own privacy and related interests, or to over-discount the dangers the data collection and use could cause them or others. We also can envision various practical difficulties in allowing minors to participate in the market we propose, including the fact that many minors presumably are unbanked (raising the question whether intermediaries should pay such minors annually for the use of their data, or rather place the money in individual trusts). Those who do have access to accounts may share control with parents or guardians whose interests are not aligned with those of the minor. And finally, we know that age verification, a seemingly necessary first step in

Even if a perfect regulator followed all the suggestions in this Article, we cannot be certain that it is possible for a market in personal data to flourish. However, all authors feel strongly that the status quo—simply ceding all the value of personalized digital advertising to a handful of big firms, allowing those firms to control the use of the data, and accepting the inefficiencies of current markets that impede high-value use of data by third parties—is not acceptable. That profit is generated by the information and activities of consumers who, for both efficiency and fairness reasons, should share in it.

Our proposal is grounded in economic theory and evidence, which we highlight in the discussion. There is still much that is unknown about the economics of data markets, which necessarily creates uncertainty and limits the specificity of our proposals. Further research and experience across different settings and jurisdictions will help to solidify our understanding in these areas. Partly for this reason, not all authors are equally enthusiastic about all the ideas in this document, but all authors agree that the proposal provides a useful starting point for debate and discussion on the future of digital markets. More importantly, each author thinks that the endeavor of exploring ways to compensate internet users for the collection and use of their data is of utmost importance from the standpoint of efficiency and fairness, in addition to concerns about competition and privacy.

Throughout the discussion, we refrain from making prescriptions about technical details required for implementation such as where data are stored or how they travel from one place to another. Our economic analysis does not depend on these choices. More importantly, if a system similar to our data-intermediation regime were to be adopted, this system should use the most efficient and appropriate technology available at that time for fulfilling the duties that our proposal assigns to various participants in the digital economy. Instead, we set out the legal and economic principles that should govern intermediaries; the technology used to implement our proposal should enable and respect these principles.

The Article proceeds as follows. We first list our policy recommendations for the reader who wants a one-page overview. Then we introduce the basic characteristics of our proposed data intermediaries and the regulatory regime in which we propose that they operate. Next, we turn to a description of personal data and privacy levels. The next Part focuses on user behavior, followed by a detailed discussion of how the purpose of data usage fits into the regulation. We then discuss the ways in which the regulator can enhance competition and conclude with issues of enforcement and future trends.

---

permitting participation by minors, presents its own set of dangers that cause most children's advocates to caution against online age verification efforts. Protecting against such dangers also lies outside the scope of this Article.

We expect that, if our proposal gains traction, others will accept the challenge to identify and solve these and other difficult issues we can only conjure, including how to ensure the market is accessible to—and not exploitative of—people with various other hurdles to full participation, including people with developmental disabilities, incarcerated people, and service members stationed abroad.

### *C. Summary List of Policies*

As mentioned, this Article should not be read as a blueprint. Rather, we offer what appear to be minimum policies a regulator must promulgate, enact, or enforce if the market we envision is to operate as intended. Our expectation is that these policies, if adopted and enacted, would create incentives that would encourage conduct by actors in digital markets that would bring about the outcomes we desire. The underlying laws that would be needed to create a market for data would first have to give consumers necessary control rights (or expressly acknowledge these rights as pre-existing) over their personal data, and second establish a regulator with the power to set rules in these markets. The following is a summary of those policies, as they relate to intermediaries, data, and the user interface.

**Intermediaries**[24]

- Each data intermediary would be required by statute to act in the fiduciary interests of its users.
- Each data intermediary would also be directed to comply with data minimization principles,[25] balancing this goal with the goal of monetizing user data and creating datasets that are valuable to consumers and society.
- Data intermediaries would be licensed and have strict regulatory requirements in terms of data security, cybersecruity, and resilience.
- Data intermediaries would be independent and could not vertically integrate with any other business whose products or services relate to the use of personal data. They could not sign exclusive deals with any firm that provides a product or service that is reasonably necessary to the business of another intermediary.
- Each user each year would contract with just one data intermediary. That intermediary would serve as the user's exclusive agent for purposes of monetizing the user's personal data generated through the use of any of their devices or web-connected products comprising the Internet of Things (IoT). Requiring that each user have only one intermediary at

24. Several proposed policies governing intermediaries are intended to promote vigorous competition among them. Doing so may be especially difficult during the early years of the market when the intermediaries have zero or only a few years of results to tout. Some may turn to third parties to whom the intermediaries would pay commissions to help recruit users. We see potential benefits to the use of recruiters—who can help users understand differences between the intermediaries—but also potential downsides. The authors agree that the regulator will need to institute some sort of policy with respect to recruiters. We cannot at this time presuppose what that policy should be. The "right" policy will depend on the state of the market at the time, including the percentage of users who have committed to sign up with a data intermediary.

25. "Data minimization" is a principle articulated in, and reflected throughout, GDPR and related regulations. The principle calls on all those who control personal data "to collect only the personal data they really need, and . . . keep it only for as long as they need it." *See Glossary D*, EUR. DATA PROT. SUPERVISOR, https://edps.europa.eu/data-protection/data-protection/glossary/d_en [https://perma.cc/8DDR-LNPV].

any given time would enable the data intermediary to build up a good picture of the user and act as a bottleneck to that user, both of which are important for maximizing the value of the user's data and also for making the complete set of their data available for research and other beneficial uses.

- Intermediaries would compete for users by offering users a share of revenue in exchange for monetizing users' data (with a minimum set by the regulator as a percentage of revenue). The cash value of that share would depend on the intermediaries' business acumen, including its ability to attract consumers, retain them, and create value from the data those consumers choose to share.
- Intermediaries also would compete along dimensions such as commitment to data security, customer service, brand, and success in facilitating innovative data use by researchers, government agencies, and other firms that can provide services to users.

**Data**

- Once a user has chosen an intermediary, the chosen intermediary would collect all online data for that user by monitoring browser and app use. The intermediary could choose the level of detail at which it collects the data, which may affect the manner and rate at which it can be monetized.
- The data collected by each intermediary would reflect a balance between its obligation to adhere to principles of data minimization, on one hand, and its incentive to promote beneficial data uses and to monetize data at rates high enough to permit competitive revenue-share returns to its users.
- The intermediary would sell access to its users based on this personal data. In order to develop cohorts for targeting of display ads, it could carry out its own data analytics across the data. The use of cohorts helps to protect individual privacy.
- Intermediaries also would sell personal data to search services that wish to advertise on the basis of the personal data (search query) entered by the user.
- Web services could collect personal data they need to provide their service, if used solely for this purpose, and could choose the level of granularity at which they collect the data that is suitable for that purpose. (As an example, a search engine may collect and use personal data to provide relevant organic results, but not to target related advertising.) They could carry out data analytics across user data, so long as this is needed for that purpose. They could not share data with third parties (or provide services to third parties based on the data) unless this is required for this purpose.
- Web services must buy personal data needed for any other purpose from the intermediary. The regulator would develop rules about how far in

advance of "use" the web services should be permitted to buy access to such data, and for how long the access lasts.

- Third parties that are not licensed data intermediaries could not transact in personal data except with a licensed data intermediary.
- Data intermediaries could assist users by making payments on their behalf to web services that charge a monetary payment and deducting the subscription or other fees from the total amount that otherwise would have been paid to users.

**User Interface**

- Intermediaries would offer consumers a choice between a small number of standardized "data sharing tiers" or "data sharing levels" that, among other things, afford different levels of remuneration.
- Users must have the ability to observe the collection and use permissions associated with each sharing tier—including their data portability choices—within a clear user-friendly interface.
- The system would encourage users to sign up to a data intermediary using nudges, defaults, and most importantly, the offer of payments. The regulator could develop additional methods to encourage participation, including public-education campaigns to dispell misinformation.
- Because data intermediaries would compete for users on the basis of payments and services, the regulator would design an environment that enables easy comparison of intermediaries, salience of terms, an open-enrollment period when offers for the coming year are made, and low switching costs.
- Intermediaries should minimize friction in switching between them, for example by including a button that effectuates a transfer of data to another intermediary. If the raw data must be downloaded, they should be downloadable to a standardized format that other intermediaries can upload easily.
- At the user's request, the intermediary would share raw data from a particular web service with third parties (e.g., a rival webs service). This feature would enable data portability between web services, intensifying competition in those markets.
- Some types of data (such as health data) are so sensitive that they should not be used for targeted advertising at all, but users should still have the option, made available via the user interface, of proactively sharing such data with services.

## II. The Data Intermediary Regime

### *A. Data Intermediaries*

Data intermediaries are needed to help consumers achieve control and remuneration, and to encourage efficient and best uses of the data that are

generated and collected. First, the fixed bargaining and transaction costs required for a single web user to be compensated from a particular web service would likely be large relative to the payment at issue. An organization representing many consumers, however, could distribute these fixed costs over a large client base. Secondly, web services tend to exploit behavioral biases and design choice architecture to exploit consumers. Individual consumers who attempt to make a decision each time they visit a web service would experience choice fatigue and overload, while each web service would have an incentive to create a choice architecture that induces consumers to pick the option that is most profitable for the web service.

Third, there exists a wedge between the marginal and average value of some sorts of consumer data to web services. This wedge arises because of scale economies in consumer data as used in web services' production of revenue. Some web services analyze or share data as a part of their core service; the application Waze, for example, provides information about traffic conditions based on data collected from its users. But the marginal value of data to such web services diminishes as it collects more data of the same type. In general, data collected from a group of consumers may be analyzed so that it can usefully predict the preferences of a user outside that group; in other words, there is a data externality.[26] Consequently, a firm's valuation of the marginal web user's data is typically low, and the web user might not get much from bargaining with the firm even if bargaining between individual consumers and firms were feasible. This outcome would occur even if the web service's value from data averaged across web users were high. If data are "social" in this way, consumers would benefit from joining together in a group to monetize their data. An organization representing many web users would be better placed than an individual web user to both create this kind of knowledge, and then bargain over the average value of consumer data with firms.

And last, firms in online industries have sufficient market power that any attempt by users to take some surplus for themselves, either by refusing to share data, or by wanting compensation for it, can be blocked. Government intervention to create control rights for users and a system in which they could express their preferences—from sharing everything to sharing nothing—is required for any change. Those control rights include permitting licensed intermediaries to collect and monetize personal data in accordance with user permissions; reserving the right to use the internet anonymously without any collection of personal data collection; prohibiting particular uses of data as the regulator learns about possible harms; and prohibiting any firm or person other

26. *See* Bergemann et al., *supra* note 3, at 264; Jay Pil Choi, Doh-Shin Jeon & Byung-Cheol Kim, *Privacy and Personal Data Collection with Information Externalities*, 173 J. PUB. ECON. 113, 115-116 (2019).

than a licensed intermediary from monetizing personal data or purporting to exercise any control over their use.[27]

For all these reasons, the notion of an intermediary is popular among policymakers. Several different proposals from academics and think tanks have put forward different versions of intermediaries. There are startups attempting to solve this problem that have positioned themselves as intermediaries. And, most promisingly, the European Union has created a legal framework for a data intermediary.[28] The lack of rules around personal data has mostly been a topic of study for lawyers rather than economists, and so they rarely include a proposal to permit the user who generates the data, or the intermediaries who manage the data, to be remunerated. Of the policy proposals, only that of Eric Posner and Glen Weyl[29] (and the subsequent and related proposal by a nonprofit founded by Glen Weyl, RadicalxChange[30]) are designed to give users monetary compensation for their data.

Specifically, Posner and Weyl envision "data coalitions" as introduced by the RadicalxChange foundation in their Data Freedom Act (DFA).[31] As is the case with our data intermediaries, Posner and Weyl's data coalitions would owe fiduciary duties to their communities of data providers (web users), and in turn, would coordinate to collectively bargain over the use of their data. Also like our data intermediaries, data coalitions would have (i) centralized bargaining power over consumer data and (ii) monetize some of their consumers' data. In a nutshell, this proposal envisions intermediaries as collective organizations through which web users could stand in solidarity with each other to advance their common interests.

We see here are two crucial distinctions between our approach and the DFA. Firstly, the DFA envisions significant participation by data-coalition members in the governance of these data coalitions.[32] By contrast, the actions of the data intermediaries in our framework require no democratic legitimation, but rather are to be launched in the same manner as other private entities, for example by incorporation or the creation of a limited partnership. The democratic processes upon which data coalitions would rely may not work well when consumers have behavioral biases and are reluctant to engage with decisions concerning their data in a substantive way. Secondly, the DFA does not impose

27. Data influence market dynamics in ways that economists are continuing to explore. For example, some data in some circumstances can facilitate competition, whereas other data in other circumstances can help enshrine market power. *See* Dirk Bergemann & Alessandro Bonatti, Data, Competition and Digital Platforms (Sep. 27, 2022) (unpublished manuscript) https://ssrn.com/abstract =4236337 [https://perma.cc/H6LS-VBY3].

28. *See Proposal for a Regulation of the European Parliament and of the Council on European Data Governance*, COM (2020) 767 final (Nov. 11, 2020).

29. *See* POSNER & WEYL, *supra* note 9, at 246-49.

30. *See* RADICALXCHANGE, *supra* note 16, at 10-18.

31. *See id.* at 5-6.

32. According to the Data Freedom Act, at least one third of the governing body of any data coalition must consist of representatives that are elected to these positions once per year. Moreover, significant collective choices require approval by a majority of the members. *Id.* at 23-24, 28.

substantial restrictions on the feasible contracts between consumers and data coalitions. Instead, these contracts would be relatively freely determined in negotiations between the data coalitions and web services. The authors of the DFA suggest that the democratic structure of data coalitions suffices to ensure that the terms of these contracts will be favorable for consumers. By contrast, we have pondered the consumer-optimal design of these contracts (compensation schemes, data-share tiers, and usage restrictions) in the face of possible agency problems.

There are several other more-nuanced differences between the DFA and our approach. For instance, the DFA focuses on relational data and claims that this has stronger relevance and value than individual data. We do not take this stance. This is why, in this Article, we address almost exclusively the nature of transactional processes for *individual* data, such as auction mechanisms and definitions of data necessary to service users. Moreover, the DFA imposes different constraints on consumers' options than we do. For example, the DFA allows consumers to be a part of multiple data coalitions, while we forbid such multi-homing. In addition, the DFA states that data-coalition members can be bound to a data coalition for six months. There is no analogue to this in our Article, in part because switching is the principal way we expect consumers to express and exert pressure as to the combination of features offered by different intermediaries.

There also have been efforts by firms to compensate web users for their data, but these efforts have either taken place at a small scale or have been short lived.[33] The Article below analyzes the issues of market design that will need to be solved for these markets to work in delivering efficient and substantial compensation to users.

