Yale
Tobin Center for
Economic Policy

DIGITAL REGULATION PROJECT

Working Paper:

Market Design for Personal Data

Policy Discussion Paper No. 6†
April 19, 2022

† The Tobin Center for Economic Policy at Yale hosts the papers of the Digital Regulation Project as a way for some of the world's leading economists and regulatory experts to present policy recommendations, based on their relevant research and expertise. The Tobin Center does not take policy positions and therefore the content does not represent the positions of the Tobin Center or Yale University.

MARKET DESIGN FOR PERSONAL DATA[1]

Authors:[2, 3, 4]

Katja Seim, Yale University
Dirk Bergemann, Yale University
Jacques Crémer, Toulouse School of Economics
David Dinielli, Yale University
Carl-Christian Groh, Yale University
Paul Heidhues, DICE, Heinrich-Heine University Düsseldorf
Maximillian Schaefer, Yale University
Monika Schnitzer, Ludwig-Maximilians-University Munich
Fiona M. Scott Morton, Yale University
Michael Sullivan, Yale University

[1] This is the sixth in a series of papers prepared by a collection of economists and policy experts in the United States, the UK, and the European Union who have studied, and are committed to the improvement of, competition in digital markets. Previous papers addressed consumer protection in online markets, regulating the market for general search services, the concepts of "fairness" and "contestability" as used in the Digital Markets Act, the use of "equitable interoperability" as a "super tool" to restore and encourage competition in online markets, and coherence between US and European approaches to digital regulation.

[2] Authors' complete affiliations and disclosures are set forth in Appendix VII.

[3] Many thanks to Brian O'Kelley and Alissa Cooper for helpful discussions that contributed to the paper.
[4] Omidyar Network, the James S. and James L. Knight Foundation, and the Alfred P. Sloan Foundation have provided funding and other support for this paper and other papers relating to regulation of digital platforms.

Table of Contents

1. Introduction

A. The problem

It now is generally understood that personal data, i.e., data that relate to individual consumers, drive digital markets. Personal data underlie targeted adverting, 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"[5] – to hone and improve them, and to recommend related products and services. The 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 researchers to better understand how humans interact with algorithms, and which policies might best serve society. And the data also facilitate competition, by improving quality, as well as 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.[6]

But there also are downsides to the collection and use of personal data on such a grand scale. "Surveillance capitalism," as Prof. Shoshana Zubroff has termed it,[7] 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 oftentimes face little to no accountability for their misuse.

Most reactions and proposed responses to the current state of affairs 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, maintenance of a stranglehold over data by a handful of large firms facing weak competition, 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 barter). This exchange stands in sharp contrast to what we see in other markets, in which those who control resources are paid for their extraction or use.
- We see a handful of firms controlling vast swathes of personal data that have significant market power. One way this market power manifests itself is in

[5] 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.
[6] *See generally* Dirk Bergemann & Alessandro Bonatti, *The Economics of Social Data: An Introduction* (March 26, 2019), COWLES FOUNDATION DISCUSSION PAPER NO. 2171, https://ssrn.com/abstract=3360352.
[7] *See generally* Shoshana Zubroff, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019).

lower quality services, including the collection of personal data without effective user control, and the use of that data to extract 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 by which to share data for valuable research or public benefit (transportation planning, or prevention of technology addictions, 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 the of market or even monopoly power.[8,9]

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 the data to benefit themselves rather than using them to advance the users' interests. Covid exposure tracking apps, which faced significant headwinds in the US especially, are a key example.[10] 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 profit in the large platforms, quality or control rights below competitive levels, and also for 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. 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. We also would expect the platforms that facilitate digital advertising, if the market were functioning as it should, to share the surplus with publishers that supply digital ad space (through higher pass-through rates of the total ad spend) and with advertisers (through lower ad prices). Firms that are hoarding data that could benefit other suppliers of services, public or private, would have incentives to share the data at reasonable prices.

---

[8] See Scott Morton and Dinielli (2020a, 2020b) for a more detailed discussion of the economic forces driving market power in digital markets.

[9] 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 European Commission, 2018). The Digital Market Act and Digital Service Act (see European Commission 2020a, 2020b) are explicitly targeted towards dominant firms defined as digital gatekeepers. Similarly, the United States House Judiciary Subcommittee on Antitrust, Commercial and Administrative Law is currently advancing a bill to curb the monopoly power of large tech companies (see House Bill 3849, 2021).

[10] *See, e.g.*, Jessica Rich, How our outdated privacy laws doomed contact tracing apps, BROOKINGS INST., January 28, 2021, https://www.brookings.edu/blog/techtank/2021/01/28/how-our-outdated-privacy-laws-doomed-contact-tracing-apps/.

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, and innovation, and also about 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. With our proposal, 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 (this failure contributes to a status quo that threatens user privacy, lowers the effective quality of online services, and also facilitates market power);
>
> (2) the failure to provide users a way to derive financial benefit from the data they generate (this failure enforces a status quo that distributes surplus from digital markets to platforms rather than consumers, and also 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 (this failure generates a status quo that implicates or even impedes innovation).

Moreover, the three problems appear to be related. The lack of effective privacy regulation or other restrictions on data collection allows the 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 deriving benefit 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 or design competing products or prepare for seamless interoperability, for example. The exclusionary approach also frustrates legitimate requests from the government or from academic researchers (the large platforms sometimes claim the data themselves constitute trade secrets and can't be disclosed, or they assert that disclosure would undermine user privacy), the platforms 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 paper explores a possible intervention that – unlike antitrust enforcement alone or enhanced privacy regulation alone – would address competition, *and* efficiency, *and* privacy concerns directly and at the same time.[11] We offer the idea of a "data intermediary," bound by

[11] 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 CMA, 2019 and Scott Morton et al., 2021). 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 needn't allow the interoperating firm to do whatever it wants with personal data to which it gains access. The interoperating firm only gets to do what the large platform gets to do; there are no new or additional data uses that could raise

fiduciary duties to users, and 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.[12]

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 US, 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 admit of such rights.[13] 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.[14] 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.[15] 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 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.

i. Our proposal governs the collection and use of all personal data

Our proposal therefore offers something additional and complementary to proposals that focus narrowly on enhancing traditional privacy protections. Private data can become non-private in an instant, thereby losing certain protections reserved only for private data. And much of the data whose extraction and use causes consternation is not considered private in the first instance.

The privacy-based proposals, if tethered to traditional understandings of privacy, often lack footing to dictate or limit the *use* of data that once were private but now are not. Equally

---

"privacy" concerns. In a recent theoretical study, Argenziano and Bonatti (2021) highlight the potential consumer welfare benefits of *data linkages*, i.e. data-sharing relationships between firms. Thus, interoperability requirements could lead to welfare benefits.

[12] 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 Posner and Weyl (2019).

[13] A case pending in Illinois state court deserves attention in this regard. A Muslim man downloaded and used an app called Muslim Pro to remind himself when and in which direction to pray. In apparent violation of a contratuctual promise, the app is alleged to have sold its users personal location data *en masse* to a data broker, who then sold it to the US Department of Defense. The plaintiff advances a number of claims against the app, two of which would, to be successful, depend on acknowledging that users have an initial right to control location information, and perhaps even control its monetization. The first is a common law unjust enrichment claim that advances the theory that the app, in arrogating to itself the decision to sell the data and to reap the benefits of the sale, acted unjustly. The second is that the app's conduct constituted misappropriation under Illinois's statute governing "trade secrets" – a category of information that derives value from its not being generally known, and whose disclosure or use by others (the app in this instance), is limited by the first party's expresss instructions, as well as principles of good faith and fealty to the public interest. *See also* Amy Kapczynski,*The Public History of Trade Secrets*, 55 UCD L. REV. 1367, 1380-90 (2022).

[14] *See* GDPR Art. 4, Definitions ("'processing' means any operation or set of operations which is performed on personal data or on sets of personal data . . . such as . . . use . . . .").