A data intermediary would carry out its tasks by installing a piece of software on an enrolled user's device(s). The software would enable the intermediary to observe the user's visits and activity online and through mobile apps. The regulator would issue rules prohibiting websites, apps, and the like from engaging in practices that inhibit such observation. The intermediary would, necessarily, have tremendous access to users' private information.[34] We recommend that data intermediaries be licensed by a regulator and adhere to the principle of data minimization. This would ensure that intermediaries satisfy any security or privacy regulations that the regulator establishes, including regulations intended to prevent intermediaries from providing personal data to one web service that is competitively sensitive to another (the gross sales figures

33. One example from the 1990s is NetZero, an internet service provider that offered its consumers free internet in exchange for the right to display targeted ads to these consumers as they browsed the web. NetZero no longer offers this service. *See* Jason K. Krause, *Last Call for Free Web Access*, CNN (Oct. 26, 1998, 1:10 PM EDT) http://www.cnn.com/TECH/computing/9810/26/lastfree.idg/index.html [https://perma.cc/2TXF-C3VH].

34. The intermediary also would gain information about one web service that would be valuable to its competitors. The regulator could guard against such indirect espionage by maintaining a data base of firms' principal competitors, which firms with such concerns could name under oath.

for a particular product, for example). It also would permit the regulator to revoke the license in the event of violations by the intermediary.

We also recommend that our data intermediaries have a fiduciary responsibility to their users. This legal designation requires that an organization act in the best interests of its users. As economists, we find this tool to be helpful because no regulation can be perfectly complete or comprehensive. In any situation where the intermediary has a choice of action, a fiduciary responsibility will discourage it from making the choice that harms consumers.

Despite the potential profitability of data intermediaries and their potential benefits for consumers, data intermediation of the sort we conceptualize has not spontaneously arisen. The discussion below should make clear that new rights and incentives must be created to change the incentives of all parties involved. The complexities of delivering competition at every level to benefit consumers requires a new regulatory framework.

### *B. Scope of Data Covered*

The European Union's GDPR defines personal data as information "related to an identified or identifiable natural person," emphasizing that their regulation covers data that allows a person's identity to be directly or indirectly inferred. Here, indirectly inferring someone's identity means inferring an identifier of the person such as their telephone number or their vehicle's license-plate number. Similarly, the California Consumer Privacy Rights Act (CPRA) of 2020 defines personal information as "information that identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household."[35]

The kinds of data and information that GDPR and CCPR describe as personal may or may not be particularly valuable. An advertiser may be more interested in knowing that a consumer is searching for yellow shoes in New York City even if they cannot be identified, than it is with information that would permit the advertiser to learn their name. If user-created data are valuable, users should be compensated even if those data cannot reasonably be used to infer a user's identity. Thus, we consider a broader class of data than the GDPR and the CPRA. We use the term "personal data" to refer to all data describing an individual's characteristics, transactions history, and browsing history that is generated by that individual, even if the data cannot lead an advertiser or other purchaser to the actual person who generated the data. An efficient data market would create the right incentives for sharing that data.

We talk of "consumer data" when we mean data that do not make possible any linking of any information that relates to a particular individual to that individual. Consider a dataset that maps the number of clicks a particular ad gets by the hour over the course of a day; such a dataset comprises consumer data. We talk of "personal data" when we mean the actual Google searches conducted

35. CAL. CIV. CODE § 1798.140(v)(1) (West 2023).

by users. These pieces of information are "personal data." Only "personal data" are covered by our proposed regulation and included in the market we envision. The regulator, based on experience over time, could issue rules further delineating this distinction.

An important concept going forward is the categorization of personal data that is needed to provide the service requested. An ecommerce site needs a physical address to deliver a package, a search engine needs a query in order to provide organic results, a social network needs the content in posts in order to share them with a user's friend, and a recommender system such as a music-streaming service needs data about selections, likes, and so forth to propose playlists.

By contrast, web services also use personal data to serve targeted advertising or paid promotions. The ecommerce site may design an ad based on a user's delivery address, a search engine may select ads to display based on the query submitted, a social network may choose ads based on the emotional valence of a user's posts. These sales of ads are not the direct service that the user requested, even though they may help fund it. Today, personalized advertising occurs in a setting that is unregulated and has not been designed to benefit consumers; this is the situation our regime is designed to make more efficient.

### *C. Monetization Function of the Data Intermediary*

We have established what constitutes the personal data in our market and have introduced the intermediary that would be the agent of the consumer. How would the intermediary both safeguard and monetize the consumer's personal data? After a user chooses an intermediary and instructs it on what to share, and technical steps have been carried out, the intermediary could see all the user's online activity. The intermediary now must try to monetize the data the user wishes to share while protecting the other data.

It is important that intermediaries face incentives to enter the intermediation industry and make investments that benefit their competitive position. Investment in the industry is important because data intermediation would require the development of new technology for transmitting, pricing, and protecting data. The benefits of the intermediation regime for web users also would depend on intermediaries' success in negotiating with web services over the surplus generated by users' data; intermediaries would earn the revenue from their users' data and would therefore be incentivized to negotiate effectively with web services. We discuss payments to users below.

The intermediaries' monetization of personal data would mostly relate to digital advertising. In the bargaining between intermediaries and web services selling advertising, it is important to understand the outside option for both sides. The system is set up so that users single home; they would enroll with one intermediary and therefore advertisers would face a monopoly seller of that

user's data. If the advertiser does not contract with the intermediary, it could not target ads using personal data to that intermediary's users.

We begin with a discussion of search advertising. If an individual searches for "white running shoes," this is a piece of personal data. The control rights established by the rules would allow "first parties" to utilize the data they collect via their own web service in order to provide that service. As such, Google could use such data in a search query to return organic search results and to train its algorithm. However, under the proposed rules, Google could not use this consumer data to serve advertising to the user because that is not the core service requested by the user, but rather a way of monetizing that service.

To use the personal data "individual X is searching for white running shoes" to advertise to that consumer, Google would be required to purchase that data from the user's data intermediary. In general, the data intermediary would not sell access to raw personal data except when it benefited the user to do so. The specific information contained in personal search queries are what lead to high prices for search advertising. The intermediary would determine that the information is in a category that a user wishes to share, establish a price through some mechanism, and collect the funds from the web service. If Google and the intermediary cannot agree on a price for the user's search query, Google could serve ads that do not require individual data—such as an ad for Insomnia Cookies if it is 2am. In that instance the intermediary would not collect any revenue from Google for the data.

The intermediary would also be able to monetize the user's possible purchase intent (individual X is searching for white running shoes) to web services more widely. Google may be well-positioned to serve an advertisement, since the user is currently on its site, but other search services—whether a search engine or an ecommerce site or other location—may also wish to purchase such data. This may be helpful in promoting much-needed competition in the search advertising market. However, such web services could not retain the personal data query for future use.

A query that is necessary to carry out the service desired by the consumer would not require the web service to purchase the personal query data. For example, consumers on an ecommerce site who search for "white running shoes" expect to be shown products that match that description so they can purchase one. The ecommerce platform's algorithm that determines which items a user sees at the top of the results may be based on personal data from many consumers' searches and purchases. We discuss the rules that would apply to data used for product improvement and analytics below.

Display advertising works differently. In the current system, a supply side platform (SSP) may sell display advertising on behalf of publishers by purchasing personal data on users from an intermediary to help it determine what ad to serve in any given slot. Advertisers who wish to advertise to certain types

of users would—likely through an SSP[36]—query the intermediary to determine which users qualify. The answer would determine whether or not the user falls in the category the advertiser wants, in which case they would see the ad. If the intermediary and the SSP cannot agree on the price of the data, the SSP could turn to a rival intermediary and seek data from its cohorts of (different) users. Again, however, because users single-home, an intermediary would be the monopoly seller of its own users and would have significant bargaining power. At all times the publisher could choose to sell a contextual ad,[37] rather than a targeted ad, and not use any personal data.

Under our regime, intermediaries would be responsible for developing, contracting, or partnering to participate in programmatic ad auctions. An intermediary could choose, for instance, to group its clients into cohorts based on interests, demographics, or locations (e.g., young male bicyclists in the U.S. Northeast) and accept bids that condition on cohort membership. Firms that specialize in designing profitable cohorts or effective algorithms could partner with intermediaries. Additionally, the structure of advertising auctions in which the publisher pays for data as outlined in the preceding paragraph is one of many possible auction structures that intermediaries could facilitate. It would also be possible for an intermediary to provide data to advertisers that seek to bid in an auction; in this case, it would be possible for the advertiser's payment for data to be conditional on winning the auction. What matters is that all uses of the web user's data would be authorized by that user's intermediary.

There are also many imaginable structures for the publisher's payment. One possible structure would involve the publisher paying a commission to the intermediary that is proportional to its revenue from the auction. Another would feature a flat fee for the data to which the intermediary provides access at the outset of the auction. Rather than enforce a particular structure, we leave it to intermediaries and publishers to bilaterally settle on payment schemes. It is possible, for instance, that the intermediary and the publisher could agree to trade access to the intermediary's clients' data for a reduced price for client access to the publisher's content. Intermediaries could differentiate along any of these dimensions.

One of the primary functions of our intermediaries is to consolidate the sources of data that can be queried by firms serving various functions in the delivery of targeted ads, and then pass on the gains from this coordination to web users. As documented by Gentzkow, Shapiro, Yang, and Yurukoglu, advertising

36. A Supply Side Platform (SSP) assists publishers who supply the space on screens where users might view advertisements. SSP's offer these "opportunities"—which generally include information about the type of ad space offered (banner ad, pop up ad, dimensions & pixel requirements, etc.) and about the "eyeballs" that could see it (male is in his 20s within 100 feet of a donut shop at 8:30 AM)—for auction through an ad exchange.

37. A contextual ad is one for which the placement decision depends exclusively on the media "context" in which it is placed—e.g., midway through the second screen of an article about the U.S. Open in the April issue of Golf Magazine at a particular time on a particular date—and not on any data about the person behind the eyeballs that might see the ad. Search advertising is not "contextual," because it necessarily is responsive to search queries, which are personal data.

channels whose viewers are difficult to reach through a particular medium (e.g., television or the internet) fetch higher ad prices.[38] Young men, for example, watch relatively little television, and much of their television watching is concentrated on sports channels; they single home on these channels. As a result, these sports channels command high ad prices because they offer access to an audience that is difficult to reach anywhere else. Given that our intermediaries would have rich data on their users, they would be able to identify users who single home and on which web services. This information would facilitate the sale of ad impressions at high prices. Additionally, a data intermediary would have exclusive control over targeted advertising to a particular web user because users would single home at intermediaries. The theoretical and empirical findings highlighted by Gentzkow and his coauthors suggest that our intermediary system would generate higher ad prices for intermediaries, most of which would be passed back to web users via competition between intermediaries.[39]

Intermediaries would serve a similar "consolidation" role for firms that want to examine personal data, or some version of that data, for market research or product development purposes. Some such uses would be efficient and consistent with users' best interests, even though the users may have no relationship with the inquiring firms and may even constitute the target market for a product or service being developed. Intermediaries would evaluate such opportunities through the lens of their fiduciary duties to users and may monetize some or all such opportunities. For example, an entrepreneur might wish to purchase data on food purchases to help design an emerging food delivery service. Intermediaries also would consider providing access to personal information for various prosocial purposes, such as medical research, research on the impacts of technology, or public purposes such as evaluating bus routes and schedules. Because each intermediary presumably would represent millions of users, their ability to facilitate arrangements for access to large data sets would open opportunities for innovation, to be balanced as always against the users' best interests.

### *D. Standardized "Data-Share" Levels*

The value of a user's personal data would depend on what sort of data they are willing to share. Data could be arrayed from least valuable to most valuable from the perspective of an advertiser. Likewise, data could be arrayed from least costly to share in terms of privacy and intrusiveness to most costly. A market mechanism would help users choose to share the data that is worth more to advertisers than it is to them. Because users visit many web services in a day,

38. Matthew Gentzkow, Jesse M. Shapiro, Frank Yang & Ali Yurukoglu, *Pricing Power in Advertising Markets: Theory and Evidence* 2-3 (Nat'l Bureau of Econ. Rsch., Working Paper No. 30278, 2022), https://www.nber.org/papers/w30278 [https://perma.cc/Q9XD-3UZF].

39. *Id*.

each of which may have different interests and a different business model, the process of determining what is efficient to share would be complex.

A major problem a consumer faces when trying to control the collection and use of their data in the status quo is that user agreements are impenetrable. They are long, contain legal concepts and jargon, and are not realistically understandable by regular users with finite time. Furthermore, spending the time and effort to understand the terms delays the use of the service, which is the user's immediate goal, so they have an incentive to skip over the task. Web services therefore design these terms of service to be ignored by consumers, which means they can contain terms that advantage the web service and harm the consumer.

Any data market that wishes to give users control over the use of their personal data must develop a system that recognizes and accommodates real-world user behavior. In our proposal a regulator would establish a standardized menu of data-share levels. The standardization of levels would allow the regulator to design descriptions that consumers can understand, does not require consumers to learn new terminology or concepts when they change web services, and removes a web service's ability to tailor data sharing descriptions so they are confusing. The idea of standardized tiers is not new; they are found, for example, in the Affordable Care Act's standardization of tiers (e.g., "Bronze," "Silver," etc.) for health insurance policies. A regulator would require that data intermediaries offer these tiers to their clients. The regulator would design the tiers to represent average user views on what constitutes decreasing levels of privacy. Users would then choose how much data to share in a standardized environment designed by a regulator to be understandable and clear.

Each data-share level would be a distinct collection of types of data a consumer would allow the intermediary to monetize. Web services would be able to contract to use these data as we describe below. Web services would not be permitted to use any data outside the categories selected by the consumer. Of course, as noted above, a web service could retain and use personal data that are essential for performing its core functions as requested by a user regardless of what tier those data fall in or the user's choice of data tier. Thus, the data-share levels would control the sorts of data that the data intermediary may make available to facilitate the personalization of advertisements and product and content recommendations (other than those recommendations made as a part of the service requested by the user). We expect that a small percentage of such transactions would be with the web service that collected the data directly (first-party requests) and that the bulk of transactions would be with web services such as advertisers or publishers for use in crafting offers and bids for ad impressions (third-party requests). We elaborate on the distinctions between uses of data later in the Article.

The first tier of data sharing we define (tier 1) includes basic demographics, which includes information on age, gender, location of residence at the ZIP code level (or some other geographical unit of similar size), and other personal characteristics that are not considered sensitive (see the discussion below). Tier

1 also includes the applications installed on a user's device, which are similarly indicative of the user's personal characteristics (e.g., an application for hockey news indicates that the user is a hockey fan, whereas general news applications indicate an interest in current affairs).

Tier 2 includes browsing and app-usage data, which refers to all data generated by the user's activities on web sites and applications that interface with the internet. Browsing history is especially valuable for two reasons: it reveals purchase intent, and it facilitates "frequency capping"—that is, the ability to limit the number of times a user sees a particular advertisement. A user's browsing data often reveals purchase intent, which allows advertisers to send ads to the people who want them. In addition, browsing data include information on whether a web user has been served a particular advertisement in the past. This information is highly valuable because advertisers' valuations of ad impressions depend on how often web users have been served their ads before, and possibly how often the web user has been served the ads of rivals. Purchases are excluded for reasons described below.