[15] *See* GDPR Art. 6.1.a.

important, privacy-based proposals, unless they rely on substantially expanded notions of what data qualify as private, have little to say about the collection or use of data that never would have been considered private in the first place (e.g., location data reflecting a two-block walk to a popular café on a busy city street). Expansive collection and use of various forms of nonprivate data, however, underlie some of the most urgent concerns with data-driven products and markets. Nonprivate data, including much location information, can be highly valuable.

ii. We postulate a control right that includes a right to be paid for data use

This is why our proposal postulates 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 General Data Protection Regulation ("GDPR'), for example, acknowledges that individuals have a continuing right to exercise some control over the use of data that relate to them. But 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 US, where individuals have few rights in respect of 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.

iii. 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 paper below traces through the economic issues that arise in creating such a system.

Creating and maintaining the institutions, procedures, and oversight necessary to permit users to exercise the new proposed control right in a manner that would yield the desired competition and privacy benefits are 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 or 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 the 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 paper 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 paper 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: the ability to limit collection and use and the ability to be paid, and that solves for the three interrelated market failures that also are 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 her personal data, or even to negotiate payment schedules with web services that might use her 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.[16] 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.

iv. 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 data intermediaries. 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

[16] 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 ad 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 her anything close to that amount, or indeed anything at all.

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 enhance 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.

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 her degree of comfort or discomfort with the collection and use of her 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 more than a thousand dollars per year per person on digital advertising that uses personal data for targeting, and the rate of growth of these revenues is high. 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.

v. 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 with greater control over their data within the existing regulatory framework.[17] 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. 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; the users choose what to put in and what to let out (and to whom and for what purpose). Our own proposal 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 policy makers 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.[18] Even if a perfect regulator followed all the suggestions in this paper, 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

[17] See EDPS (2016) and EDPS (2020) for the European Union, CDEI (2021) for the UK Centre for Data Ethics and Innovation, and the "Data Freedom Act" for the proposal of RadicalxChange (2020).

[18] 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 paper. 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 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 paper.

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.

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 paper 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 behavior of users is the focus of the next section, followed by a detailed discussion of how the purpose of using the data fits into the regulation. The ways in which the regulator can enhance competition follows, and the paper concludes with issues of enforcement and future trends.

C. Summary list of policies

As mentioned, this paper 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, will create incentives that will encourage conduct by actors in digital markets that will bring about the outcomes we desire. The underlying laws that will be needed to create a market for data will 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[19]

- Each data intermediary is required by statute to act in the fiduciary interests of its users.
- Each data intermediary also is directed to comply with data minimization principles,[20] balancing this goal with the goal of monetizing user data and creating datasets that are valuable to consumers and society.
- Data intermediaries are licensed and have strict regulatory requirements in terms of data/cybersecruity and resilience.
- Data intermediaries are independent and may not vertically integrated with any other business whose products or services relate to the use of personal data. They cannot 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 contracts with just one data intermediary. That intermediary serves as the user's exclusive agent for purposes of monetizing the user's personal data generated through the use of any of her devices or web-connected products comprising the Internet of Things (IoT). Requiring that each user have only one intermediary at any giveb time enables 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 her data avaiilable for research ad other beneficial uses.
- Intermediaries compete for users by offering users a share of revenue in exchange for monetizing users' data (with a minimum set by regulator as a percentage of revenue). The cash value of that share will 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.

[19] 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.

[20] "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* European Data Protection Supervisor, Glossary, at "Data minimization," https://edps.europa.eu/data-protection/data-protection/glossary/d_en.

- Intermediaries also 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 and intermediary, the chosen intermediary collects all online data for that user by monitoring browser and app use. The intermediary can 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 bv each intermediary will reflect its balancing the obligation to adhere to principles of data minimization, on one hand, with 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 sells access to its users based on this personal data. In order to develop cohorts for targeting of display ads it can carry out its own data analytics across the data. The use of cohorts helps to protect individual privacy.
- Intermediaries also 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 may collect personal data they need to provide their service, if used solely for this purpose, and can choose the 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 can carry out data analytics across user data, so far as this is needed for that purpose. They cannot share data with third parties (or provide services to third parties based 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 will develop rules about how far in advance of "use" 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 cannot transact in personal data except with a licensed data intermediary.
- Data intermediaries may 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 the users.

User Interface

- Intermediaries 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 encourages users to sign up to a data intermediary using nudges, defaults, and most importantly, the offer of payments. The regulator may develop additional methods to encourage participation including public education campaigns to dispell misinformation.
- Because data intermediaries compete for users on the basis of payments and services, the regulator designs 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 will share raw data from a particular web service with third parties (e.g. a rival webs service). This feature enables 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.

2. 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 will experience choice fatigue and overload, while each web service will 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.[21] Consequently, a firm's valuation of the marginal web user's data is typically low, and web user might not get much from bargaining with the firm even if bargaining between individual consumers and firms were feasible. This outcome would obtain even if the web service's value from data averaged across web users were high. If data are "social" in this way, consumers will benefit from joining together in a group to monetize their data. An organization representing many web users will 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.

[21] *See* Dirk Bergemann, Alessandro Bobatti, Tan Gan, *The Economics of Social Data*, https://www.mit.edu/~bonatti/social.pdf (Jube 23, 2021); Jay Pil Choi, Doh-Shin Jeon, and Byung-Cheol Kim, *Privacy and Personal Data Collection with Information Externalities*, 173 J. PUB. ECON. 113 (2019).

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 in creating control rights for users and a system in which they can express their preferences – share everything, share nothing – will be required for any change. Those control rights will include permitting licensed intermediaries to collect and monetize personal data in accordance with user permissions; reserving the right use the internet anonymously without any collection of personal data collection; prohibit particular uses of data as the regulator learns about possible harms; and prohibiting any firm or person other than a licensed intermediary from monetizing personal data or purporting to exercise any control over their use.

For all these reasons the notion of an intermediary is popular among policy makers. 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.[22] 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 (2019) (and the subsequent and related proposal by a nonprofit founded by Glenn Weyl, see RadicalxChange (2020)) are designed to give users monetary compensation for their data. There 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.[23] The document 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 activities 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.[24] We recommend that data intermediaries be licensed by a regulator and adhere to the principle of data minimization. This will 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 for a particular product, for example). It also permits 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.

---

[22] See the proposal for a European Data Governance Act (European Commission, 2020c), https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113.

[23] One example from the 1990s is NetZero, an internet service provider tha 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 here: http://www.cnn.com/TECH/computing/9810/26/lastfree.idg/index.html. See also the discussion of Permission.io in Section 2.

[24] 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 to which firms with such concerns could name under oath their principal competitors and

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."

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 she cannot be identified, than it is with information that would permit the advertiser to learn her 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 will create the right incentives for sharing of 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 by young woman worried that her brother may be drinking too much, or information about which of the suggested articles or ads she clicked on. 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, may issue rules further delineating this distinction.

An important concept going forward will be 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 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 will be the agent of the consumer. How will the intermediary both safeguard and monetize the consumers' personal data? After a user chooses an intermediary and instructs it on what to share, and technical steps have been carried out, the intermediary can 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 will require the development of new technology for transmitting, pricing, and protecting data. The benefits of the intermediation regime for web users also depend on intermediaries' success in negotiating with web services over the surplus generated by users' data; intermediaries earn the revenue from their users' data and are therefore incentivized to negotiate effectively with web services. We discuss payments to users below.

The intermediaries' monetization of personal data will mostly relate to digital advertising. In the bargaining between intermediaries and web services selling advertising it is important to understand the outside option of both sides. The system is set up so that users single-home; they enroll with one intermediary and therefore advertisers face a monopoly seller of that user's data. If the advertiser does not contract with the intermediary, it cannot 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 allow 'first parties' to utilize the data they collect via their own web service in order to provide that service. As such, Google can use such data in a search query to return organic search results and to train its algorithm. However, under the proposed rules, Google cannot 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 must 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 determines that the information is in a category that user wishes to share, establishes a price through some mechanism, and collects the funds from the web service. If Google and the intermediary cannot agree on a price for the user's search query, Google may serve ads that do not require individual data -- such as an ad for Insomnia Cookies if it is 2am. In that instance the intermediary does not collect any revenue from Google for the data.

The intermediary is also 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 may not retain the personal data query for future use.

A query that is necessary to carry out the service desired by the consumer does 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 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 will – likely through an SSP[25] – query the intermediary to determine which users qualify. The answer determines whether or not the user falls in the category the advertiser wants, in which case she sees the ad. If the intermediary and the SSP cannot agree on the price of the data, the SSP may turn to a rival intermediary and seek data from its cohorts of (different) users. Again, however, because users single-home, an intermediary is the monopoly seller of its own users and has significant bargaining power. At all times the publisher could choose to sell a contextual ad,[26] rather than a targeted ad, and not use any personal data.

Under our regime, intermediaries are responsible for developing, contracting, or partnering to participate in programmatic ad auctions. An intermediary may choose, for instance, to group its clients into cohorts based on interests, demographics, or locations (e.g., young male bicyclists in the US Northeast) and accept bids that condition on cohort membership. Firms that specialize in designing profitable cohorts or effective algorithms may 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 are authorized by that user's intermediary.

There are also many imaginable structures for the publishers' payment. To fix ideas, one possible structure would involve the publisher paying a commission charge to the intermediary that is proportional to its revenue from the auction. Another would feature a flat

[25] An SSP, or Supply Side Platform, 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 the about the "eyeballs" that could see it (male is his 20s within 100 feet of a donut shop at 8:30 am.) – for auction through an ad exchange.
[26] 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 US 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.

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 on any of these dimensions.

One primary function of our intermediaries is to consolidate the source 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 et al. (2021), advertising channels whose viewers are difficult to reach through a particular medium (e.g., television or the internet) fetch higher ad prices. 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 have rich data on their users, they will be able to identify users who single home and on which web services. This information will facilitate the sale of ad impressions at high prices. Additionally, a data intermediary will exclusively control targeted advertising to a particular web user because users single home at intermediaries. The theoretical and empirical findings highlighted by Gentzkow et al. (2021) suggest that our intermediary system will generate higher ad prices for intermediaries, most of which will be passed back to web users via competition between intermediaries.

Intermediaries will 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 can 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 the product or service being development. Intermediaries will 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 will 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 will represent millions of users, their ability to facilitate arrangements for access to large data sets will 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 depends on what sort of data she is willing to share. Data can be arrayed from least valuable to most valuable from the perspective of an advertiser. Likewise, data can be arrayed from least costly to share in terms of privacy and intrusiveness to most costly. A market mechanism helps 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, each of which may have different interests and a different business model, the process of determining what is efficient to share can be complex.

A major problem a consumer faces when trying to control the collection and use of her data in the status quo is that user agreements are impenetrable. They are long, contain legal concepts and terminology and are not realistically understandable by regular users in 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 she has 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 establishes a standardized menu of data-share levels. The standardization of levels allows 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 the metal tiers —e.g., "Bronze," "Silver," etc.—for health insurance policies. A regulator will require that data intermediaries offer these tiers to their clients. The regulator designs the tiers to represent average user views on what constitutes decreasing levels of privacy. Users then choose how much data to share in a standardized environment designed by a regulator to be understandable and clear.

Each data-share level will be a distinct collection of types of data a consumer would allow the intermediary to monetize. Web services will be able to contract to use these data as will be described below. Web services will not be permitted to use any data outside the categories selected by the consumer. Of course, as noted above, a web service may 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 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 will be with the web service that collected the data but must pay for the right to use it for any purpose other than performing or improving the service the user requested (first-party requests) but that the bulk will 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 paper.

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 1also 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 (2), 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," i.e., the ability to limit the number of times a user sees a particular advertisement. A user's browsing data will often reveal 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 will generate 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 e-mail receipts. While financial records are sensitive, our justification for including this category is that attribution is highly valued in digital advertising, and thus it is 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 will 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 will need to develop a plan for supporting this level of data sharing. Web services will 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 will be important in the negotiation between the web service and the intermediary because it is the outside option if negotiations over the price to pay for data break down.

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 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 aware of studies or data that reveal consumers' actual relative preferences and discomfort with the various types of data collections and uses with certainty, and so these

proposed tiers reflect our best effort to create tiers of increasing "invasiveness" and "prospects for monetization."[27] A regulator would consult experts in the design of these levels and 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 is 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 that society might wish to exclude from data markets include: political party affiliation, trade union affiliation, sexual orientation and sexual history, religious belief, addictions, and health information more generally. These characteristics are similar to the categories sensitive data under the GDPR and the CPRA.[28] Relatedly, the regulator may 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.

The rules will also provide "allowances" of a certain number of hours during which a consumer may browse anonymously, regardless of which data sharing level she has selected. In that case the consumer's data intermediary will not transmit any of the consumer's data to any web service and there will be no collection of the consumer's data by any data intermediary. This feature allows 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.

---

[27] Cite to recent study Amelia circulated as among the most detailed info on how consumers value collection/use of various kinds of data.

[28] See the following webpage for the GDPR sensitive data categories: 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. See Section 1798.140 of the CPRA for a list of its sensitive data categories and the following webpage for a comparison of the protection of sensitive data under the GDPR and the CPRA: https://iapp.org/news/a/new-categories-new-rights-the-cpras-opt-out-provision-for-sensitive-data/. Note that the CPRA considers the contents of a users' messages to be sensitive information.

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 provides 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 let the intermediary make different payments to web users who choose different data-share levels; it also allows 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 74% 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 lower risk but generate several problems. First, determining the exact value of any individual user's data is likely to be challenging given the presence of complementarities between the data of different users, e.g., 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 she is not interested in because providing data to these sites earns her a higher payment from the intermediary. Although it is possible that data intermediaries will be able to develop methods that detect devious web use and discourage or eliminate such use, it seems simpler to choose compensation scheme that does 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 will receive different levels of remuneration. Recall that we allow web users of any share level to use the internet anonymously, i.e., 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 her 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 will 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.

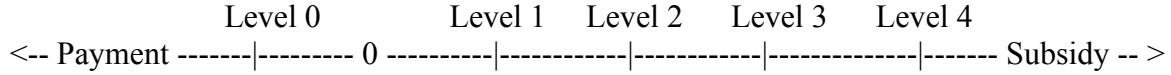
Irrespective of the chosen regulations for intermediaries' compensation schemes, competition between intermediaries is desirable in that it will help ensure that the surplus generated from web users' data is largely returned to web users.

We conclude this section by raising a tricky economic issue: web users who choose Level 0 will fail to generate revenue for their data intermediaries but will 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, by, e.g., 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.[29]

[29] It may seem incongruous that Level Zero users may have to pay a fee to their intermediaries, given that the control right we propose to recognize is defined to include 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*, *e.g.*, Matthew Menendez & Lauren-Brooke Eisen, *Report: The Steep Costs of Criminal Fees and Fines* (Executive Summary), BRENNAN CENTER FOR JUSTICE (Nov. 21, 2019), https://www.brennancenter.org/our-work/research-reports/steep-costs-criminal-justice-fees-and-fines (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 and higher-than-average rates. Intermediaries could offer a "no-fee" introductory rate, for example, which itself would spur competition among the intermediaries.

Payment scheme – illustration

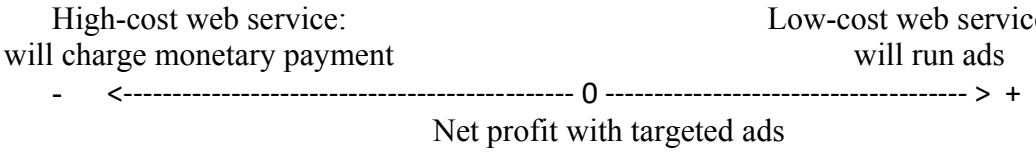
Payments to users



The optimal level of quality of a web service might be sustainable with contextual advertising, it might require the higher level of revenue generate 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.

Categories of web content – illustration

Remuneration of web services



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 funds with the intermediary due to sharing her personal data, there is no reason that user couldn't ask the intermediary to spend some of those funds to allow her to access websites that charge subscriptions. An intermediary can negotiate a price schedule with web services like newspapers. Because the intermediary reduces 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 will 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 will have to signal to users that they have arrived at a web service that will cost money and tell them the net price. The user can then decide to go ahead or not. In this way consumers will be offered choices to pay with money when there is available expensive content that cannot be bartered for. The price schedule available to users and the range of content covered by an intermediary will 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 will also provide a version of their service that is available without monetary or data charge.