Tier 3 includes the approximate real-time location of a user. Extremely precise locations, such as which floor of a particular building a user is on may be unduly intrusive. However, an approximate real-time location generates value to advertisers in the local area. For example, a shoe store or pizza restaurant in New Haven, Connecticut might be willing to pay for users within a few miles but would be unwilling to advertise to a person in California.

The final category of data sharing (Tier 4) allows the intermediary to attribute purchases to advertisements or other content that a web user has seen online. These data include the consumer's financial records, data on online transactions, and the user's email receipts and other electronic transaction confirmations. While financial records are sensitive, our justification for including this category is that attribution is highly valued in digital advertising, and thus it would likely to be lucrative for a consumer to share the data that will allow purchases to be attributed to online content. Because financial data must be kept secure, the methods for tracking attribution would need to be developed by intermediaries and the regulator, according to consumer preferences and the value of the data. An intermediary, for instance, could offer web users the ability to link their credit card accounts with their intermediary accounts and, in doing so, authorize the intermediary to search for attribution data and monetize it on behalf of these users.

Last, we create a Tier 0 for consumers who wish to remain anonymous. An anonymity option has several benefits. First, society may want to establish a right for a consumer to anonymously access the web, so a market design should accommodate that. Second, web services would need to develop a plan for supporting this level of data sharing. Web services would be able to show contextual advertising to this potentially large group of anonymous users, but not personalized ads. Those contextual ads may support the web service, or the service may want to offer a lower quality version to Level 0 users. A Level 0 option would be important in the negotiation between the web service and the

intermediary because it would be the outside option if negotiations over the price to pay for data were unsuccessful.

In summary, our proposed data-share levels are:

- **Level 0**: This level features zero data monetization. The intermediary does not pass on or monetize any data to web services for purposes other than servicing users. The data intermediary's role is only to provide first-party information to web services that is necessary for web services to carry out the core functions that the intermediary's users have requested them to perform.
- **Level 1**: The intermediary is additionally permitted to monetize the user's basic demographic information and the set of applications installed on the user's devices to web services.
- **Level 2**: The intermediary is additionally permitted to monetize the user's browsing and app-usage data and data that personally identify the user.
- **Level 3**: The intermediary is additionally permitted to monetize the user's approximate real-time location.
- **Level 4**: The intermediary is additionally permitted to monetize data that facilitate attribution of the user's purchases to digital advertising and the display of other online content.

We offer these definitions in the spirit of providing an example of how a market for data could work, rather than as the final word on either the number of levels or the content of each level. We are not experts in consumers' actual relative preferences and discomfort with the various types of data collections and uses, so these proposed tiers reflect our best effort to create tiers of increasing "invasiveness" and "prospects for monetization." Experts in the design of these levels could modify them based on new learning. And as new functionalities and business models develop, a regulator might want to alter the levels to accommodate new ways of creating value. The regulator could even design tiers as defaults, with further personalization allowed within each tier. However, because the use of data is confusing to consumers and web services' current data policies are impenetrable, there is great value in simplifying the consumer's choice by limiting the number of levels and standardizing them.

Indeed, under our system, an intermediary might even decide not to offer Tier 4 services, for example, on account of a philosophical opposition to the collection or use of data that can be linked to an actual human. That intermediary would make available cohort-level data available to SSPs and advertisers but would not provide access to data that would permit attribution (and could not provide the data because it would not collect them in the first instance). That decision could provide an advantage—users who admire the principled decision might flock to the intermediary—or a disadvantage—the intermediary would not participate in some of the most lucrative arrangements involving personal data. The decision might or might not be sustainable.

A system of levels would be consistent with privacy laws that might ban the collection and use of certain data or establish certain consumer rights or protections (e.g., a ban on targeting cigarette or vaping ads to people who are trying to quit smoking). Indeed, we are in favor of a regulator carefully considering whether there are categories of data to exclude from markets altogether because the chance of harm to consumers is too high. The regulator should be able to protect consumers from exploitation by prohibiting the intermediary from gathering or sharing these data. Categories of data that society might wish to exclude from data markets include: data that reveal political party affiliation, trade union affiliation, sexual orientation and sexual history, religious belief, addictions, and other characteristics that historically have served as the basis for discrimination or disfavored treatment. These characteristics are similar to the categories of sensitive data under the GDPR and the CPRA.[40] Relatedly, the regulator might wish to exclude categories of firms (e.g., online gambling firms) or firms in violation of certain laws or standards such as trade policies or economic sanctions issued by the home country of the regulator.[41]

The rules would also provide "allowances" of a certain number of hours during which a consumer could browse anonymously, regardless of which data sharing level they have selected. In that case the consumer's data intermediary would not transmit any of the consumer's data to any web service, and there would be no collection of the consumer's data by any data intermediary. This feature would allow web users who have chosen to monetize their data to keep certain web activities private. This could be desirable, for instance, for someone diagnosed with a medical condition who seeks to conduct online research about the condition, but who does not want any web service to collect data related to this diagnosis.

40. *See What Personal Data is Considered Sensitive*, EUR. COMM'N, https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en [https://perma.cc/Z3QD-PYCN] (listing categories of sensitive data under the GDPR); CAL. CIV. CODE § 1798.140(v)(1) (West 2023) (listing categories of sensitive data under the CPRA); Katelyn Ringrose, *New Categories, New Rights; The CPRA's Opt-out Provision for Sensitive Data*, IAPP (Feb. 8, 2021) https://iapp.org/news/a/new-categories-new-rights-the-cpras-opt-out-provision-for-sensitive-data [https://perma.cc/4Z64-77TP] (comparing the protection of sensitive data under the GDPR and CPRA). Note that the CPRA considers the contents of a users' messages to be sensitive information.

41. Regulators may also want to restrict the ability of intermediaries to sell data that facilitate targeting consumers based on their vulnerabilities—for example, because they are in financial distress, struggle with addiction, or are "financially unsavvy." *See* Jon Keegan & Joel Eastwood, *From "Heavy Purchasers" of Pregnancy Tests to the Depression-Prone: We Found 650,000 Ways Advertisers Label You*, THE MARKUP (June 8, 2023), https://themarkup.org/privacy/2023/06/08/from-heavy-purchasers-of-pregnancy-tests-to-the-depression-prone-we-found-650000-ways-advertisers-label-you [https:// perma.cc /U7DD-MFN3] (explaining how advertisers use personal data to segment consumers into groups with the same or similar undesirable characteristics and vulnerabilities). For an economic analysis of the welfare consequences of "steering" fallible consumers to products based on errors these consumers make in evaluating offers, see Paul Heidhues, Mats Köster, & Botond Kőszegi, *Steering Fallible Consumers*, 133 THE ECON. J. 1430 (2023) (concluding the welfare effects of "steering" based on good information about consumer errors are likely significantly negative).

### *E. Combining Data and Dollars*

Our proposed regulation would require each data intermediary to transparently state the amount of compensation that it offers to any consumer that chooses a given privacy level in calendar year. They would choose these prices at the start of an "open-enrollment period" when consumers would choose an intermediary for the coming year.

One challenge in designing the market for data intermediaries lies in choosing how intermediaries may set their compensation schedules. The scheme must be incentive compatible, promote competition, and be understandable to consumers. We propose a system in which each intermediary would provide the same base level of compensation to every consumer within a particular data-share level, subject to some adjustments as explained in the following paragraphs. Flat rates would let the intermediary make different payments to web users who choose different data-share levels; they would also allow compensation to vary across intermediaries within a particular tier.

Because the value of users' data and their activity is uncertain, an intermediary might want to guarantee a certain monthly payment and once the year is over, top up each user proportionately with a bonus that depends on total revenue. An intermediary could offer a payout of Y% of revenue or $X per month, whichever was larger. The $X per month would be guaranteed and would therefore be the price displayed on the choice screen. However, intermediaries might develop reputations for good performance and positive bonuses.

Another possible compensation system would involve intermediaries paying each web user a share of the revenue that this web user's data generates. Such compensation schemes would lower risk but generate several problems. First, determining the exact value of any individual user's data would likely to be challenging given the presence of complementarities between the data of different users, for example, in analytics. Additionally, this system could induce moral hazard. Moral hazard, in this setting, refers to the possibility that web users could increase the revenue that their data generates, and thus their compensation, by adjusting their browsing patterns. As an example, the consumer may browse websites that they are not interested in because providing data to these sites earns them a higher payment from the intermediary. Although it is possible that data intermediaries would be able to develop methods that detect, discourage, and eliminate devious web use, it seems simpler to choose compensation schemes that do not generate any moral hazard.

An exception to our stipulation of constant remuneration for web users belonging to the same privacy level and intermediary is that users who spend different shares of browsing time in anonymous mode would receive different levels of remuneration. Recall that we would allow web users of any share level to use the internet anonymously—that is, to use the internet without passive data collection or personalization. A web user could select a share level allowing for a great extent of data monetization (e.g., Level 4) and always browse anonymously, which would prevent the user's data from being monetized. Such

behavior would undermine the value of share levels that allow for substantial data sharing for users who seek to monetize their data. Therefore, we propose allowing intermediaries to mark down web users' compensation by the fraction of time that they spend in anonymous mode. A Level 4 web user who spends half of their time in anonymous mode, for instance, would receive half of the advertised payment for Level 4 under this system. Such a policy would have no practical consequence for users who only occasionally browse anonymously.

There are significant uncertainties inherent in starting up a whole new market, and firms may either lose or profit more than they forecast until an equilibrium is achieved. To forestall excess profits, we suggest the regulator apply a loss ratio rule familiar from the health insurance context. Such a requirement would require intermediaries to pay out a minimum level of revenues to users (e.g., 70%) by tier. At the end of the year, the intermediary would assess its revenues and payments and, if needed, increase payments to its clients to meet the threshold. This type of rule would ensure that web users are remunerated for their data as the regulator fine tunes market design.

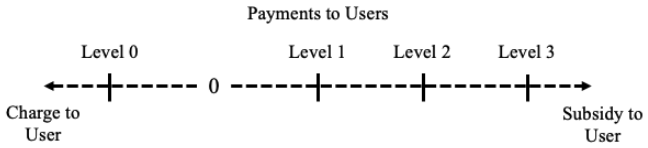
Intermediaries or the regulator may want to design details of this compensation system to increase web users' enthusiasm about data intermediation. The regulator could set a common schedule for the payment of annual lump sums or bonuses to occur on December 1 for example, which would help consumers yearning for liquidity around the holidays and increase the salience of the payment. Intermediaries would be free to design a combination of monthly and end-of-year components. The regulator could also allow intermediaries to offer payments to consumers in the form of credits for certain products (e.g., the mobile phone bill). Intermediaries could also form around causes and contribute their profits to that cause, rather than to consumers. However, every intermediary would be required to state a clear expected annual dollar amount for each privacy tier.

Irrespective of the chosen regulations for intermediaries' compensation schemes, competition between intermediaries is desirable in that it will help ensure that the surplus generated from web users' data is largely returned to web users.

We conclude this Section by raising a tricky economic issue: web users who choose Level 0 would fail to generate revenue for their data intermediaries but would create costs (e.g., for managing the data that web services use for servicing their users). Thus, data intermediaries that do not receive other revenue from Level 0 users would face an incentive to discourage users from selecting Level 0—for example, by offering a poor quality of service to these users. To address this perverse incentive, we propose allowing data intermediaries to charge fee to users selecting Level 0. Such a fee is akin to charging web users a positive price

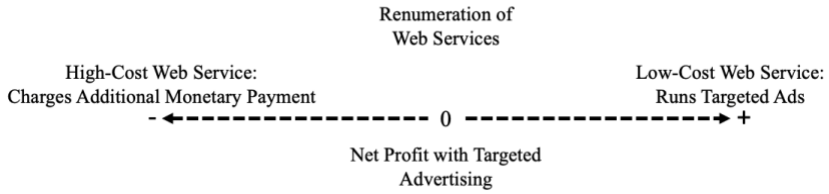
(rather than compensating, i.e., charging a negative price). Any such fee would also need to be transparently displayed at the time of annual enrollment.[42]

Figure 2. Payment Scheme: Illustration




The optimal level of quality of a web service might be sustainable with contextual advertising, it might require the higher level of revenue generated by targeted advertising, or it might be sufficiently expensive (or have consumers who are not valuable) that it requires another revenue source like a subscription to cover its costs. In general, one should think about the net cost of web services under a targeted advertising regime as ranging from positive to negative. Services that can earn more than their costs through advertising are on the positive side, while those that require a cash payment to run are on the negative side.

Figure 3. Categories of Web Content: Illustration




A user may want to access both types of content, and the data intermediary should facilitate that. The proposal thus far has discussed the business on the right, the positive side, because these businesses generate a surplus under targeted advertising and can share it with users. However, if a user is building up

42. It may seem incongruous that Level 0 users may have to pay a fee to their intermediaries, given that the control right we propose includes the right to decline any collection or use of personal data. Deeming something a "right," however, doesn't imply that exercising that right must be free in all instances. Citizens in most countries may have the right to drive a car on public roads after passing a test, but they nonetheless might have to pay a fee to obtain the license that confirms the right. Persons charged with serious crimes have a right to counsel; but if counsel is appointed, counsel fees may be assessed as court costs upon conviction. *See* Matthew Menendez, Lauren-Brooke Eisen, Noah Atchison & Michael Crowley, *The Steep Costs of Criminal Fees and Fines*, BRENNAN CTR. FOR JUST. (Nov. 21, 2019), https://www.brennancenter.org/our-work/research-reports/steep-costs-criminal-justice-fees-and-fines [https://perma.cc/VTV6-Z6ED] (noting that court fees "cover almost every part of the criminal justice process and can include court-appointed attorney fees"). It also is possible that Level 0 consumers would not have to pay a fee. People who select this level are likely to be wealthier than the average user, in that their selection indicates they have sufficient economic freedom to choose anonymity over payment. Intermediaries would want to attract such users with the hope of "upselling" them in future years into levels that permit more sharing of their relatively high value data, which the intermediary would hope to be able to monetize at higher-than-average rates. Intermediaries could offer a "no-fee" introductory rate, for example, which itself would spur competition among the intermediaries.

funds with the intermediary due to sharing their personal data, there is no reason that the user could not ask the intermediary to spend some of those funds to allow their to access websites that charge subscriptions. An intermediary could negotiate a price schedule with web services like newspapers. Because the intermediary would reduce transaction costs, users need not sign up for annual expensive subscriptions, but would be able to pay $.10 to read an article in a newspaper. We call this a "microsubscription."

The microsubscription price would be negotiated by the intermediary and then discounted according to the data tier of the user—because the data barter the user has chosen offsets the monetary cost. The intermediary would signal to users that they have arrived at a web service that costs money and tell them the net price. The user could then decide to go ahead or not. In this way consumers would be offered choices to pay with money when there is expensive content available that cannot be bartered for. The price schedule available to users and the range of content covered by an intermediary would be one of the dimensions over which the intermediary bargains. As noted above, to prevent exploitation of Level 0 users, web services that offer content in exchange for personal data only would also provide a version of their service that is available without monetary or data charge.

The possibility of incorporating microsubscriptions might also solve some of the business challenges created by Level 0 users. These users do not want to barter for services with their data. Therefore, they would need to pay more in money. The regulator might want to permit intermediaries to charge a transaction fee (e.g., 5%) for payments made by Level 0 users. If this were clearly stated on the choice screen, intermediaries would compete on this dimension. But intermediaries would have a way to earn profit from Level 0 users and would not avoid serving them.

### *F. Competition Among Data Intermediaries for Consumers*

Data intermediaries would compete for consumers on a number of dimensions. First and most obviously, consumers would be attracted by the level of remuneration offered by the intermediary. Higher payments (or payments in attractive forms) would attract more users, who would generate revenues and economies of scale. If a consumer plans to choose Level 0 data sharing, for which we expect intermediaries will charge a fee, lower *fees* would attract users to intermediaries in the same way that higher or better forms of *payments* would attract users who choose to share more personal data. Intermediaries might even offer an introductory rate of zero fees for Level 0 for the first year, if the user remains with the intermediary for the following year.[43]

43. We assume that competition will generate consumer benefits along dimensions including price and innovation because, as antitrust economists, we know this to be the case. This is precisely why we do not recommend a single, government-operated or government-controlled entity to act as intermediary: it would not face the competitive forces that generate the same consumer benefits as competition will, and so consumers would be worse off.

Consumers would also choose intermediaries based on the quality of service and the user interface. For example, data intermediaries would likely provide identity and log-in services to their users as these are very convenient services. In the status quo, large digital platforms including Google and Facebook allow their users to log into third-party web services using their platform credentials, which allows those platforms to track users. Our data intermediaries could offer their own log-in services to facilitate the sharing of data for servicing users and protect it. To provide an example, the intermediary could share the web user's shipping address, credit card information, and language preferences with ecommerce sites at which the web user places orders. Thus, data intermediaries would replace the status quo regime in which web users either need to create distinct profiles at many web sites, providing each with a copy of the user's data, or provide their data to a large platform that may use the web user's data for other purposes.

Consumers might also want to join an intermediary that is differentiated in some way, for instance its payments support a cause the consumer values. Or perhaps the intermediary would enter high-profile partnerships with research institutes that rely on personal data to design solutions to any number of social or health-related problems—for example, to increase voter participation in local elections, combat climate change, and so forth.

Intermediaries might engage in marketing to inform consumers about their attributes and benefits. A regulator would need to mandate transparency to aid competition. For example, a regulator could require data intermediaries to provide a data dashboard that clearly communicates to web users how their data have been used online. Although we propose that the regulator mandate intermediaries to report certain types of information on their data dashboards—namely, which data have been shared, with whom the data have been shared, and the level of compensation that the user has received as a consequence of data sharing—intermediaries should be encouraged to innovate in designing their dashboards.

The higher the per-person revenue of the intermediary, the larger the remuneration it could offer. Intermediaries would therefore compete on the basis of innovation in monetizing user data. This might occur through creation of more effective cohorts for display advertising, or better algorithms to evaluate consumer data. To the extent the intermediary can extract more value from consumer data, it can raise the amount of remuneration it offers and gain market share and profit. Intermediaries would also have an incentive to invest in contracting and negotiation with web services as this would increase their revenue.

We are concerned that web services might discriminate against users selecting restrictive privacy levels. Under our regulatory regime, web services would be capable of inferring web users' privacy levels by sending data intermediaries queries whose results depend on the web users' selected privacy levels. If those users are lower value, web services might have an incentive to charge higher prices or offer lower service quality to users who select these

levels. In a competitive market, data intermediaries would be able to address the problem outlined above by negotiating with web services to provide their clients with quality service at a reasonable price. Data intermediaries would have an incentive to ensure all their clients have positive experiences using the internet as long as they profit from serving a client irrespective of the client's type. When data intermediaries are free to charge fees, or positive prices, to Level 0 clients, we expect this to be the case.

A data intermediary would interact with any web service used by one of the intermediary's users that requests access to personal data. Similarly, a web service that seeks to use personal data for each of its users would have to interact with each intermediary used by that group. One of the dimensions on which intermediaries would compete is their ability to design efficient contracts and negotiate with web services. Larger web services would likely want unique bilaterally negotiated contracts with intermediaries. Small web services would likely choose from a menu of standardized contracts specifying payment rates and other terms. (This pattern resembles the contracting system in the food delivery industry; under this system, delivery platforms offer restaurants contracts specifying levels of promotion and commission rates.[44])

In order to understand the parties' bargaining positions and the extent to which contracts would favor intermediaries versus web services, we need to specify the services available to users in the case in which the parties fail to reach an agreement to sell personal data. It is critical that users who choose not to share any personal data, or users belonging to intermediaries that do not have a contract with a web service, could still access the site. A system where those users were shut out of the internet would give undue bargaining power to the web service. In order to ensure that the web remains widely accessible to web users without personal data to sell, the rules must require that any web service that offers a version of its service at a zero cash price must make that or a similar version of its service also available to users who have selected Tier 0 (and therefore have declined to sell access to their data) and to users whose intermediary has failed to reach an agreement with that web service (and therefore, despite their presumed willingness to sell access to their data, can't effectuate such a trade until an agreement is reached). Such services could still require users to provide personal data necessary to provide the service the user requests. An online store, for example, could require users in Tier 0 and users whose intermediaries have no agreement with the online shop to provide an address for delivery of the items the user purchases without running afoul if this rule.

## III. The User Interface

This Part discusses how the user-facing aspects of the data intermediary regime can be designed to promote our proposal's goals. Behavioral economics

44. *See Grow Online Sales with Doordash*, DOORDASH (2022), https://get.doordash.com/en-us/products/marketplace [https://perma.cc/WT4E-Z4QF].

literature suggests various reasons why the current regime based on informed consent for data sharing is unlikely to be exercised in a fruitful manner by consumers. We emphasize in particular the desirability of a clear and simple choice architecture in light of web users' attention costs and behavioral biases.

### *A. Choice Architecture*

Behavioral economics emphasizes the role of choice architecture—that is, the way alternatives are presented as opposed to intrinsic characteristics of the alternatives—in consumer decision-making. Default options, for instance, are important drivers of consumer choice even when it is not costly to switch from defaults.[45] The visual presentation of information about choices also matters. Today, important information about the use of consumer data is concealed in visually unappealing terms and conditions notices.[46] The psychological literature on decision fatigue (alternatively called *ego depletion*) also suggests that a consumer's decision-making ability worsens as the consumer makes more successive decisions; thus, we would expect that the abundance of data-use choices, each specific to a particular web service, leads to suboptimal consumer choices.[47] Regulation of digital markets should take these behavioral aspects of decision-making into account by establishing defaults that are in line with consumer preferences, ensuring that important information about alternatives is salient wherever consumers make decisions about their data, and limiting the number of separate choices required of consumers.

Furthermore, behavioral Industrial Organization literature points out that profit-maximizing firms have an incentive to find that decision-frame in which

45. *See* Eric J. Johnson, Suzanne B. Shu, Benedict G. C. Dellaert, Craig Fox, Daniel G. Goldstein, Gerald Häubl, Richard P. Larrick, John W. Payne, Ellen Peters, David Schkade, Brian Wansink & Elke U. Weber, *Beyond Nudges: Tools of a Choice Architecture*, 23 MKTG LETTERS 487, 488-93 (2012) (citing numerous studies in behavioral economics and psychology providing evidence on the role of choice architecture—including defaults—in consumer decision making).

46. *See* Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty & Arvind Narayanan, *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*, PROC. ACM ON HUM.-COMPUT. INTERACTION, Nov. 2019, art. 81, at 81:2. (providing evidence of the use of dark patterns by commercial websites to influence users into disclosing information. The authors define dark patterns as "user interface design choices that benefit an online service by coercing, steering, or deceiving users into making unintended and potentially harmful decisions").

47. *See* Roy F. Baumeister, Ellen Bratslavsky, Mark Muraven & Dianne M. Tice, *Ego depletion: Is the active self a limited resource?*, 74J. PERSONALITY & SOC. PSYCH. 1252, 1253 (1998) (defining ego depletion as "temporary reduction in the self's capacity or willingness to engage in volitional action . . . caused by prior exercise of volition," and providing experimental evidence for this phenomenon). There is also considerable evidence of ego depletion in behavioral economics. *See, e.g.*, David Hirschleifer, Yaron Levi, Ben Lourie & Siew Hong Teoh, *Decision Fatigue and Heuristic Analyst Forecasts*, 133 J. FIN. ECON. 83 (2019) (finding that the accuracy of financial analysts' forecasts decreases as the number of previous forecasts made in the same day increases); Emil Persson, Kinga Barrafrem, Andreas Meunier & Gustav Tinghög, *The Effect of Decision Fatigue on Surgeons' Clinical Decision Making*, 28 HEALTH ECON. 1194 (2019) (finding that surgeons are less likely to schedule a patient for an operation late in their shifts); Ned Augenblick & Scott Nicholson, *Ballot Position, Choice Fatigue, and Voter Behaviour*, 83 REV. ECON. STUD. 460 (2016) (finding evidence of decision fatigue in voting by exploiting variation in choices' positions on ballots caused by differences in the number of local ballot measures).

consumers are most likely to make the decision the firm prefers.[48] For example, if consumers tend to understand most ways of presenting the choice correctly but, due to naivete, misunderstand one framing thereof, this will eventually drive firms to select the frame consumers misunderstand. By regulating the frame—the data-share levels and how they are presented—we circumvent this problem.

Although we emphasize the role of behavioral biases, it is worth noting that even a consumer without these biases is likely to be overwhelmed: the real costs of learning about terms of service and switching costs are high. Reading through lengthy terms of service agreements requires users' time and effort and may also require legal training to understand. These costs discourage web users from attempting to make good decisions about their data. By obfuscating terms of service agreements, a web service can induce consumers to consent to provisions that they would otherwise oppose. Web services may further change their terms in unfavorable ways after users are locked in. Because these tactics make it infeasible for most consumers to forecast the sorts of data they are sharing and the value of these data, these consumers cannot effectively make choices, let alone negotiate with web services, over the value of their data.[49]

Web users are unlikely to understand the technical language that technologists or privacy experts may use to describe the privacy levels we introduced in the previous Part. Inconsistency across intermediaries in the language used to describe these concepts could further confuse consumers. Data intermediaries must accommodate limited consumer literacy and actively reduce obstacles to consumer comprehension of the privacy levels. The most vulnerable consumers may have the most difficulty understanding the tradeoffs involved in

48. *See, e.g.*, Tom Blake, Sarah Moshary, Kane Sweeney & Steve Tadelis, *Price Salience and Product Choice*, 40 MKTG SCI. 619, 625 (2021) (presenting evidence suggesting online vendors have an incentive to display mandatory fees at the end of the purchasing process because consumers are less responsive to fees displayed at this stage than fees prominently displayed at the outset of the purchasing process); *see also* Jennifer Brown, Tanjim Hossain & John Morgan, *Shrouded Attributes and Information Suppression: Evidence from the Field*, 125 Q. J. ECON. 859, 870-71 (2010) (noting that sellers can increase overall profits by lowering their advertised prices while adding shipping fees and/or other add-on fees); Liran Einav, Theresa Kuchler, Jonathan Levin & Neel Sundaresan, *Assessing Sale Strategies in Online Markets Using Matched Listings*, 7 AM. ECON. J.: MICROECON. 215, 239-41 (May 2015) (suggesting that consumers do not rationally internalize shipping fees into products' overall prices when making purchasing decisions, and that sellers can exploit these biases by lowering their advertised prices while adding shipping fees and/or other add-on fees).

49. This concern appears in California's Consumer Privacy Rights Act of 2020, which states:

> In practice, consumers are often entering into a form of contractual arrangement in which, while they do not pay money for a good or service, they exchange access to that good or service in return for access to their attention or access to their personal information. Because the value of the personal information they are exchanging for the good or service is often opaque, depending on the practices of the business, consumers often have no good way to value the transaction. In addition, the terms of agreement or policies in which the arrangements are spelled out, are often complex and unclear, and as a result, most consumers never have the time to read or understand them. . . . This asymmetry of information makes it difficult for consumers to understand what they are exchanging and therefore to negotiate effectively with businesses."

California Privacy Rights Act of 2020, 2020 Cal. Stat. A-84, A-86 (codified at CAL. CIV. CODE § 1798.100-99).

their choices. It is therefore imperative to ensure that the privacy levels are described and framed in a way that is intelligible to the population at large.

The regulator could promote comprehension of the privacy levels by developing descriptions written in clear and plain language—and informed by insights from psychology—and by standardizing these descriptions across data intermediaries. We similarly see value in the development of graphical designs describing the privacy levels that would be standardized across intermediaries. Our insistence that intermediaries describe the privacy levels in plain language resembles the GDPR's stipulation that web services communicate how they use personal data in "clear and plain language."[50]

Unlike the GDPR, however, our approach would provide data intermediaries with specific and standardized descriptions of data-share levels. The constancy of data-share level descriptions across intermediaries would make it clear to consumers that all intermediaries offer the same privacy levels and facilitate learning what the levels mean over time. The standardized clear-and-plain-language descriptions of the data-share levels should also describe, and provide examples of, the risks associated with monetizing their data. Better choices would result from consumer understanding of both the benefits and costs of data sharing.[51]

Our data-share levels would not allow users to customize the types of data that individual web services can use. A user, for example, could not choose to permit one web service to access their Level 3 data, but another service to access only their Level 2 data. The majority of authors rule out this form of customization to simplify web users' choice problem, which is desirable for the reasons enumerated above. A minority of the authors, however, believe an ability to personalize is critical in securing widespread use of the intermediaries and effectuating the control right our proposal promises. We also, note, however, that a user who did not wish to share data with a particular web service could choose to access that web service anonymously under our proposal.

One criticism of our proposal is that consumers would overvalue the tangible, short-run benefits of remuneration from data sharing relative to the risks of data sharing, which are uncertain and may only be realized after many years. Although we take this criticism seriously, we hope that the intermediary regime's data security regulations mitigate the criticism. We also hope that our standardized, plain-language descriptions of the privacy levels would help consumers understand the risks associated with each privacy level. We view our proposal as an improvement over the status quo in this regard, as the status quo

50. It also resembles the CPRA's insistence that "Consumers should be entitled to a clear explanation of the uses of their personal information." *Id.*

51. *Active Online Choices: Designing to Empower Users*, THE BEHAVIOURAL INSIGHTS TEAM, DOTEVERYONE & CTR. FOR DATA ETHICS & INNOVATION (Nov. 2020) https://www.bi.team/wp-content/uploads/2020/11/CDEI-Active-Online-Choices-Desk-Research-Write-up-FOR-PUBLICATION-1.pdf [https://perma.cc/BXH6-W3W9] (suggesting a series of principles for designing choice architecture that promotes users' ability to make active choices, including making the trade-offs involved in a choice interactive—i.e., allowing the user to interact with, or experience, what the trade-off means).

offers consumers little effective control over their exposure to risks associated with data sharing.