The possibility of incorporating microsubscriptions may 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 will need to pay more in money. The regulator may want to permit intermediaries to charge a transaction fee, e.g., 5%, for payments made by Level 0 users. If this is clearly stated on the choice screen, intermediaries will compete on this dimension. But intermediaries will have a way to earn profit from Level 0 users and will not avoid serving them.

F. Competition among data intermediaries for consumers

Data intermediaries will compete for consumers on a number of dimensions. First and most obviously, consumers will be attracted by the level of remuneration offered by the intermediary. Higher payments (or payments in attractive forms) will attract more users who will 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* will attract users to intermediaries in the same way that higher or better forms of *payments* will attract users who choose to share more personal data. Intermediaries might even offer an introductory rate of zero fees for Level 0 for first year, if the user remains with the Intermediary for the following year.

Consumers will 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 e-commerce 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 its data to a large platform that may use the web user's data for other purposes.

Consumers may 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 will enter high-profile partnerships with research institutes that rely on personal data to design solutions to any number of social or health-related problems, or to increase voter participation in local elections, combat climate change, and so forth.

Intermediaries may engage in marketing to inform consumers about their attributes and benefits. A regulator will 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 mandates intermediaries to report certain types of information on their data dashboards—namely, which data has been shared, with whom has the data 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. Consumers surveyed by the Federal Trade Commission (2009) reported concerns about opaqueness in data collection and the disclosure of the use of their data. Providing consumers access to detailed reports about the use of their data should assuage these concerns.

The higher the per person revenue of the intermediary, the larger the remuneration it can offer. Intermediaries will therefore compete on the basis of innovation in monetizing user data. This may 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 will also have an incentive to invest in contracting and negotiation with web services as this will increase their revenue.

We are concerned that web services might discriminate against users selecting restrictive privacy levels. Under our regulatory regime, web services are 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, it may have an incentive to charge higher prices or offer lower service quality to users who select these levels. In a competitive market, data intermediaries will 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 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 will 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 on each of its users will have to interact with each intermediary used by that group. One of the dimensions on which intermediaries will compete is their ability to design efficient contracts and negotiate with web services. Larger web services will likely want unique bilaterally negotiated contracts with intermediaries. Small web services will 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.[30])

In order to understand the parties' bargaining positions and the extent to which contracts will 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, can 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.

[30] See here: https://get.doordash.com/en-us/products/marketplace.

3. The user interface

This section discusses how the user-facing aspects of the data intermediary regime can be designed to promote our proposal's goals. The behavioral economics literature suggests various reasons why the current regime based on informed consent of data sharing is unlikely to be exercised in a fruitful manner by consumers. We particularly emphasize the desirability of a clear and simple choice architecture in light of web users' cost of attention 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.[31] 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.[32] A 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.[33] Regulation of digital markets should take these behavioral aspects of decision-making into account by establishing defaults that are in line with consumer preferences, by ensuring that important information about alternatives is salient wherever consumers make decisions about their data, and by limiting the number of separate choices required of consumers.

Furthermore, the behavioral IO literature points out that profit-maximizing firms have an incentive to find that decision-frame, for which consumers are most likely to make the decision the firm prefers.[34] If, for example, consumers would tend to understand most ways of

[31] Johnson et al. (2012) cite numerous studies in behavioral economics and psychology providing evidence on the role of choice architecture (including defaults) in consumer decision making.
[32] Mathur et al. (2019) provide evidence of the use of dark patterns by commercial websites to influence users into disclosing information. They 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."
[33] Baumeister et al. (1998) define ego depletion as "temporary reduction in the self's capacity or willingness to engage in volitional action […] caused by prior exercise of volition," and they provide experimental evidence for this phenomenon. There is considerable evidence of ego depletion in behavioral economics. To provide some examples, Hirschleifer et al. (2019) find that the accuracy of financial analysts' forecasts decreases in the number of previous forecasts made in the same day; Persson et al. (2019) find that surgeons are less likely to schedule a patient for an operation late in their shifts; and Augenblick and Nicholson (2016) find evidence of decision fatigue in voting by exploiting variation in choices' positions on ballots caused by differences in the number of local ballot measures.
[34] For example, Blake et al. (2021) present evidence suggesting online vendors strategically choose 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. Similarly, the literature on add-on fees—see, for example, Brown et al. (2010) and Einav et al. (2015)—suggests that consumers do not rationally internalize shipping fees into products' overall prices when making purchasing decisions. Sellers can exploit these biases by lowering their advertised prices while adding shipping fees and/or other add-on fees.

presenting the choice correctly but due to naivete misunderstand one framing thereof, competition 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: 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 the consumer opposes but of which the web user remains unaware after accepting the terms. 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.[35]

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 section. Inconsistency between intermediaries in the language used to describe these concepts could further confuse consumers. Obstacles to consumer comprehension of the privacy levels, and of data intermediaries must accommodate limited consumer literacy. The most vulnerable consumers may have the most difficulty understanding the tradeoffs involved in 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 for the levels 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."[36]

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

---

[35] 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." See the following link for the full text of the Act: https://oag.ca.gov/system/files/initiatives/pdfs/19-0021A1%20%28Consumer%20Privacy%20-%20Version%203%29_1.pdf.

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

describe, and provide examples of, the risks associated with monetizing their data. Better choices will result from consumer understanding of both the benefits and costs of data sharing.[37]

Our data-share levels do 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 her level three data, but another service to access only her level two 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 a additional ability to personal 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 does not wish to share data with a particular web service may choose to access that web service anonymously under our proposal.

One criticism of our proposal is that behavioral consumers will overweight the tangible, short-run benefits of remuneration from data sharing relative to the risks of data sharing, which are uncertain and may only realize themselves 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 will help consumers understand the risks associated with each privacy level. Also, we view our proposal as an improvement over the status quo in this regard, as the status quo 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.[38] This is particularly concerning because malicious forms of targeting often prey on people of lower socioeconomic status (e.g., predatory loans).[39] 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

To fix ideas, we outline a possible mechanism by which data intermediaries could be adopted but recognize that this is only one possibility of many. We recommend that the regulator create a standardized application (we will refer to this as the Data Manager going forward) that will come pre-installed on devices with capabilities for accessing the internet. If this is part of US regulation, a developer of any operating system for consumer devices sold in the US 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 will present the user with a choice menu displaying the privacy levels, the standardized descriptions of each level,

[37] Behavioural Insights Team and Doteveryone (2020) suggest a series of principles for designing choice architecture that promotes users' wherewithal to make active choices. One of the principles is to make the trade-offs involved in a choice interactive, allowing the user to interact with, or experience, what the trade-off means.
[38] See Elvy (2017).
[39] Wills and Tatar (2012) showed the relevance of this concern for Facebook; in particular, they found that Facebook targeted ads to users on the basis of sensitive personal characteristics such as health status and sexual orientation.

and the dollar range of compensation for each level. Once the user has selected a privacy level, the Data Manager will 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 will, despite the compensation offered by intermediaries, not wish to engage with Data Manager and will instead quickly click through its choice menus. To protect these users, we suggest that the regulator specifies 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 will 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 will also voluntarily connect their older devices to intermediaries as information about intermediaries' benefits spreads. Nudges and incentives may be a better method of establishing this market than a mandate. If competition between intermediaries generates payments to consumers that are a significant share of digital advertising, payments will be many hundreds of dollars per person which will create a financial incentive to use a data intermediary.

Data intermediaries' offerings to advertisers become less valuable if their consumers *also* join competing intermediaries, as this nullifies the intermediary's status as a local monopolist over its consumers' data. If a user were paid her value, she would have a strong incentive to be exclusive with one intermediary. For this reason, an intermediary should be able to offer terms that apply 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 are likely to always be some users who do not belong to a data intermediary. Such a user may actively provide information to web services for the purposes of servicing the user, but the web service cannot monetize that personal data in exactly the same way it cannot monetize enrolled users' data. It is critical that these users are not profitable for web services in order to ensure that there is no incentive for platforms to discourage users from joining an intermediary. If users who wish to share their data earn significant payments from an intermediary and those who prefer limited sharing can manage their privacy more effectively through an intermediary, consumers will have incentives to enroll. Above we explain why we propose that new devices come with a default enrollment stage to raise participation.

In the case that the data of a web user who does not register with an intermediary is treated similarly to that of a web user who selects an intermediary's Level 0, it may 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 is especially true if the intermediary, which cannot monetize its Level 0 clients, charges 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 subsection, we discuss several of these services.