A related concern is that a system allowing consumers to monetize their data could increase inequality in access to privacy or freedom from being targeted, by providing poorer consumers with greater incentives to sell their data.[52] This is particularly concerning because malicious forms of targeting often prey on people of lower socioeconomic status (e.g., predatory loans).[53] However, our proposal limits web services' abilities to target users relative to the status quo in which targeting is unavoidable for most web users and does not involve compensation for these users. Additionally, our proposal is compatible with consumer protection regulations intended to limit malicious targeting.

### *B. Mechanics of the User Interface and Adoption*

It is important to consider the mechanism by which users would sign up with a data intermediary. There are many options, and we don't claim expertise in marketing, user interface design, or other knowledge sets that might inform the design decision. With these caveats in mind, however, we recommend that the regulator create a standardized application (we will refer to this as the Data Manager going forward) that would come pre-installed on devices with capabilities for accessing the internet. If this is part of U.S. regulation, a developer of any operating system for consumer devices sold in the U.S. must install the regulator's Data Manager program. When the user first connects to the internet on a device with a Data Manager installation, the Data Manager would present the user with a choice menu displaying the privacy levels, the standardized descriptions of each level, and the dollar range of compensation for each level. Once the user has selected a privacy level, the Data Manager would display a listing of data intermediaries and the compensation level and terms offered by each for the user's selected privacy level.

Evidence suggests that a significant share of users, despite the compensation offered by intermediaries, would wish not to engage with the Data Manager and would instead quickly click through its choice menus. To protect these users, we suggest that the regulator specify default choices. One default, for example, could be the selection of Level 2 and the data intermediary that offers the highest level of cash compensation for that privacy level.

Over time people would replace their devices with new ones that open with a Data Manager, prompting them to register with an intermediary when they initialize the device. Web users would also voluntarily connect their older devices to intermediaries as information about intermediaries' benefits spreads.

---

52. *See* Stacy-Ann Elvy, *Paying for Privacy and the Personal Data Economy*, 117 COLUM. L. REV. 1369, 1406-13 (2017).

53. *See* Craig E. Wills & Can Tatar, *Understanding What They Do with What They Know* 13 (Workshop on Priv. in the Elec. Soc'y, 2012) (showing the relevance of this concern for Facebook; in particular, finding that Facebook targeted ads to users on the basis of sensitive personal characteristics such as health status and sexual orientation).

Nudges and incentives may be a better method of establishing this market than a mandate. Growing awareness of the intermediaries' payments to consumers could give those who are hesitant a financial incentive to sign up. And if competition among the intermediaries works as it should, we would expect the intermediaries that pay the most to their users, or return the highest percentage of advertising spend, would trumpet that fact in advertising, creating an incentive not just to participate in the system, but to sign up with the particular intermediaries that consumers think will pay them the most.

Data intermediaries' offerings to advertisers would become less valuable if their consumers *also* joined competing intermediaries, as this would nullify the intermediary's status as a local monopolist over its consumers' data. If a user were paid their value, they would have a strong incentive to be exclusive with one intermediary. For this reason, an intermediary should be able to offer terms that applied to an exclusive consumer contract at a given point in time, although it should be straightforward for the consumer to switch to a new intermediary using the Data Manager application described in the next Section.

There would likely always be some users who do not belong to a data intermediary. Such a user might actively provide information to web services for the purposes of servicing the user, but the web service could not monetize that personal data in exactly the same way it could not monetize enrolled users' data. It is critical that these users would not be profitable for web services in order to ensure that there would be no incentive for platforms to discourage users from joining an intermediary. If users who wished to share their data earned significant payments from an intermediary and those who preferred limited sharing could manage their privacy more effectively through an intermediary, consumers would have incentives to enroll. Above we explain why we propose that new devices come with a default enrollment stage to raise participation.

If the data of a web user who does not register with an intermediary were treated similarly to that of a web user who selects an intermediary's Level 0, it might seem that a privacy-concerned user who would select Level 0 over the other levels would have little reason to sign up with an intermediary at all. This would be especially true if the intermediary, which could not monetize its Level 0 clients, charged a fee to these clients rather than remunerate them. Various services offered by intermediaries to Level 0 clients, however, would make it worthwhile for web users who would select this level to sign up for intermediaries. In the remainder of this Part, we discuss several of these services.

### *C. Switching Among Intermediaries*

In order to incentivize intermediaries to compete for consumers based on the size of their payments, it is important that competition between them be vigorous. This means that the regulator must set up rules to lower switching costs and create salience around the choice of intermediary.

The Data Manager would facilitate choice when a user first operates a new device. Likewise, the Data Manager would be the tool users employ to enroll in

an intermediary each year. The regulator could intensify competition by choosing an "open enrollment" period during which users would be presented with salient information about the remuneration they obtained (or fees they paid) in the current year and the offers available, at their current privacy level, for the coming year. The regulator could develop the information and messaging for intermediaries to deliver to users to facilitate their choices. Users would be directed to the Data Manager where they could make the choice in a controlled environment.

Based on evidence from other markets, there may be many users who are passive and do not make an active choice of intermediary.[54] The regulator may wish to develop an automatic enrollment scheme that is fair to users and intensifies competition in the marketplace. As already mentioned, auto enrollment could make a conservative choice such as privacy Level 2 and the highest cash price in that level. If multiple intermediaries had similar cash prices, permitting them all to be allocated a share of passive users would intensify competition. A user who is automatically enrolled should have the chance to make a subsequent active choice of intermediary. An alternative system would shift many of the users who exert minimal effort to the selection process from the lowest-compensating intermediary to the highest-compensating intermediary. Users other than those signed up with the highest-compensating intermediary would receive a notification during open enrollment of the form:

> Your current provider offers $X less than the best offer for the same level of privacy you have currently use. Do you want to:
>
> A: Look at Choice Screen
> B: Switch to Best Paying Offer
> C: Stay with Current Provider

Such a notice would operate as a semi-automatic switch, resulting in users who exercise minimal effort selecting their data intermediary switching in large numbers the intermediary that pays the most to users.

To keep service levels high, users would want to bring their web-usage data with them when they switch intermediaries. To lower switching costs, the

54. *See, e.g.*, Kate Ho, Joseph Hogan & Fiona M. Scott Morton, *The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program*, 48 RAND J. ECON. 877 (2017) (demonstrating that consumer inertia in the U.S. healthcare market allows insurers to charge higher premiums); Ali Hortaçsu, Seyed Ali Madanizadeh & Steven L. Puller, *Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market*, 9 AM. ECON. J. 192, 220-24 (2017) (documenting significant consumer inertia in the choice of electricity provider and estimating that information interventions can significantly raise consumer surplus); Lukasz Grzybowski & Ambre Nicolle, *Estimating Consumer Inertia in Repeated Choices of Smartphones*, 69 J. INDUS. ECON. 33, 47-54 (2021) (documenting how consumer inertia facilitates concentration in the smartphone industry); Alexander MacKay & Marc Remer, *Consumer Inertia and Market Power* 13-19 (Harv. Bus. Sch., Working Paper No. 19-111, 2022), https://www.hbs.edu/ris/Publication%20Files/19-111_298206b6-5217-4905-a381-d7173ae957cc.pdf [https://perma.cc/FFL3-UE75] (documenting how consumer inertia impacts the simulated price effects of a merger in the gasoline industry).

regulator would need to require that data intermediaries transfer raw consumer data in a standardized format upon request. The data format should be specified by the regulator, so it would be consistent across intermediaries. If a user's data were transferred in an unusable format, it would frustrate the goal of easy switching and vigorous competition. To encourage the development of technologies that maximize the value of web users' data, we would not require intermediaries to share any secondary analysis that they have applied to a web user's data when that web user switches to a different intermediary. The intermediary would retain the property rights to the algorithms or learning it created from its former users, but not their raw data. Once a user is no longer enrolled with an intermediary, it would be required to delete that user's data. If a user did not re-enroll with an intermediary, prior to deletion the intermediary would need to send the user their raw data in the standardized format, so the consumer controlled the data, or transfer them to a different intermediary.

A regulator may need to address the possibility of consumers multihoming across devices. One could imagine, for instance, a consumer using one intermediary for accessing the internet on their laptop and a different intermediary for accessing the internet on their mobile phone. The consumer's data specific to any particular device would not be as rich as the pooled version. Ideally, this behavior should be prohibited because it would make the consumer's data less valuable and harms them financially. Additionally, the consumer's platform would lose its position as a local monopolist over that particular consumer's data when the consumer used different intermediaries on different devices. The policy response to this problem would depend on whether equilibrium prices result in a financial gain or penalty to users who run a different account for each device as well as on technological solutions for monitoring it.

Conversely, multiple consumers (e.g., members of the same household) could share a device. One possibility is to allow consumers to choose different intermediaries by using different device accounts through a browser, for example. Another is to establish a "one device-one intermediary" system in which data from a smart refrigerator, for example, is collected by a single intermediary, which avoids the unworkable solution of asking different household members to "log in" each time they peek in the fridge.

### *D. Enabling Data Portability*

One of the more useful capabilities of a data intermediary would be its ability to lower switching costs between web services. For example, if a user moves from one ecommerce site to another, the new site would not initially have data on the user's methods of payment, frequent mailing addresses, past purchases, and other information that would improve the quality of service. For this reason, the user might not want to switch ecommerce sites, and, in turn, new entrants into ecommerce would be discouraged from entering. A consumer's data intermediary would often have the data needed to lower these switching costs.

We propose that users be able to instruct intermediaries to transfer relevant data to approved web services. For example, the entering ecommerce site could prompt a new user to authorize their data intermediary to send categories of data from their existing ecommerce site. Such a tool must be carefully overseen by the regulator because it would be capable of transferring large amounts of the consumer's data. The regulator might consider requiring that web services wishing to receive ported data acquire a license. Certifying the security of web services should increase their attractiveness to consumers and thereby increase contestability. An alternative method might involve downloading to a uniform format that then is easily uploadable to a different intermediary.

## IV. Types of Data Use

In this Part, we discuss various uses of data generated by web users from the perspective of web services. The three major uses of web users' data that we identify are (i) servicing users, (ii) targeting users, and (iii) conducting analytics. Each type of use would imply different restrictions in terms of data-access rights. Additionally, the treatment of data used for analytics would bear important consequences for the competition policy implications of our proposed data intermediary regime.

### *A. First-Party Data for Servicing Users*

The emergence of the internet as a means of communication has created significant economic value, which sensible regulation should aim to safeguard. The internet requires a minimal exchange of data to function and achieve gains of trade between web users and firms, and these essential minimal data flows should be free: firms should not face restrictions in accessing data that fulfill the purpose of enabling the technology and the core value proposition that the firm offers to the web user.

Within the context of a web user's interaction with a web service, the category of *data for servicing users* encompasses all data required to allow the web service to provide its core services as requested by the web user. Examples of data for servicing users include data required to order products (e.g., shipping and billing addresses), receive directions (e.g., real-time location data), and send messages (e.g., directories of social connections). In order to allow web services to offer attractive products and to accommodate the manifold beneficial uses of web users' data, the privacy levels defined in the web user section therefore would not apply to data for servicing users. That is, web services could generally use data for providing core services requested by users without paying intermediaries for the right to use these data. Note that the classification of information as data for servicing users does not reflect any intrinsic characteristic of the information but rather the purpose for which the information is used in a particular context. The fact that LinkedIn uses employment-history data for servicing users by providing them with digital resumes does not mean that

LinkedIn's use of that data for other purposes (e.g., selling the data to recruiters) would qualify as data for servicing users.

The less stringent rules on data needed for servicing the user may be abused by web services that make overbroad claims about what they need. A regulator would inevitably need to investigate difficult cases and disallow unnecessary data collection. A regulator might want to engage in rulemaking to create clarity among market participants. We recommend that no use of data for targeting users or for analytics can qualify as a use of data for servicing users, even if it is requested by the user. This delimitation provides a broad definition of data for servicing users while enabling our goals of web user remuneration and privacy control.

### *B. First-Party Data for Targeting Users*

We define targeting as encompassing all instances in which web users' data are used to assist in promoting services that would cost the user additional resources, whether the payment takes the form of money or data. Recommendations that do not lead to more user expenditures, such as film recommendations for a user that already has a Netflix subscription or Spotify's compilation of playlists, would be exempted. Likewise, an exercise app that reminds the user to stretch or take a walk after periods of physical inactivity would be exempted. Recommendations on an ecommerce site would, however, count as data for targeting users because following the recommendations of the platform by buying the highest-ranked product would cost the consumer money while raising the revenue of the ecommerce site. Likewise, recommendations for apps that charge in some way (e.g., in-app purchases as well as the app itself) would also be commercial recommendations.

A helpful way to think about targeting is that it leverages private signals about consumer intent. Search advertising is targeted for this reason. Search ads are chosen based on information the consumer provided to the web service; that information constitutes a private signal of consumer intent. A user who types a query into a search engine such as "best quality running shoes" is actively giving specific information to the web service. In this example, the same piece of personal data could be used to target display advertising to the user. We would categorize advertising based on the query as using personal data and therefore subject to the regime.

By contrast, a user who is browsing an online newspaper and chooses to read a story about running has revealed only limited information about herself––she is interested in a story about running. That piece of information is fairly vague since the reader could be a runner, could be related to the writer of the story, could live near the route of the run, or have another reason for reading the story. An indirect revelation of interests that is mediated through the publisher, and *which requires no data generated by the web user*, would be contextual. Drawing the line between contextual and personalized advertisement using the criterion of the involvement of a publisher would create good economic

incentives for publishers to create compelling and specific material. But we expect the line between these types of advertising to be unclear at times, and the regulator would have to develop guidelines to help web services comply with the rules.

### *C. Data for Analytics*

The interactions between web users and web services generates a constant flow of data that are potentially useful for conducting analytics and improving service quality. For example, a company that uses a recommendation system needs access to data about the success of previous recommendations in order to further develop and improve its algorithm. The treatment of data for analytics relates to data that have been generated by web users in the past. The use of data for analytics does not rely on the identity of the user who generated the data. Nor do the analytics directly cause the user to be contacted to make any additional purchases.

It is not obvious what the correct policy for data analytics is, and more research in this area would be welcome. If web services were permitted to freely collect first-party data on their own users for analytics and product improvement, then consumers would gain from the resulting innovation and quality. However, there are two advantages to requiring the web service to purchase access to these data to carry out its analytics. First, it would increase the intermediary's revenue to be passed on to web users. Second, the regulator could issue rules governing the circumstances by which intermediary could make the data available for analytics by entrants, presumably at the same cost. For example, the rules might prohibit a dating site from gaining access to raw data collected from other dating sites, even for analytics, because such data reveal highly sensitive regarding preferences and proclivities. Competitors could train their algorithms on rivals' databases and overcome scale-disadvantages.

We note that a system that gives the firm that created the data a lower cost of analyzing it would create good incentives for that firm to enter and innovate in the first place. Moreover, consistent with users' data-share levels, data intermediaries may sell access to these data so that entrants and competitors may also improve their analytics. If the intermediary sells data in an equitable and nonexclusive manner, all entrants would be able to pay to learn about the market—through existing users. The regulator could promote entry by establishing rules that prevent exclusives and discrimination in the use of data for analytics.