C. Switching among intermediaries

In order to incentivize intermediaries to compete to gain 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 will facilitate choice when a user first operates a new device. Likewise, the Data Manager will be the tool users employ to enroll in an intermediary each year. The regulator can intensify competition by choosing an 'open enrollment' period during which time users are 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 can develop the information and messaging for intermediaries to deliver to users to facilitate their choices. Users will be directed to the Data Manager where they can 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.[40] 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 can make a conservative choice such as privacy Level 2 and the highest cash price in that level. If multiple intermediaries have similar cash prices, permitting them all to be allocated a share of passive users will 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 will want to bring their web usage data with them when they switch intermediaries. To lower switching costs, the regulator will need to require that data intermediaries transfer raw consumer data in a standardized format upon a user request. The data format should be specified by the regulator, so it is 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

[40] Several papers highlight the relevance of consumer inertia for firm behavior. Ho, Hogan, and Scott Morton (2017) document that consumer inertia in the US healthcare market allows insurers to charge higher premiums. Hortacsu et al. (2017) document significant consumer inertia in the choice of electricity provider and estimate that information interventions can significantly raise consumer surplus. Grzybowski and Nicolle (2021) reveal that consumer inertia facilitates concentration in the smartphone industry. MacKay and Remer (2022) show that consumer inertia impacts the simulated price effects of a merger in the gasoline industry.

technologies that maximize the value of web users' data, we do 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 retains 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 must delete that user's data. If a user does not re-enroll with an intermediary, prior to deletion the intermediary must send the user their raw data in the standardized format, so the consumer controls 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 her laptop and a different intermediary for accessing the internet on her mobile phone. The consumer's data specific to any particular device will not be as rich as the pooled version. Ideally, this behavior should be prohibited because it makes the consumer's data less valuable and harms her financially. Additionally, the consumer's platform loses its position as a local monopolist over that particular consumer's data when the consumer uses different intermediaries on different devices. The policy response to this problem will 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) may 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" 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 is its ability to lower switching costs between web services. For example, if a user moves from one ecommerce site to another, the new site will 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 may not want to switch ecommerce sites, and, in turn, new entrants into ecommerce will be discouraged from entering. A consumer's data intermediary will 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 her data intermediary to send categories of data from her existing ecommerce site. Such a tool must be carefully overseen by the regulator because it is 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

4. Types of data use

In this section, 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 implies different restrictions in terms of data-access rights. Additionally, the treatment of data used for analytics bears 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 user's data, the privacy levels defined in the web user section therefore do not apply to data for servicing users. That is, web services may 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. [redacted] A regulator will inevitably need to investigate difficult cases and disallow unnecessary data collection. A regulator may 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, are exempted. Likewise, an exercise app that reminds the user to stretch or take a walk after periods of physical inactivity is included functionality. 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 will 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) are also 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 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*, is contextual. Drawing the line between contextual and personalized advertisement using the criterion of the involvement of a publisher creates 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 will 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 are permitted to freely collect first-party data on their own users for analytics and product improvement, then consumers will 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. (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 creates 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 non-exclusive manner, all entrants will be able to pay to learn about the market – through existing users. The regulator can promote entry by establishing rules that prevent exclusives and discrimination in the use of data for analytics.

5. 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 holds data intermediaries and web services to the principle of data minimization, which is one of the central principles of the EU'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 have a role to play in respecting data minimization; each type of agent is 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 can effectively respond to conduct by intermediaries that violates their fiduciary duties or other rules, we recommend that data intermediaries are required to hold a license from the regulator. A system of licensing would give the regulator the ability to create standards, strict data protection regulations, and responsibilities for intermediaries. Such a system also facilitates the punishment of intermediaries that violate these standards by revoking their licenses or levying less severe sanctions, e.g., fines.

Last, our proposed regulation holds 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.

An issue of concern is whether a web service will discriminate based on 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 will 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.[41]

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, 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 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 will 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 will 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 will also have poor incentives to engage in sophisticated analytics which enable it to provide effective targeted advertising, and this will in turn be bad for both advertisers and users. In general, all these markets will perform better if there is 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.

[41] Larson et al. (2015) find that households in high-income neighborhoods are systematically charged more for online tutoring packages. Scherer and Siddiq (2019) provide empirical evidence that socioeconomic status (SES) positively correlates with information and computer technology (ICT) 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 of how their data are used.

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 is 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, e.g., white men between 25 and 29 years old residing in the greater New York City area. To accurately estimate this quantity with an intermediary's data, the intermediary must have tracked a sufficient number of members of this group who 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 will reward an intermediary that serves a large share of a cohort. By internalizing the data externality, the users and the data intermediary will 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 are valuable on their own.

Most authors expect that differentiation among intermediaries will naturally limit concentration, although some have significant concerns about the tendency to tip. Intermediaries may contribute their profits to particular causes that attract a subset of users but not all. Intermediaries may specialize in serving certain types of users. That specialization may cause 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.[42] In general, avoiding vertical integration by the intermediary will prevent some competitive issues that could arise. Similarly, competition will be enhanced by a prohibition on exclusive contracts between intermediaries and any other parties such as consultants, DSPs,[43] 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.

6. 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 forgotten" as defined by the GDPR, which gives web users the right of instructing web services to delete their information about the user. The practical implementation of this right has turned out to be

[42] 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 Data Ethics Commission 2019, section 4.3.
[43] DSPs, or Demand Side Platforms, assist advertisers 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.

challenging. As a response, there are initiatives such as the data rights protocol that aim to facilitate such requests.[44] 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 with in cases where a right to be forgotten applies. Furthermore, we describe above certain data – 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 for which the regulator should impose time limitations on data access. If enforced, these too should ease implementation of a right to be forgotten in their effect is to reduce the number of possible data custodians over time.

B. Violations

Large web services will 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 in 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 condition 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 intermediaries with 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.

7. 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.

A. The Internet of Things

The internet of things – the "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 devices considered within the IoT 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

[44] Details of the data rights protocol can be found here under Data Rights Protocol (2021).

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 avoid inconvenient situations in which the consumer finds no more milk in his 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 paper. 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. (This problem may apply to personal computers and may be mitigated by software that facilitates switching between intermediation profiles on a particular device.)

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 her jar of chili crisp and her tup of white miso, the company presumably should be required to purchase that information.

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, i.e., 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.

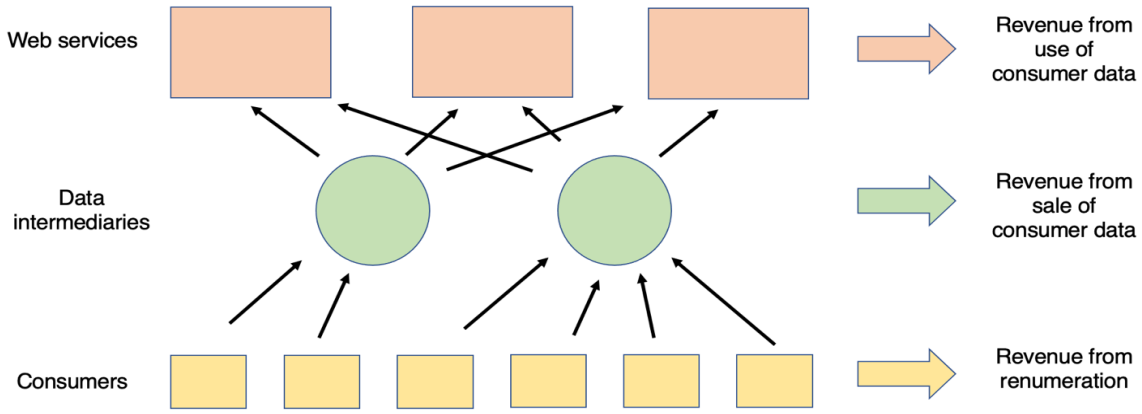
8. Conclusion

The above discussion demonstrates that establishing a regulatory environment for successful data markets is complex. We hope the ideas in this paper 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 a 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, or for other market participants, or even the citizenry who might benefit from innovative data uses. It is unlikely to be efficient either, as competitive remuneration for the content that draws users to the web will stimulate the amount and variety of content consumers want, while entry of web services that use personal data will be enabled when access to that data is available at competitive rates to both entrants and incumbents.

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 situation 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 mechanism to innovate in the prosocial uses of personal data. A market for personal data, properly constructed, will 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 also gives 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: Data intermediation schematic

Schematic identifying principal categories of participants in data intermediary markets



Appendix 2: 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 section, we review alternative ideas and proposals related to the regulation of web users' data.

Fukuyama et al. (2020) envision a regime in which consumers access digital platforms through third-party web services called middleware. 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 et al. 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 choice between privacy settings and other terms of service. Fukuyama et al.'s 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. Balkin (2016) and Balkin and Zittrain (2016), for instance, suggest treating web services as fiduciaries in their handling of their users' data, i.e., 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 hold fiduciary duties in the status quo.[45] Under Balkin and Zittrain's proposal, web services would be similarly 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 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 Burners-Lee, is an example of a PIMS. Unlike our proposal or Fukuyama et al.'s 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 web users to import their data to the platform, where these data are then encrypted. Users may then decide with which web services that integrate with digi.Me they want to share

[45] Moreover, it is emphasized that assigning web services fiduciary duties would deal with an argument 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.

their data. 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. 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. Finally, there is an app called UBDI based on digi.Me's platform that allows its users to choose to share their data in an anonymized fashion with third parties conducting research in exchange for monetary compensation.

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 in EDPS (2016) and EDPS (2020).[46] 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 EDPS (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" in subsection 3.10. Congruent with this notion, most existing PIMS do not entail remuneration of consumers for their data.

Our data intermediaries are similar to PIMS in that they consolidate and handle consumers' data. However, our data intermediaries 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. Thus, consumers only need to make a few decisions in our approach — they can 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 codifies 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.

[46] The German Ministry of Economics is currently working on a directive detailing the standards PIMS have to fulfil.

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.IO. The web browser Brave blocks all third-party ads. 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. Users of Brave can choose to receive targeted ads facilitated by Brave themselves in exchange for cryptocurrency. Similarly, Permission.IO's browser extension tracks a consumer's browsing behavior and suggests targeted pop-up video ads to the consumer based on the former. A consumer receives cryptocurrency when watching a video and additional compensation when clicking on a link suggested by such ad videos. Advertisers engaging with web users through Permission.IO never receive access to the consumers' data.

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 in the report CDEI (2021). In this 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, i.e., 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.[47] As such, it is related to recent work by Martens et al. (2021), who suggest the establishment of an *in situ* data access right for consumers on platforms. 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 et al. (2021) differs from ours in the sense that data portability relies on web users' efforts. By contrast, we allocate this responsibility to the data intermediaries in our framework.

In the 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". 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 – henceforth DFA (RadicalxChange 2020). The authors suggest the establishment of data coalitions with fiduciary duties, where communities of data providers (web users) coordinate to collectively bargain over the use of their data. 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 stand in solidarity with each other to advance their common interests.

However, there are two crucial distinctions between our approach and the DFA. Firstly, the DFA envisions a significant participation of data coalition members in the

[47] Article 26 of the German TKG law also gives the regulator the right to mandate interoperability of dominant players in the digital market.

governance of these data coalitions.[48] By contrast, the actions of the data intermediaries in our framework require no democratic legitimation. Such democratic processes 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 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, 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 reflected by our paper, in which we discuss the nature of the transaction processes for individual data (auction mechanisms, 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 paper.

Another natural solution to the problem of online firms exploiting consumers' data is for government to tax the use of data.[49] 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 does not provide consumers with control over their data. Additionally, it sets the incentive for web services not to advertise or collect data to a uniform arbitrary selected 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), who propound their proposal in Feygin et al. (2021). This working group 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 the use or store consumer data, with the amount of the tax depending on the extent to which the company uses or stores consumer data.[50] 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.

---

[48] 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.

[49] 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 (https://www.nytimes.com/2019/05/06/opinion/tax-facebook-google.html) and expanded on the proposal in a longer essay (https://adtax.paulromer.net/).

[50] 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.

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 feared 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 to 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 code, 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 their 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 3: Middleware

Fukuyama et al. (2020) propose a regime under which consumers access digital platforms through third-party web services called *middleware*. Under this regime, each middleware service would filter, sort, fact-check, or control the display of platform content to users. One example of middleware could 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 interests. Allowing consumers to choose between middleware services offering alternative presentations of platforms' content would reduce the influence of major platforms over public discourse and online content consumption. In the status quo, major platforms' procedures for recommending and displaying content to users are opaque and therefore difficult to scrutinize and evaluate. Middleware services that transparently report their algorithms for delivering content to users would mitigate this problem. An advantage of Fukuyama et al.'s middleware proposal relative to structural remedies imposed by regulators is that competing middleware firms would have incentives to provide content in line with consumer preferences, which the regulator may have difficulty ascertaining. Also, a middleware market with free entry allows for more innovation in content delivery than regulations stipulating how platforms should deliver their content to users.

Our proposal focuses on a different problem that of Fukuyama et al.; we focus on providing consumers control over and remuneration for their data as opposed to reducing the power of major platforms over online content consumption. The adoption of either our proposal or Fukuyama et al.'s proposal does not rule out the adoption of the other. Both our proposals allow consumers to choose how they engage with online content, but in distinct ways; data intermediaries allow consumers to choose their levels of data sharing whereas middleware services allow consumers to choose the type and presentation of information served by web services. With that being said, middleware could give consumers more control over privacy settings in certain cases; a middleware service used for interacting with Facebook, for example, could offer a consumer finer control over how that consumer's data are used than the privacy settings currently available on Facebook.

Our data intermediaries could benefit middleware services by selling them access to consumer data for conducting analytics. Middleware services would seek to analyze external data this point because, under Fukuyama et al.'s proposal, middleware firms would be barred from collecting data from their users. Additionally, competition between middleware services would not increase consumers' control over the terms under which they access online services because consumers pay little attention to terms of service as they are currently presented. Thus, middleware services will not be able to make themselves more attractive to consumers by offering superior privacy settings. Consequently, middleware services are unlikely to drive each other to adopt terms of service preferred by consumers. Our data intermediation proposal would help consumers choose data-share settings in line with their preferences by centralizing and simplifying their terms of service and by making their features highly salient.

Appendix 4: Solid and data storage pods

Solid is a platform that aims to decentralize the web by providing its users ownership and control over their data. Tim Burners-Lee, the inventor of the World Wide Web, created Solid in 2016. 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. The sort of 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. A user's Pod data container can take the form of a cloud data storage unit or a physical drive (e.g., a USB drive). Additionally, Solid features services for encrypting its users' personal data. 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. The vision underlying Solid is the creation of a Web in which a standardized system for storage and control of consumer data provides the basis for the interoperability of web services.

Solid is intended to provide consumers with extensive control over the storage and use of their data. Few web services currently support Solid, which is in a prototype phase. Our data intermediaries are similar to Solid in that both consolidate and process consumer data. Unlike our intermediaries, which 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. Indeed, data intermediaries' management of many users' data will allow them to bargain on their users' behalf with web services in a way that a single Solid user could not bargain with web services. Web services will not value a single Solid user's data that highly for reasons discussed in the main text. They will, however, value the massive amounts of data that a data intermediary stands to offer. A key difference between our data intermediaries and Solid is that our proposal results in compensation to consumers who choose to share their data, which owes to the fact that data intermediaries will be able to bargain with web services. Solid does not currently feature any scheme allowing consumers to be remunerated for their data.

The adoption of either Solid or our proposed data intermediation system could result in increased innovation and creativity amongst developers and platform producers. 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. Although some advanced web users may desire to personally manage their data, 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.

Appendix 5: 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 section, 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. The study of He et al. (2017) finds evidence consistent with improvements in search result accuracy from additional users providing feedback about search results. Yoganarasimhan (2020) finds a clear positive relationship between the length of personal data records and various measures of search result accuracy. Additionally, Schaefer and Sapi (2020) 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.