## V. Controlling the Behavior of Parties

### *A. Risks to Users*

Data intermediaries would possess extensive control over consumer data, which raises the concern that they would become targets for attackers aiming to

abuse or steal data. One protection a regulator could mandate is strict compliance with best practices for data protection and regular self-assessments and third-party audits. We additionally propose that the regulator hold data intermediaries and web services to the principle of data minimization, which is one of the central principles of the European Union's GDPR. In our setting, enforcing data minimization by regulation means establishing rules that prevent intermediaries from sharing or using more data than is pertinent to the purpose of a particular application of web users' data. It also means minimizing the storage of users' data in places where it faces the risk of a breach. Both intermediaries and web services would have a role to play in respecting data minimization; each type of agent would be capable of exposing its clients' data to risks and should actively minimize the exposure of these clients' data to risks.

The specifics of the regulations intended to promote data minimization should be based on the advice of data security experts, and deliberations about these regulations should acknowledge the importance of not creating unnecessary entry barriers. With these regulations in hand, the regulator can use a combination of audits, investigations into complaints, and technological solutions to obtain compliance.

To ensure that the regulator could effectively respond to conduct by intermediaries that violates their fiduciary duties or other rules, we recommend that data intermediaries be required to hold a license from the regulator. A system of licensing gives the regulator the ability to create standards, strict data protection regulations, and responsibilities for intermediaries. Such a system would also facilitate the punishment of intermediaries that violate these standards by revoking their licenses or levying less severe sanctions, for example, fines.

Last, our proposed regulation would hold data intermediaries to strict standards for the protection of their users' data to which other firms in the digital economy may not be subject. A web user who signs up with an intermediary and selects Level 0 would therefore receive a higher level of data protection than a user who does not sign up with intermediary.

Another issue of concern is whether a web service would discriminate based on a user's data-share level choice. When data-share level choices are correlated with web users' personal characteristics, then a web service's knowledge that a web user has selected a particular privacy level would be a noisy measure of that web user's characteristics. Suppose, for example, that higher-income web users are more likely to choose Level 0. This could be because they fear facing price discrimination based on their incomes, or because they understand the risks of data sharing more deeply than the population on average.[55]

55. For example, households in high-income neighborhoods are systematically charged more for online tutoring packages. Jeff Larson, Surya Mattu & Julia Angwin, Unintended Consequences of Geographical Targeting (Aug. 31, 2015) (unpublished manuscript) https://techscience.org/a/2015090103 [https://perma.cc/KEC3-ZS4X]. There is also empirical evidence that socioeconomic status positively correlates with information and computer technology literacy, and it seems plausible that agents with higher computer literacy levels may also have a stronger desire for privacy due to a deeper understanding

Such a link between income and privacy choices implies that online firms with knowledge about privacy choices would be tempted to offer higher prices to consumers that select stricter privacy levels. This form of price discrimination may be undesirable for several reasons. First, it may reduce consumer welfare (in that paying a higher price without benefiting from output or quality improvements necessarily reduces welfare), even holding privacy choices fixed. Second, it could lead to an inefficiently low uptake of the strictest privacy choices. Consumers may anticipate that selecting strict privacy levels adversely affects the prices they receive down the line. For these reasons, we suggest that the regulator prohibit web services from attempting to infer web users' privacy choices in the data intermediation regime.

### *B. Monopolization of the Intermediary Market*

Data intermediaries could themselves become large digital platforms that exercise market power in ways that affect other market players. A monopolist intermediary would hold a better bargaining position with large digital platforms than a small intermediary in a competitive market. Therefore, the monopolist intermediary may be better able to extract surplus from platforms on behalf of its users. This is, however, not likely to deliver the best outcomes for users. A fiduciary duty to enrollees would be much harder to enforce if users have little choice of data intermediary and if the regulator has little visibility of alternatives. Furthermore, a monopolist intermediary may be inefficient, have high costs, and generally provide a low rate of compensation to consumers. (A loss ratio rule will limit this problem.) A monopolist would also have poor incentives to engage in sophisticated analytics that enable it to provide effective targeted advertising, and this would in turn be bad for both advertisers and users. In general, all these markets would perform better if there were robust competition between intermediaries.

We have described above what the regulator can do to create as much competition as possible between intermediaries. This includes a standardized set of privacy levels, a regulated dashboard to provide users with clear and simple information, salient prices and price competition, a specific period during which all users make their annual choice of intermediary, and mandatory portability of data from one intermediary to another.

However, as with any new market, it is not clear ex ante how much concentration may be driven by economies of scale. For example, some scale would be required of an intermediary for its data to be used in analytics. Suppose, for instance, that a firm wishes to estimate the click-through rate for home appliance ads for members of a specific demographic group—for example, white men between twenty-five and twenty-nine years old residing in the greater New

of how their data are used. *See* Ronny Scherer & Fazilat Siddiq, *The Relation Between Students' Socioeconomic Status and ICT Literacy: Findings from a Meta-Analysis*, 138 COMPUTS. & EDUC. 13, 21-28 (2019).

York City area. To accurately estimate this rate, the intermediary must have tracked a sufficient number of members of this group who have encountered a home appliance ad. Given that the gains to accuracy from a larger sample size are diminishing, however, the importance of scale in conducting analytics may still allow for several intermediaries with the requisite data for conducting these analyses. Likewise, "social data" that has an externality on other users would reward an intermediary that serves a large share of a cohort. By internalizing the data externality, the users and the data intermediary would gain. For targeting advertising to a user, however, scale is less important. If ads displayed to a particular user sell for a higher price when combined with a user's demographic information and browsing history, then this single user's data would be valuable on their own.

Most authors expect that differentiation among intermediaries would naturally limit concentration, although some have significant concerns about the tendency to tip. Intermediaries may contribute their profits to causes that attract a subset of users but not all. Intermediaries may specialize in serving certain types of users. That specialization may cause the development of proprietary and useful algorithms permitting effective monetization of those consumers.

We propose a prohibition on web services holding ownership stakes in data intermediaries and on data intermediaries owning web services.[56] In general, avoiding vertical integration by the intermediary would prevent some competitive issues that could arise. Similarly, competition would be enhanced by a prohibition on exclusive contracts between intermediaries and any other parties such as consultants, demand side platforms (DSPs),[57] and so forth. A regulator tasked with maintaining competition could be empowered with tools it could use if any one intermediary became dominant. For example, if an intermediary passed a certain market share threshold the regulator could be authorized to divide that intermediary into two independent data intermediaries, each with the software and algorithms of the original, but only half of the users.

## VI. Pertinent Legal Issues

### *A. Right To Be Forgotten*

A central goal of this proposal is to give consumers greater control of their data. An important part of this endeavor is the implementation of the "right to be

56. This is in line with the recommendation of the German Data Ethics Commission that "privacy management tools/personal information management systems must continue to serve as dedicated custodians of data subjects' interests, and . . . conflicts of interest must be ruled out." *See Opinion of the Data Ethics Commission*, DATA ETHICS COMM'N. 135 (Dec. 2019), https://www.bmi.bund.de/SharedDocs/downloads/EN/themen/it-digital-policy/datenethikkommission-abschlussgutachten-lang.pdf;jsessionid=4045AFAA7DE42634170DBC87D0936584.1_cid322?__blob=publicationFile&v=5 [https://perma.cc/V3YA-5L2V].

57. DSPs assist advertisers to implement programmatic ad campaigns by determining whether and what to bid on opportunities offered for auction through an exchange, with the goal of placing a large number of high value ads at a low price to their advertising clients.

forgotten" as defined by the GDPR, which gives web users the right to instruct web services to delete their information about the user.[58] The practical implementation of this right has turned out to be challenging. As a response, there are initiatives such as the Data Rights Protocol that aim to facilitate such requests.[59] Given their central position in the data markets we envision, data intermediaries may have an easier time handling these requests and ensuring that web services comply when a user exercises their right. Furthermore, we describe certain data above—personal data generated by one web service but used for analytics by another—for which the regulator should establish time boundaries, after which such data must be destroyed. There may other circumstances in which the regulator should impose time limitations on data access. If enforced, these too should ease implementation of a right to be forgotten by reducing the number of services holding data over time.

### *B. Violations*

Large web services would have strong monetary incentives to find ways of avoiding the need to pay intermediaries to access web users' data. For instance, they could construct separate databases on web users that could be used for targeting without the need of seeking approval from these web users' intermediaries. Such practices, which would be illegal under our proposal, would undermine our envisioned system of intermediation. One way to combat attempts to circumvent intermediaries while accessing web users' data is to design data-access systems in ways that prevent the leakage of raw data. Advertisers, for instance, could be required to programmatically submit bids to intermediaries that are conditioned on web users' characteristics, rather than receive the data of these web users in raw form from intermediaries. Additionally, web services seeking to perform analytics could be required to conduct these analytics on intermediaries' computers. We do not take a stand on which technologies should be employed to minimize risks of data leaks, but we assign to intermediaries the responsibility of using the most suitable technology for this purpose. The regulator should be empowered to penalize intermediaries and any other market participant at a level that creates deterrence. Other ways to avoid illegal use of web users' data include the use of investigations, third-party audits, and rewards for whistleblowers.

## VII. Extensions of the Data Intermediary Framework

Our basic intermediary framework can be extended in a straightforward way to address problems posed by novel technologies, non-standard types of data, and existing digital intermediaries and platforms. We sketch out a few of these here without providing substantial detail on these extensions.

58. Council Regulation 2016/679, art. 17, 2016 O.J. (L 119) 1, 43.
59. DATA RIGHTS PROTOCOL, https://datarightsprotocol.org [https://perma.cc/BK5B-EKEA].

### *A. The Internet of Things*

The internet of things (IoT) refers to devices other than conventional computing devices (e.g., computers, tablets, and mobile phones) that interact with other devices over the internet. Examples of IoT devices include smart refrigerators and televisions that feature internet capabilities. A natural question is how to integrate IoT devices into our data intermediation framework. These devices generate data that, like data generated by the user of computers and mobile phones, can increase overall surplus. For instance, a smart refrigerator might register that a consumer is running out of milk and transfer this information to the consumer's online grocery store, which could then send a targeted recommendation to this consumer's smartphone. Because such a recommendation may prevent an inconvenient situation in which the consumer finds no more milk in their fridge, utilizing this data may raise total welfare.

Given that IoT devices are internet capable, the status quo entails the possibility that the manufacturers of these appliances already collect and utilize such data without restrictions. Allowing this to continue unchecked would stand in contrast to the goals of this Article. Additionally, there are several challenges involved with integrating web users' devices with their intermediation accounts. First, a web user may own many smart devices. Configuring all these devices such that their data may be used by the intermediary may be inconvenient for the consumer. Other challenges result from the fact that IoT devices may be shared by different consumers living in the same household. For example, such consumers may have joined different data intermediaries and may also select different privacy settings. Moreover, the data generated by IoT devices is not necessarily specific to any given individual that uses them but reflects the habits of all users jointly.[60]

The solution to this problem needs more research and exploration. The regulator may want to give web users the choice to configure these devices with intermediaries, however, in which case the same rules should apply to data generated from IoT devices as apply to data generated from conventional devices. To understand how this would apply in practice, consider again the example of a smart refrigerator that seeks to notify its owner to purchase a generic grocery item. As long as the web user explicitly requests recommendations of these forms, the recommendations would only rely on data for servicing users under our proposal given that the provision of such recommendations is one of the smart refrigerator's core purposes. But if the refrigerator company wants to serve an owner an ad for ghee, because the refrigerator knows the owner keeps a bottle of coconut oil in between their jar of chili crisp and their tub of white miso, the company presumably should be required to purchase that information.

60. This problem may apply to shared personal computers in particular and may be mitigated by software that facilitates switching between intermediation profiles on a particular device.

### *B. Internet Service Providers*

Internet service providers (ISPs) are also able to collect and use data on their customers' online activities. Allowing them to do so under our data intermediation regime would undermine the goals of this proposal. Thus, we specify that a consumer's ISP may not use or share any data on its clients without the consent of these clients' intermediaries. The only exception to this rule concerns data that the ISP requires to provide the consumer with high-quality internet access (i.e., data for servicing its users). For example, an ISP knows a consumer's location and will need to refer to it in order to provide service of equipment. To summarize, we treat ISPs as we treat other web services.

### *C. Relational Data*

Online interactions generate relational data, that is information characterizing the relationship of a given person with other web users. We propose that a user's intermediary may share relational data involving the user as long as the identities of other people described by the data are suppressed. As an example, the user's intermediary could share that the user commented on a friend's Facebook photo but not the identity of that friend. More research is needed to understand the implications of regulations in this area.

## VIII. Conclusion

The above discussion demonstrates that establishing a regulatory environment for successful data markets is complex. We hope the ideas in this Article serve as a useful starting point for policy makers thinking through how to make these markets better serve consumers. The status quo—wherein almost all the profit from digital advertising flows to a few large companies, users exercise little to no control over the collection and use of their personal data, and large platforms maintain a stranglehold over the use of collected data, preventing their beneficial use—does not represent a fair outcome for either consumers or advertisers, other market participants, or even the citizenry who might benefit from innovative data uses. It is unlikely to be efficient either: Competitive remuneration for the content that draws users to the web would stimulate the amount and variety of content consumers want, while the availability of their personal data at competitive rates to both entrants and incumbents would enable the entry of web services that use these data.

A market for personal data that allows users to control how their data are used, as well as the chance to benefit from their monetization, addresses multiple interrelated problems at once: the market failure whereby large platforms obtain valuable inputs without user choice and compensation—which reinforces their market power and insulates them from competitive pressures to reform their data practices; the resulting lower quality of user experience and privacy online; and the lack of a data-sharing mechanism to facilitate prosocial innovation that

leverages personal data. A market for personal data, properly constructed, would allow web services to engage in socially valuable advertising as well as investment in product improvements. And by offering a range of sharing tiers, a well-designed market for personal data would also give effect to heterogeneous user preferences regarding the use of their data. We stress, however, that there are many tricky economic issues involved in making such markets work. More research by economists is needed in this important area.

**Appendix 1: Narrative Summary of Related Ideas and Proposals**

We are not the first to suggest an alternative regulatory regime for the collection and use of web users' data, and many other proposals suggest the use of intermediaries. In this Appendix, we review alternative ideas and proposals related to the regulation of web users' data.

Francis Fukuyama, Barak Richman, Ashish Goel, Roberta R. Katz, A. Douglas Melamed, and Marietje Schaake envision a regime in which consumers access digital platforms through third-party web services called middleware.[61] Their proposed middleware services would filter, sort, fact-check, and otherwise control the display of platforms' content to users. An example of middleware would be a website that provides Twitter posts to users after filtering out content determined to be factually incorrect. Another example is a service that sorts Google News articles according to their quality or their match with the user's stated interests. Fukuyama and coauthors argue that middleware would reduce the influence of major platforms over public discourse and the nature of content consumption online. They also argue that competition between middleware providers would offer consumers a choice between privacy settings and other terms of service. Fukuyama and coauthors' proposal is primarily intended to reduce the control that platforms exercise over public discourse and the content consumed online. It does not directly address privacy, nor does it provide remuneration to users.