Several other studies assess the value of data in contexts other than search. Neuman et al. (2019) 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. Another general study of the value of consumer data is provided by Azevedo et al. (2019), 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. One study of data's value in the content recommendation context is Claussen et al. (2019), who compare the quality of algorithmic and editorial news recommendation. They find that the length of personal data records leads the algorithm to outperform 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 et al. (2019) provides one of the rare studies measuring the price differences between these sorts of advertisement and finds that ads served to users who opt out of behavioral targeting—that is, ads based on past browsing behavior—yield 52% less revenue than ads shown to consumers who do not opt out. An earlier study of Beales and Eisenach (2013) 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. The authors find that the addition of a 90-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 (Ravichandran and Korula, 2019, Eisenach (2013) and Johnson et al. (2019). 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.).[51] A data markets 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%.[52] According to the *Los Angeles Times*, the data brokerage industry is thought to be worth around $200 billion ($526 per capita) as of November 2019.[53]

A recent court settlement provides further indication about the valuation of user data resulting from a bargaining process: A class action lawsuit 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.[54] Facebook is also currently facing a $15 billion dollar probe for illegally tracking and selling user data.[55] 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.

[51] The full story may be viewed under Edelman (2020).
[52] See here: https://s3.amazonaws.com/media.mediapost.com/uploads/InternetAdvertisingRevenueReportApril2021.pdf
[53] See here: https://www.latimes.com/business/story/2019-11-05/column-data-brokers
[54] See here: https://www.theverge.com/2021/2/27/22304618/judge-approves-facebook-privacy-settlement-illinois-facial-recognition
[55] See here: https://www.reuters.com/article/us-usa-court-facebook-idUSKBN2BE1TX

Table 1: Back-of-the-envelope estimates of the value of data						
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
Illinois Lawsuit††	2021	p	$66–132B	$200–400*	$267–534B	$34–68

Table notes:

- All figures are in US dollars. The "total" columns report aggregate valuations of data for the respective geography, whereas the "p.c." columns report per capita valuations. When information on the size of the user base of the web service for which we make a back-of-the-envelope calculation is available, we use this user base size to compute per capita values and obtain the total valuation by multiplying the per capita values with 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 (US 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 US, and a population of 7.9 billion for the world. We use the ratio between the nominal global GDP expressed in USD and the nominal US GDP to convert the total values between different geographies. The nominal world GDP is 85 trillion USD, the nominal US GDP is 21 trillion USD. The resulting ratio is 4.04. (Source: World Bank estimates for 2020 as reported in https://data.worldbank.org/indicator/NY.GDP.MKTP.CD).
- 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. The list following these table notes provides the source corresponding to each alphabetic label.
- All rows with a setting labelled "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.
- †The Facebook study project offered $10–20 compensation per month in exchange for access to data.
- ††As a result of the Illinois class-action lawsuit settlement, affected Facebook users were entitled to a $200–400 dollar compensation from Facebook.

List of links to sources:

a) US online ad spending in 2020: 140B USD (page 4, https://www.iab.com/wp-content/uploads/2021/10/IAB-PWC-Outlook-2022-The-Digital-Advg-Ecosystem-Oct-2021.pdf)
b) Global digital ad spending in 2020: 378B USD (https://www.emarketer.com/content/worldwide-digital-ad-spending-2021)
c) US data brokerage industry in 2020: 200B USD (https://www.latimes.com/business/story/2019-11-05/column-data-brokers)
d) Google US ad revenue in 2020: 40B USD (https://www.statista.com/statistics/469821/google-annual-ad-revenue-usa/)
e) Google unique visitors in US in 2020: 270M (https://www.statista.com/topics/1001/google/)

f) Google global ad revenue in 2020: 147B USD (page 33, https://abc.xyz/investor/static/pdf/20210203_alphabet_10K.pdf)
g) Amazon US ad revenue in 2020: 16B USD (https://www.wsj.com/articles/amazon-surpasses-10-of-u-s-digital-ad-market-share-11617703200)
h) Twitter US ad revenue in 2020: 1.6B USD (https://www.emarketer.com/content/will-the-twitterpurge-bolster-ad-growth)
i) Twitter active US users in 2021: 77.8M (https://datareportal.com/essential-twitter-stats?rq=twitter)
j) Twitter global ad revenue in 2020: 3.2B USD (page 38, https://s22.q4cdn.com/826641620/files/doc_financials/2020/ar/FiscalYR2019_Twitter_Annual_Report.pdf)
k) Twitter global users in 2020: 187M (https://datareportal.com/essential-twitter-stats?rq=twitter)
l) Facebook US & Canada ad revenue in 2020: 40B (https://www.statista.com/statistics/223280/facebooks-quarterly-revenue-in-the-us-and-canada-by-segment/)
m) Facebook US & Canada monthly active users in 2020: 255 million (https://s21.q4cdn.com/399680738/files/doc_financials/2021/q3/FB-Earnings-Presentation-Q3-2021.pdf)
n) Facebook average revenue by user in 2020: 32USD (https://www.statista.com/statistics/234056/facebooks-average-advertising-revenue-per-user/)
o) Facebook study project (2019): monthly compensation of 10–20 USD per month (https://techcrunch.com/2019/06/11/study-from-facebook/)
p) Illinois lawsuit (2021): 200–400 USD (https://www.facebookbipaclassaction.com)

Note: All links were last accessed on 14 February 2022

Works cited

Argenziano, Rosella and Alessandro Bonatti. 2021. “Data linkages and privacy regulation”. Working paper accessed at http://www.mit.edu/~bonatti/protection.pdf on January 17, 2022.

Augenblick, Ned and Scott Nicholson. 2016. “Ballot position, choice fatigue, and voter behaviour.” *The Review of Economic Studies* 83 (2): 460–480.

Azevedo, Eduardo M., Alex Deng, José Luis Montiel Olea, Justin Rao, and E. Glen Weyl. 2020. "A/b testing with fat tails." *Journal of Political Economy* 128 (12): 4614-4672.

Balkin, Jack M. 2016. “Information Fiduciaries and the First Amendment.” UC Davis Law Review 49: 1183-1234.

Balkin, Jack M. and Jonathan Zittrain. 2016. “A grand bargain to make tech companies trustworthy.” *The Atlantic*. Accessed at https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346/ on January 17 2021.

Baumeister, Roy F., Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice. 1998. “Ego depletion: Is the active self a limited resource?” *Journal of Personality and Social Psychology* 74 (5): 1252–1265.

Beales, Howard, and Jeffrey A. Eisenach. 2014. "An empirical analysis of the value of information sharing in the market for online content." Working paper. Accessed at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2421405 on January 17, 2022.

Behavioural Insights Team and Doteveryone. 2020. “Active Online Choices: Designing to Empower Users - Summary of desk research.” Accessed at https://www.bi.team/wp-content/uploads/2020/11/CDEI-Active-Online-Choices-Desk-Research-Write-up-FOR-PUBLICATION-1.pdf on January 17, 2022.

Blake, Tom, Sarah Moshary, Kane Sweene, and Steven Tadelis. 2021. “Price salience and product choice”. *Marketing Science* 40(4):619-636.

Brown, Jennifer, Tanjim Hossain, and John Morgan. "Shrouded attributes and information suppression: Evidence from the field." *The Quarterly Journal of Economics* 125.2 (2010): 859-876.

CDEI (Centre for Data Ethics and Innovation). 2021. “Unlocking the value of data: Exploring the role of data intermediaries.” Accessed at: 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 on January 17, 2022.

Claussen, Jörg, Christian Peukert, and Ananya Sen. 2019. "The editor vs. the algorithm: Targeting, data and externalities in online news." Accessed at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3399947 on February 10, 2022

CMA (Competition and Market Authority). 2019. “Online platforms and digital market study.” Accessed at https://www.gov.uk/cma-cases/online-platforms-and-digital-advertising-market-study on February 10, 2022.

Data Ethics Commission. 2020. “Opinion of the Data Ethics Commission.” Accessed at https://www.bmjv.de/DE/Themen/FokusThemen/Datenethikkommission/Datenethikkommission_EN_node.html on January 17, 2022.

Data Rights Protocol. 2021. “Data Rights Protocol: Standardizing and streamlining consumer data rights requests”. Accessed at https://datarightsprotocol.org/ on January 17, 2022.

Edelman, Gilad. 2020. “Can killing cookies save journalism”. Accessed at https://www.wired.com/story/can-killing-cookies-save-journalism/ on January 17, 2022.

EDPS (European Data Protection Supervisor). 2016. “EDPS Opinion on Personal Information Management Systems.” Accessed at https://edps.europa.eu/sites/edp/files/publication/16-10-20_pims_opinion_en.pdf on January 17, 2022.

EDPS (European Data Protection Supervisor). 2020. “Personal Information Management Systems.” EDPS TechDispatch on Personal Information Management Systems. Accessed at https://edps.europa.eu/data-protection/our-work/publications/techdispatch/techdispatch-32020-personal-information_en on January 17 , 2022.

Einav, Liran, Theresa Kuchler, Jonathan Levin and Neel Sundaresan. 2015. Assessing sale strategies in online markets using matched listings. *American Economic Journal: Microeconomics*, *7*(2), 215-47.

Elvy, Stacy-Ann. 2017. "Paying for privacy and the personal data economy." *Columbia Law Review* 117: 1369–1459.

European Commission. 2018. “Commission Decision AT.40099 -- Google Android.” Accessed at https://ec.europa.eu/competition/antitrust/cases/dec_docs/40099/40099_9993_3.pdf on January 17, 2022.

European Commission. 2020a. “A proposal for a regulation of the European Parliament and of the council on contestable and fair markets in the digital sector (Digital Markets Act).” Accessed at https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A842%3AFIN on January 17, 2022.

European Commission. 2020b. “A proposal for a regulation of the European Parliament and of the council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC.” Accessed at https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A825%3AFIN on January 17, 2022.

European Commission. 2020c. “A proposal for a regulation of the European Parliament and of the council on European Data Governance (Data Governance Act)” Accessed at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767 on February 10, 2022.

Federal Trade Commission. 2009. “Self-regulatory principles for online behavioral advertising.” FTC Staff Report. Accessed at https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-staff-report-self-regulatory-principles-online-behavioral-advertising/p085400behavadreport.pdf on March 8, 2021.

Feygin, Yakov, Hanlin Li, Chirag Lala, Brent Hecht, Nicholas Vincent, Luisa Scarcella, and Matthew Prewitt. 2021. “A data dividend that works: Steps toward building an equitable data economy.” Berggruen Institute White Paper. Accessed at https://www.berggruen.org/ideas/articles/a-data-dividend-that-works-steps-toward-building-an-equitable-data-economy/ on January 17, 2022.

Fukuyama, Francis, Barak Richman, Ashish Goel, Roberta R. Katz, A. Douglas Melamed, and Marietje Schaake. 2020. “Report of the working group on platform scale.” Accessed at https://cyber.fsi.stanford.edu/publication/report-working-group-platform-scale on January 17, 2022.

Gentzkow, Matthew, Jesse M. Shapiro, Frank Yang, and Ali Yurukoglu. 2021. “Advertising prices in equilibrium: Theory and evidence.” Working paper. Accessed at https://www.brown.edu/Research/Shapiro/pdfs/reduced.pdf on January 17, 2022.

Grzybowski, Lukasz, and Ambre Nicolle. 2021. "Estimating Consumer Inertia in Repeated Choices of Smartphones." *The Journal of Industrial Economics* 69.1 (2021): 33-82.

He, Di, Aadharsh Kannan, Tie-Yan Liu, R. Preston McAfee, Tao Qin, and Justin M. Rao. 2017. “Scale effects in web search.” *Web and Internet Economics*: 294–310.

Hirschleifer, David, Yaron Levi, Ben Lourie, and Siew Hong Teoh. 2019. “Decision fatigue and heuristic analyst forecasts.” *Journal of Financial Economics* 133 (1): 83–98.

Ho, Kate, Joseph Hogan, and Fiona Scott Morton. 2017. "The impact of consumer inattention on insurer pricing in the Medicare Part D program." *The RAND Journal of Economics* 48(4): 877-905

Hortacsu, Ali, Seyed Ali Madanizadeh, and Steven Puller, 2017. “Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market.” *American Economic Journal: Economic Policy* 9(4): 192-226.

House Bill 3849. 2021. “A Bill To promote competition, lower entry barriers, and reduce switching costs for consumers and businesses online.” Accessed at https://www.govtrack.us/congress/bills/117/hr3849/text on January 17, 2022.

Johnson, Eric. J., 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, and Elk U. Weber. 2012. “Beyond Nudges: Tools of a choice architecture.” *Marketing Letters* 23: 487–504. Accessed at https://link.springer.com/article/10.1007/s11002-012-9186-1 on August 20, 2021.

Johnson, Garrett A., Scott K. Shriver, and Shaoyin Du. 2019. “Consumer privacy choice in online advertising: Who opts out and at what cost to industry?” *Marketing Science* 39 (1): 33–51.

Larson, Jeff, Surya Mattu, and Julia Angwin. 2015. “Unintended Consequences of Geographic Targeting.” *Technology Science*. Available at: http://techscience.org/a/2015090103/.

MacKay, Alexander, and Marc Remer. 2022. "Consumer inertia and market power." *Working Paper*. Accessed https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3380390 on February 11, 2022.

Mathur, Arunesh, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. "Dark patterns at scale: Findings from a crawl of 11K shopping websites." *Proceedings of the ACM on Human-Computer Interaction* 3(CSCW): 1–32.

Martens, Bertin, Geoffrey Parker, Georgios Petropoulos, Marshall van Alstyne. 2021. "Towards Efficient Information Sharing in Network Markets”, working paper.

Neumann, Nico, Catherine E. Tucker, and Timothy Whitfield. 2019. "Frontiers: How effective is third-party consumer profiling? Evidence from field studies." *Marketing Science* 38 (6) : 918–926.

Persson, Emil, Kinga Barrafrem, and Andreas Meunier, and Gustav Tinghög. 2019. “The effect of decision fatigue on surgeons’ clinical decision making.” *Health Economics* 28 (10): 1194–1203.

Posner, Eric and E. Glen Weyl. 2019. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*: [need page numbers]

RadicalxChange. 2020, “The Data Freedom Act.” Accessed at https://www.radicalxchange.org/media/blog/solving-the-social-dilemma/ on January 17, 2022.

Ravichandran, Deepak and Nitish Korula. 2019. “Effect of disabling third-party cookies on publisher revenue”. Last accessed at https://apachedigital.io/wp-content/uploads/2020/01/disabling_third-party_cookies_publisher_revenue.pdf on February 10, 2022.

Schaefer, Maximilian, and Sapi, Geza. 2020. “Learning from data and network effects: The example of internet search.” DIW Discussion Paper No. 1894.

Scherer, Ronny and Fazzat Siddiq. 2019. “The relation between students’ socioeconomic status and ICT literacy: Findings from a meta-analysis.” *Computers & Education* 138: 13-32.

Scott Morton, Fiona and David Dinielli. 2020a “Roadmap for a digital advertising monopolization case against Google” *Omidyar Network.* Accessed at https://omidyar.com/wp-content/uploads/2020/09/Roadmap-for-a-Case-Against-Google.pdf on January 17, 2022.

Scott Morton, Fiona and David Dinielli. 2020b “Roadmap for an antitrust case against Facebook.” *Omidyar Network.* Accessed at https://www.omidyar.com/wp-content/uploads/2020/06/Roadmap-for-an-Antitrust-Case-Against-Facebook.pdf on January 17, 2022.

Scott Morton, Fiona, Gregory Crawford, Jacques Crémer, David Dinielli, Amelia Fletcher, Paul Haidhues, Monika Schnitzer, and Katja Seim. 2021. "Equitable Interoperability: the "Super Tool" of Digital Platform Governance". *Yale Tobin Center for Economic Policy Discussion Paper No. 4*. Accessed at https://tobin.yale.edu/sites/default/files/Digital%20Regulation%20Project%20Papers/Digital%20Regulation%20Project%20-%20Equitable%20Interoperability%20-%20Discussion%20Paper%20No%204.pdf on January 17, 2022.

Wills, Craig E. and Can Tatar. 2012. "Understanding what they do with what they know." *Proceedings of the 2012 ACM workshop on privacy in the electronic society (WPES 2012)*. Accessed from http://web.cs.wpi.edu/~cew/papers/wpes12.pdf on January 17, 2022.

Yoganarasimhan, Hema. 2020. "Search personalization using machine learning." *Management Science* 66 (3): 1045–1070.