An alternative approach to the regulation of online markets is to assign fiduciary duties to existing web services. Jack Balkin[62] and Jonathan Zittrain,[63] for instance, suggest treating web services as fiduciaries in their handling of their users' data, that is, as *information fiduciaries*. This suggestion is motivated by the fact that the relationship between web users and web services bears similarities to those that exist between people and their doctors or accountants, who do hold fiduciary duties.[64] Under Balkin and Zittrain's proposal, web services would similarly be expected to ensure the security of their clients' data and to only use their data to their users' benefit. Web services would not be able, for instance, to sell their users' data to firms with weak data-security measures. They would also be barred from using data on consumers' political beliefs to promote their private political objectives. Our proposal would similarly limit web services' use of consumer data, although fiduciary responsibilities under our

61. Francis Fukuyama, Barak Richman, Ashish Goel, Roberta R. Katz, A. Douglas Melamed & Marietje Schaake, *Report of the Working Group on Platform Scale*, CTR. ON PHILANTHROPY & CIV. SOC'Y 30-38 (Nov. 2020), https://pacscenter.stanford.edu/wp-content/uploads/2020/11/platform_scale_whitepaper_-cpc-pacs.pdf [https://perma.cc/N72G-SVTD].

62. Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1186 (2016).

63. Jack M. Balkin & Jonathan Zittrain, *A Grand Bargain to Make Tech Companies Trustworthy*, THE ATLANTIC (Oct. 3, 2016), https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346 [https://perma.cc/3XVV-N7UT].

64. Moreover, assigning web services fiduciary duties would address an argument that web services have used in the past to oppose privacy legislation in the United States, namely that their usage of data is protected by the First Amendment.

regime would rest with the intermediaries providing web services with data rather than with the web services themselves.

Related to our idea are Personal Information Management Systems (PIMS), which (i) store users' data in a safe way (e.g., via encryption) and (ii) allow consumers to specify exactly who can access what parts of their data (in place of cookies). Solid, an initiative headed by World Wide Web inventor Tim Berners-Lee, is an example of a PIMS.[65] Unlike our proposal or Fukuyama and coauthors' middleware proposal, which are proposed regulatory regimes, Solid is an existing online platform (although it is has not yet been widely adopted). Members of Solid store data that they use on the internet in virtual data containers called Pods from which authorized third parties may access members' data. Customizability is an important principle for Solid, which aims provide its users with extensive control over both the storage and use of their data.

There are several other examples of existing PIMS. For example, the platform digi.Me allows users to decide with which web services that integrate with digi.Me they want to share their data.[66] Additionally, digi.Me allows its users to view how a given web service would use their data, and it allows them to specify which parts of their data they want to share. This kind of service is also offered by the UK-based community-interest company MyDex.[67] The app Mine provides a list of web services that possess data on a consumer by accessing the web user's email account and enables the web user to request the deletion of this data.[68]

The European Data Protection Supervisor (EDPS) has identified PIMS as a possible policy solution and has laid out a set of features that PIMS should ideally have.[69] The desiderata laid out in these papers include some of the desirable properties of the intermediaries we envision, namely (i) enabling consumers to control access to their data, (ii) fostering transparency and traceability of data usage, and (iii) facilitating data portability and data minimization. At odds with our goals, the 2016 Report states that "as a matter of principle PIMS will not be in a position to 'sell' personal data, but rather, their role will be to allow third parties to use personal data, for specific purposes."[70] Congruent with this notion, most existing PIMS do not entail remuneration of consumers for their data.

65. *About Solid*, SOLID, solidproject.org/about [https://perma.cc/9WST-Z5D7].

66. *The Digi.me Core Tech Stack*, DIGI.ME, digi.me/features-core [https://perma.cc/X7PR-YTKS].

67. *Mydex Charter*, MYDEX, https://mydex.org/about-us/mydex-charter [https://perma.cc/295B-U92K].

68. MINE, saymine.com [https://perma.cc/685G-AKFB].

69. *Opinion 9/2016: Personal Information Management Systems*, EUR. DATA PROT. SUPERVISOR (Oct. 20, 2016), https://edps.europa.eu/sites/edp/files/publication/16-10-20_pims_opinion_en.pdf [https://perma.cc/J4WR-4RDP] [hereinafter *2016 Report*]; *TechDispatch #3/2020 - Personal Information Management Systems*, EUR. DATA PROT. SUPERVISOR (Jan. 6, 2021), https://edps.europa.eu/data-protection/our-work/publications/techdispatch/techdispatch-32020-personal-information_en [https://perma.cc/X9Y3-GYD7].

70. *See 2016 Report*, *supra* note **Error! Bookmark not defined.**, at 13.

Our data intermediaries are similar to PIMS in that they would consolidate and handle consumers' data. However, our data intermediaries would act on behalf of consumers in a fiduciary capacity, while PIMS provide consumers with direct control over all parts of their data. This distinction is substantial, given the goals of this proposal. First, we aim to establish well-functioning remuneration schemes that endow consumers with a fair share of the value of their data. The main reason why our intermediaries would be able to remunerate consumers is that they control access to many consumers' data, which gives the intermediaries both bargaining power and economies of scale in transacting with web services. By contrast, PIMS users separately control their data. Economic theory suggests that remuneration may thus be much lower due to the absence of bargaining power on the consumer side.

Secondly, we develop a system explicitly designed for web users with behavioral biases and decision fatigue. We also note that the behavioral-economics literature indicates that most consumers do not want to spend substantial time managing the finer details of how their data are used and stored.[71] Thus, consumers would only need to make a few decisions in our approach—they could set a general data-share level and then allow specialized intermediaries to handle data management on their behalf, subject to the selected data-share level's constraints. By contrast, PIMS require much more active decision making; users have to make conscious decisions both when joining a PIMS and any time they grant or revoke data access privileges. Given the behavioral evidence we cite, this may severely impede the adoption and consumer welfare benefits of PIMS.

Last, existing conceptions and implementations of PIMS do not offer detailed discussions of how to avoid outcomes in which web services simply ignore the data stored in PIMS and conduct targeting based on consumer data that is not controlled by a PIMS. By contrast, our regulation would codify that web services can only offer personalized ads or recommendations with data they receive through the intermediary. Any other forms of targeting would be illegal.

Two existing technical solutions that aim to foster consumer privacy or provide remuneration without creating data intermediaries or PIMS are the web browser Brave and the browser extension of Permission. The web browser Brave blocks all third-party ads.[72] All the consumer's browsing data is stored on the consumer's device and is inaccessible to third parties. In addition, Brave has developed several technical solutions to prevent consumer tracking across websites.[73] Users of Brave can choose to receive targeted ads facilitated by Brave itself in exchange for cryptocurrency.[74] Similarly, Permission's browser

71. For a succinct review of the empirical research on privacy behavior *see* Alessandro Acquisti, Laura Brandimarte & George Loewenstein, *Privacy and Human Behavior in the Age of Information*, 347 SCI. 509 (2015).

72. *Advanced Privacy*, BRAVE, https://brave.com/privacy-features [https://perma.cc/5F4W-3TS8].

73. *Id*.

74. *Brave Rewards*, BRAVE, https://brave.com/brave-rewards [https://perma.cc/3H8T-2Y33].

extension tracks a consumer's browsing behavior and suggests targeted ads to the consumer based on the former.[75] A consumer receives cryptocurrency when they view and engage with ads.

The UK Centre for Data Ethics and Innovation has categorized different forms of data intermediaries and the value they can create in the digital economy.[76] In its report, the term "data intermediary" includes, among others, (i) PIMS, (ii) data trusts with fiduciary responsibilities for the users they represent, and (iii) data custodians—that is, platforms that enable the analysis of sensitive data by third parties in a secure way. Our data intermediaries would be endowed with responsibilities similar to those of data trusts and data custodians in the language of this report.

An important desideratum of our data intermediary proposal is the facilitation of data portability.[77] As such, it is related to recent work by Bertin Martens, Geoffrey Parker, Georgios Petropoulos, and Marshall Van Alstyne, who suggest the establishment of an "*in-situ* data access right" for consumers on platforms.[78] Under this right, the users of a given platform would be able to run algorithms designed by third parties on their data stored within the platform. This would facilitate the portability of data and enable the interpretation of data in the context of a given network. The approach of Martens and coauthors differs from ours in the sense that data portability relies on web users' efforts.[79] By contrast, we would allocate this responsibility to the data intermediaries in our framework.

In the proposed Data Governance Act of 2020, the European Commission has also mentioned the role of data intermediaries to help individuals "exercise their rights under the EU GDPR."[80] This act mandates that data intermediaries should take neutral positions in the data markets and should be endowed with fiduciary responsibilities. We build on this proposal by sketching how to design the market for data intermediaries in a consumer-optimal way.

Most closely related to our approach is the idea of data coalitions as introduced by the RadicalxChange foundation in their Data Freedom Act (DFA).[81] We discuss data coalitions and their similarities with and differences

75. *Permission Whitepaper*, PERMISSION, https://permission.io/whitepaper [https://perma.cc/WR6N-RKS9].

76. *Unlocking the Value of Data: Exploring the Role of Data Intermediaries*, CTR. FOR DATA ETHICS & INNOVATION 9-12 (July 22, 2021), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1004925/Data_intermediaries_-_accessible_version.pdf [https://perma.cc/DQ5L-A6RW].

77. Section 26 of the German Telecommunications Act gives the regulator the right to mandate interoperability of dominant players in the digital market. Telekommunikationsgesetz [TKG] [Telecommunications Act], June 23, 2021, BUNDESGESETZBLATT, Teil I [BGBL I] at 1858, § 26.

78. Bertin Martens, Geoffrey Parker, Georgios Petropoulos & Marshall Van Alstyne, *Towards Efficient Information Sharing in Network Markets* 4 (Bruegel, Working Paper 12/2021, Nov. 10, 2021), https://www.bruegel.org/sites/default/files/wp_attachments/WP-12-101121-1.pdf [https://perma.cc/H2UY-FZMA].

79. *See id.* at 4-5.

80. *Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act)*, at 2, COM(2020) 767 final (Nov. 25, 2020).

81. RADICALXCHANGE, *supra* note 16.

from our data intermediaries at some length in the main text.[82] The two most important distinctions are that data coalitions require substantial participation by their members, whereas our intermediaries would require no such direct democratic legitimation, and we envision intermediaries as having the incentive to require users to do little more than select their data sharing tiers. Second, data coalitions would be free to enter whatever contractual relations with their members arise out of the democratic processes used to manage the business of the coalition. By contrast, we would impose significant substantive and procedural restrictions and mandates with respect to the relationship between consumers and their intermediaries, including compensation schemes, data-share tiers, and usage restrictions.

Another natural solution to the problem of online firms exploiting consumers' data is for government to tax the use of data.[83] If data collection and use is more expensive, firms will do less of it. Consumers, however, would not receive direct compensation through a tax. Instead, they would be remunerated indirectly through government expenditure benefiting these web users and others. In addition, a data tax system would not provide consumers with control over their data. Additionally, it would set the incentive for web services not to advertise or collect data based on a uniform, arbitrary tax rate instead of a market price reflecting firms' valuations of consumer data and consumers' valuations of privacy.

One taxation-based proposal that seeks to indirectly provide consumers with remuneration for the use of their data is that of the Berggruen Institute's California Data Dividends Working Group (henceforth the Berggruen Working Group).[84] This working group was formed in response to California Governor Gavin Newsom's proposal for payments from firms using personal data to the public. The Berggruen Working Group proposes that California institute a Data Dividend Tax; this tax would apply both to sales of consumer data and to companies that use or store consumer data, with the amount of the tax dependent on the extent to which the company uses or stores consumer data.[85] Under the Berggruen Working Group's proposal, California would fund public spending that broadly benefits the public using the proceeds from the Data Dividend Tax instead of directly passing these proceeds on to the state's web users.

---

82. *See supra* Section II.A.

83. Paul Romer, for instance, has proposed a tax on targeted digital advertising intended to limit such advertising's political harms. He proposed this tax in a New York Times op-ed and expanded on the proposal in a longer essay. *See* Paul Romer, Opinion, *A Tax that Could Fix Big Tech*, N.Y. TIMES (May 6, 2019), https://www.nytimes.com/2019/05/06/opinion/tax-facebook-google.html [https://perma.cc/PW2R-B78L]; Paul Romer, *Taxing Digital Advertising*, PAULROMER.NET (May 17, 2021), https://adtax.paulromer.net [https://perma.cc/GW8T-ZKHJ].

84. Yakov Feygin, Hanlin Li, Chirag Lala, Brent Hecht, Nicholas Vincent, Luisa Scarcella & Matthew Prewitt, *A Data Dividend that Works: Steps Toward Building an Equitable Data Economy*, BERGGRUEN INST. 5 (May 5, 2021), https://www.berggruen.org/ideas/articles/a-data-dividend-that-works-steps-toward-building-an-equitable-data-economy [https://perma.cc/6BUR-TZC7].

85. The authors of the Berggruen Working Group's report consider various types of taxes that depend on companies' usage and storage of consumer data without taking a stand on which should be adopted. *Id.* at 14.

The Berggruen Working Group's data dividends scheme would provide the public with a benefit proportional to the value generated using its data. But, unlike the establishment of data intermediaries, the introduction of a Data Dividends Tax would not provide web users with a simple way to limit the extent to which their data are shared. Thus, the Berggruen Working Group's proposal does not provide an option for consumers who do *not* wish to monetize their data to opt out of its remuneration scheme. Additionally, consumers would not be able to increase their remuneration by sharing more data under the Data Dividends Tax. If we fear that consumers would share their data out of financial desperation, then this feature of the Data Dividends Tax could be desirable.

Another existing regulatory proposal is the implementation of a cohort learning system. Cohort learning is an alternative to the contextual regime that strengthens consumer privacy relative to the status quo while retaining a degree of personalization. Under cohort learning, a digital platform groups web users into cohorts and personalizes web users' online experiences based on the cohort to which they belong instead of their individual identities. An existing version of a cohort learning system has been developed by Google—namely, the Federated Learning of Cohorts (FLoC) system. This would represent a means of targeting ads without using third-party cookies. Under FLoC, each web user would be assigned a cohort populated by web users with similar web browsing patterns. Each cohort would have a minimum size and an ID code. Ad targeting would be conducted on the basis of web users' cohort ID codes but not their individual characteristics.

This expansion would generally allow web services to target consumers based on their cohort ID codes, both for advertising and for content recommendations. Cohort learning as implemented by an expansion of FLoC would limit the use of personal data and prevent targeting on the basis of the consumer's individual web use history. It would additionally preserve possibilities for personalization and the associated benefits.

A regulator could expand FLoC by requiring Google to allow other web services to purchase access to the consumer data used in implementing FLoC. The regulator could additionally require that Google sells this access in a secure manner at fair, reasonable, and non-discriminatory rates. Regulation could permit or incentivize rival cohort learning regimes; cohort learning regimes could differ in many dimensions, including the procedure for constructing cohorts, the size of cohorts, and the protocols for accessing a user's cohort identifier. A cohort learning regime could also feature competition between cohort learning services and allow for innovation in data management services. Designing regulation that would establish a cohort learning environment to keep users' data secure, permit innovation in the creation of cohorts, and control the price at which they were licensed would surely be difficult. Other than making the point that Google should not be an unregulated supplier of cohort data, we do not further explore alternative cohort learning regimes.

## Appendix 2: Narrative Exploration of the Monetary Value of Personal Data

One argument against establishing a regulatory regime based on data intermediaries is that the establishment of such a regime would entail fixed costs and transaction costs that are large relative to the surplus available to be transferred from firms to web users. In this Appendix, we review the available empirical evidence on the value of user data, which indicates that the surplus firms derive from user data is substantial.

There is a nascent literature studying the value of web users' data for the quality of online services offered by firms. This literature typically assesses improvements of service quality in statistical terms: recommending products/content can be viewed as a prediction task with the goal of achieving the highest possible prediction accuracy. Several studies assess the value of data in the context of search engines. He and coauthors find evidence consistent with improvements in search result accuracy from additional users providing feedback about search results.[86] Yoganarasimhan finds a clear positive relationship between the length of personal data records and various measures of search result accuracy.[87] Additionally, Schaefer and Sapi provide evidence consistent with complementarity effects between the richness of personal data records and the number of users providing feedback in the search engine context.[88]

Several other studies assess the value of data in contexts other than search. Neuman and coauthors demonstrate that data profiles provided by data brokers improve accuracy in identifying a user with a particular attribute by up to 77% relative to random selection, with substantial heterogeneity in effect size across data brokers.[89] Another general study of the value of consumer data is provided by Azevedo and coauthors, who establish a theoretical foundation for slowly decaying returns from user-generated data to scale for firms running large-scale experiments, such as A/B testing.[90] One study of data's value in the content recommendation context is Claussen and coauthors, who compare the quality of algorithmic and editorial news recommendations.[91] They find that the length of

86. Di He, Aadharsh Kannan, Tie-Yan Liu, R. Preston McAfee, Tao Qin & Justin M. Rao, *Scale Effects in Web Search* (Int'l Conf. on Web & Internet Econ. No. 294, 2017).

87. Hema Yoganarasimhan, *Search Personalization Using Machine Learning*, 66 MGMT. SCI. 1045 (2020).

88. Maximilian Schaefer & Geza Sapi, *Learning from Data and* Network *Effects: The Example of Internet Search* (DIW Berlin, Discussion Paper No. 1894, 2020), https://www.diw.de/documents/publikationen/73/diw_01.c.798442.de/dp1894.pdf [https://perma.cc/5F3X-AG6B].

89. Nico Neuman, Catherine E. Tucker & Timothy Whitfield, *Frontiers: How* Effective *is Third-Party Consumer Profiling? Evidence from Field Studies*, 38 MKTG. SCI. 918 (2019).

90. Eduardo M. Azevedo, Alex Deng, José Luis Montiel Olea, Justin Rao & E. Glen Weyl, *A/B Testing with Fat Tails*, 128 J. POL. ECON. 4614 (2020).

91. Jörg Claussen, Christian Peukert & Ananya Sen, *The Editor vs. the Algorithm: Targeting, Data and Externalities in Online News* (CESifo, Working Paper No. 8012, 2019), https://www.cesifo.org/en/publications/2019/working-paper/editor-vs-algorithm-returns-data-and-externalities-online-news [https://perma.cc/J3SR-NBC7].

personal data records leads the algorithm to outperform the editorial news recommendation in terms of user engagement.

Although there is clear and mounting evidence that user-generated data improve the quality of online services, a lack of data about this relationship has limited researchers' efforts to estimate the pecuniary value of web users' data. One area in which such efforts have been fruitful is digital advertising, in which the relative prices of targeted and non-targeted (or contextual) advertisements are informative about the value of consumer information to advertisers. Johnson, Shriver, and Du provide one of the rare studies measuring the price differences between these sorts of advertisements and find that ads served to users who opt out of behavioral targeting—that is, ads based on past browsing behavior—yield fifty-two percent less revenue than ads shown to consumers who do not opt out.[92] An earlier study by Beales and Eisenach corroborates a similar average effect size with substantial heterogeneity depending on the age of the cookie; older cookies, which convey more information, increase the price of advertisements displayed to the user.[93] The authors find that the addition of a ninety-day-old cookie increases the price of an advertisement displayed to the user by 200% relative to the mean price in their data.

Targeted advertisements seem to be valuable according to Ravichandran and Korula,[94] Eisenach,[95] and Johnson and coauthors.[96] However, there is reason to believe that publishers do not receive a meaningful share of this value. For example, the main Dutch national public broadcaster completely abolished the use of targeted advertisements in January 2020, replacing these with fully contextual ads. The profits this publisher received from advertising increased because this decision enabled the publisher to cut all payments to companies in the "ad tech stack" (demand-side platforms, supply-side platforms, etc.).[97] A data market regime that reduces the market power of large platforms and establishes competitive markets is likely to benefit high-quality publishers.

The value of user data can also be gauged by analyzing revenues of firms mainly engaged in the extraction and monetization of user-specific data. The Interactive Advertisement Bureau (IAB) reports that spending on internet advertising in the United States reached $140 billion ($426 per capita) in 2020 with a year-over-year increase of 12.2%.[98] According to the *Los Angeles Times*,

92. Garrett A. Johnson, Scott K. Shriver & Shaoyin Du, *Consumer Privacy Choice in Online Advertising: Who Opts Out and at What Cost to Industry?*, 39 MKTG. SCI. 33 (2019).

93. Howard Beales & Jeffrey A. Eisenach, *An Empirical Analysis of the Value of Information Sharing in the Market for Online Content*, NAVIGANT ECON. (Apr. 8, 2014), https://ssrn.com/abstract=2421405 [https://perma.cc/7ZSA-SWFV].

94. Deepak Ravichandran & Nitish Korula, *Effect of Disabling Third-Party Cookies on Publisher Revenue*, GOOGLE (Aug. 27, 2019).

95. Beales & Eisenach, *supra* note 93.

96. Johnson, Shriver & Du, *supra* note 92.

97. Gilad Edelman, *Can Killing Cookies Save Journalism?*, WIRED (Aug. 5, 2020, 7:00 AM), https://www.wired.com/story/can-killing-cookies-save-journalism [https://perma.cc/YR45-55JQ].

98. *Internet* Advertising *Revenue Report*, PWC & IAB 3 (Apr. 2021), https://s3.amazonaws.com/media.mediapost.com/uploads/InternetAdvertisingRevenueReportApril2021.pdf [https://perma.cc/9PZS-M22D].

the data brokerage industry is thought to be worth around $200 billion ($526 per capita) as of November 2019.[99]

A recent court settlement provides further indication about the valuation of user data resulting from a bargaining process: A class action lawsuit filed against Facebook for illegally storing the biometric information of its users resulted in a settlement worth between $200–$400 for every affected Facebook user in the state of Illinois.[100] Facebook is also currently facing a fifteen billion dollar probe for illegally tracking and selling user data.[101] While the exact value of user data is context-specific and likely to depend on the socio-demographic background of the user, the above figures suggest a sizeable and one-sided appropriation of the surplus generated from user data.

Table 1 reports back-of-the-envelope estimates of the value of data based on setting-specific valuations of data. It suggests that data is of considerable value to web services, even when computed per capita.

99. David Lazarus, *Shadowy Data* Brokers *Make the Most of their Invisibility Cloak*, L.A. TIMES (Nov. 5, 2019, 5:00 AM PT), https://www.latimes.com/business/story/2019-11-05/column-data-brokers [https://perma.cc/XN55-V9TZ].

100. Kim Lyons, *Judge Approves $650 million* Facebook *Privacy Settlement over Facial Recognition Feature*, THE VERGE (Feb. 27, 2021, 12:09 PM EST), https://www.theverge.com/2021/2/27/22304618/judge-approves-facebook-privacy-settlement-illinois-facial-recognition [https://perma.cc/V389-KKSV].

101. Andrew Chung, *U.S. Supreme Court Rebuffs Facebook Appeal in User Tracking Lawsuit*, REUTERS (Mar. 22, 2021, 9:48 AM), https://www.reuters.com/article/us-usa-court-facebook-idUSKBN2BE1TX [https://perma.cc/H6DD-9ZDA].

Table 1. Back-of-the-Envelope Estimates of the Value of Data[102]

| **Setting** | **Year** | **Sources** | **US, total** | **US, p.c.** | **Global, total** | **Global, p.c.** |
|---|---|---|---|---|---|---|
| US online ad spending | 2020 | a | $140B* | $424 | $567B | $72 |
| Global digital ad spending | 2020 | b | $93B | $282 | $378B* | $48 |
| US data brokerage industry | 2020 | c | $200B* | $606 | $810B | $102 |
| Google ad revenue | 2020 | d, e, f | $49B | $148* | $147B* | $19 |
| Amazon ad revenue | 2020 | g | $16B* | $49 | $65B | $8 |
| Twitter ad revenue | 2020–2021 | h, i, j, k | $7B | $21* | $134B | $17* |
| Facebook ad revenue | 2020 | l, m, n | $52B | $157* | $253B | $32* |
| Facebook Study Project | 2019 | o | $40–79B | $120–240* | $162–320B | $20–40 |

102. All figures are in U.S. dollars. The "total" columns report aggregate valuations of data for the respective geography, whereas the "p.c." columns report per capita valuations. Where information on the size of the user base of the web service is available, we use this user base size to compute per capita values and obtain the total valuation by multiplying the per capita values by the size of the population of the geography. If there is no information about the size of the relevant user base for a setting on which we base a row of the table, and the source refers to the total over a geographical entity (United States or global), then we use this value for the total valuation, and we compute the per capita value by dividing the total valuation by the population of this geographical entity.

We use a population of 330 million for the United States, and a population of 7.9 billion for the world. We use the ratio between the nominal global GDP expressed in U.S. dollars and the nominal U.S. GDP to convert the total values between different geographies. The nominal world GDP is $85 trillion, and the nominal U.S. GDP is $21 trillion. The resulting ratio is 4.04. *GDP (current US$)*, WORLD BANK (2020), https://data.worldbank.org/indicator/NY.GDP.MKTP.CD [https://perma.cc/R4B9-YW7J].

An asterisk denotes a value drawn directly from the source on which we base the entry's row of the table.

The date in the "Year" column indicates the year corresponding to each data valuation.

The "Sources" column reports the labels of the sources on which we base our valuations. Table 2 provides the source corresponding to each alphabetic label.

All rows with a setting labeled "ad revenue" contain data valuations that reflect the value of data as used in advertising only. We report valuations based on the overall revenue of web services' digital advertising operations. We acknowledge the corresponding failure of these estimates to exclude revenue from purely contextual advertisements that do not rely on consumer data.

Table 2. Sources

| | **Data** | **Source** |
|---|---|---|
| a | US online ad spending in 2020: $140B | *Outlook 2022: The US Digital Advertising Ecosystem*, PWC & IAB 4 (Oct. 2021), https://www.iab.com/wp-content/uploads/2021/10/IAB-PWC-Outlook-2022-The-Digital-Advg-Ecosystem-Oct-2021.pdf [https://perma.cc/564L-7N5C]. |
| b | Global digital ad spending in 2020: $378B | Ethan Cramer-Flood, *Worldwide Digital Ad Spending 2021*, INSIDER INTELLIGENCE (Apr. 29, 2021), https://www.insiderintelligence.com/content/worldwide-digital-ad-spending-2021 [https://perma.cc/WAW4-DAFL]. |
| c | US data brokerage industry in 2020: $200B | Davad Lazarus, *Column: Shadowy data brokers make the most of their invisibility cloak*, L.A. TIMES (Nov. 5, 2019), https://www.latimes.com/business/story/2019-11-05/column-data-brokers [https://perma.cc/9KBF-ZHYM]. |
| d | Google U.S. ad revenue in 2020: $40B | Julia Faria, *Google ad revenue in the U.S. 2019-2024*, STATISTA (Jan. 6, 2023), https://www.statista.com/statistics/469821/google-annual-ad-revenue-usa [https://perma.cc/748V-WVQ2]. |
| e | Google unique visitors in The United States in 2020: 270M | Tiago Bianchi, *Google – Statistics & Facts*, STATISTA (Jan. 3, 2023), https://www.statista.com/topics/1001/google/#topicOverview [https://perma.cc/4AC4-PNS4]. |
| f | Google global ad revenue in 2020: $147B | Alphabet Inc., Annual Report, 33 (Form 10-K) (Feb. 2, 2021). |
| g | Amazon U.S. ad revenue in 2020: $16B | Alexandra Bruell, *Amazon Surpasses 10% of U.S. Digital Ad Market Share*, WALL ST. J. (Apr. 6, 2021), https://www.wsj.com/articles/amazon-surpasses-10-of-u-s-digital-ad-market-share-11617703200 [https://perma.cc/2Z2P-WTFS]. |
| h | Twitter U.S. ad revenue in 2020: $1.6B | Blake Droesch, *Why Have We Raised Our Twitter Forecast?*, INSIDER INTELLIGENCE (Feb. 22, 2019), https://www.insiderintelligence.com/content/will-the-twitterpurge-bolster-ad-growth [https://perma.cc/DPE6-ECP4]. |
| i | Twitter active U.S. users in 2021: 77.8M | *Twitter Statistics and Trends*, DATAREPORTAL (Aug. 15, 2022), https://datareportal.com/essential-twitter-stats?rq=twitter [https://perma.cc/U5HX-MF52]. |
| j | Twitter global ad revenue in 2020: $3.2B | Twitter, Inc., Annual Report (Form 10-K) (Feb. 17, 2021). |

| | **Data** | **Source** |
|---|---|---|
| k | Twitter global users in 2020: 187M | *Twitter Statistics and Trends*, DATAREPORTAL (Aug. 15, 2022), https://datareportal.com/essential-twitter-stats?rq=twitter [https://perma.cc/VDK2-TZAQ]. |
| l | Facebook U.S.& Canada ad revenue in 2020: $40B | S. Dixon, *Facebook: quarterly revenue in the U.S. and Canada 2010-2021, by segment*, STATISTA (May 2, 2022), https://www.statista.com/statistics/223280/facebooks-quarterly-revenue-in-the-us-and-canada-by-segment [https://perma.cc/Q8YJ-AMBD]. |
| m | Facebook U.S. & Canada monthly active users in 2020: 255M | *FB Earnings Presentation Q3 2021*, FACEBOOK, https://s21.q4cdn.com/399680738/files/doc_financials/2021/q3/FB-Earnings-Presentation-Q3-2021.pdf [https://perma.cc/76CR-DSYM]. |
| n | Facebook average revenue by user in 2020: $32 | S. Dixon, *Meta: average revenue per user 2011-2021*, STATISTA (Dec. 16, 2022), https://www.statista.com/statistics/234056/facebooks-average-advertising-revenue-per-user [https://perma.cc/V288-DYPW]. |
| o | Facebook study project (2019): monthly compensation of $10–20 per month | Josh Constine, *Facebook's New Study app pays adults for data after teen scandal*, TECHCRUNCH (June 11, 2019), https://techcrunch.com/2019/06/11/study-from-facebook [https://perma.cc/6JZ4-7EL7]. |