Privacy & market concentration: Intended & unintended consequences of the GDPR

Garrett A. Johnson, Scott K. Shriver & Samuel G. Goldberg*

January 13, 2021

Abstract

We show that the European Union's General Data Protection Regulation (GDPR) reduced data sharing online, but had the unintended consequence of increasing market concentration among technology vendors that provide support services to websites. We collect panel data on the web technology vendors selected by more than 27,000 top websites internationally. The week after the GDPR's enforcement, website use of web technology vendors falls by 15% for EU residents. Websites that would face greater penalties under the GDPR drop more vendors. Websites are more likely to drop smaller vendors, which increases the relative concentration of the vendor market by 17%. Increased concentration predominantly arises among vendors that use personal data such as cookies, and from the increased relative shares of Facebook and Google-owned vendors, but not from website consent requests. These findings suggest increases in concentration are driven by website's vendor choices rather than changes in user behavior.

Keywords: Privacy, GDPR, Competition, Web technology, Regulatory compliance

*Johnson: Questrom School of Business, Boston University <garjoh@bu.edu>; Shriver: Leeds School of Business, University of Colorado – Boulder <Scott.Shriver@colorado.edu>; Kellogg: School of Management, Northwestern University <Samuel.Goldberg@kellogg.northwestern.edu>. We gratefully acknowledge the financial support of the Marketing Science Institute and the Program on Economics & Privacy at George Mason University. We thank Avi Goldfarb, Ginger Jin, Laura Kornish, Jura Liaukonyte, Jordan Mitchell, Rob Porter, Liad Wagman and Ran Zhuo for helpful comments & discussions.

1 Introduction

Academics and policymakers worry that privacy regulation could harm competition. For example, large firms may have more technical and financial resources to comply with regulation (Brill, 2011; Phillips, 2019). Further, where regulations require consent for personal data processing, large firms can more easily obtain consent from individual consumers (Campbell et al., 2015). In this paper, we suggest that policies limiting business-to-business data sharing could also benefit large firms. Potential business partners may favor large vendors because they offer a better product or better regulatory compliance, thereby limiting legal risk. Our empirical work provides novel evidence of these tradeoffs between privacy and competition policy.

Both privacy and competition concerns are acute in our study of the web technology market, where vendors provide technology support services to websites. These services include: raising ad revenue, hosting audiovisual content, measuring visitor activity, and facilitating social media sharing. Web technology is an area of concern for privacy regulators because of its large-scale personal data processing (Commission Nationale de l'Informatique et des Libertés, 2019; Information Commissioner's Office, 2019). Web technology's largest companies—Google and Facebook—capture 56% of global digital advertising spend (WARC, 2019), and draw privacy and competition scrutiny from regulators on both sides of the Atlantic.

To empirically investigate the impact of privacy policy on competition, we examine website choices of web technology vendors in response to the European Union (EU) enforcing the General Data Protection Regulation (GDPR). Europe's GDPR serves as a model for privacy regulation in Brazil, Japan, South Korea and several American states. These state-level regulations, like the California Consumer Privacy Act, may herald privacy regulation at the federal level. Despite this policy momentum, a growing literature reveals unintended consequences of privacy policy. Privacy policy can slow technology diffusion (Miller and Tucker, 2009, 2017; Adjerid et al., 2016) and even increase data breaches (Miller and Tucker, 2011). The GDPR coincided with lower venture capital investment for EU technology firms (Jia et al., 2019b,a), reduced web traffic and revenue (Goldberg et al., 2020; Aridor et al., 2020), and limited personalized marketing effectiveness (Godinho de Matos and Adjerid, 2019). If privacy policy harms competition, this compounds concerns about market power in the economy (Council of Economic Advisors, 2016; Berry et al., 2019).

Online web technology interactions provide an opportunity to measure the otherwise opaque data transfers between firms that the GDPR seeks to limit. When users visit a website, their browsers also interact with the third-party domains of web technology vendors and often share user identifiers (e.g. stored on cookies) that the GDPR considers to be personal data. We exploit this behavior in our data collection, which periodically crawls a sample of websites and directly observes their web technology vendor usage. We examine a panel of over 27,000 websites drawn from the top 2,000 sites in each EU country, the US, Canada and globally. We

browse each site throughout 2018 using a specialized tool to record web technology vendor interactions and using a VPN service to appear as a French user, yielding an initial set of over 375,000 website-vendor ties.

To guide our analysis, we develop a theoretical model of the web technology market under the GDPR. Our model assumes that websites purchase services from technology vendors that are differentiated by heterogeneous data processing costs. GDPR enforcement introduces a chance that websites could be penalized for sharing data with web technology vendors. Websites trade off the benefit of flouting a law and the chance of a penalty, as in the tax evasion literature (Becker, 1968; Allingham and Sandmo, 1972; Yitzhaki, 1974; Slemrod, 2019). We derive comparative statistics that provide testable predictions for how the GDPR will affect a website's choice of vendors and concentration in the vendor market. Our main result shows that vendor market concentration increases as the probability of GDPR enforcement increases. We also explore key moderators of GDPR enforcement effects. For example, we show that websites with higher revenue shifters—site traffic, ad count, and user income—make deeper cuts to their vendors post-GDPR.

Our empirical analysis confirms the theory model predictions. We find that the GDPR restricts website use of web technology vendors, leading to the unintended consequence of increased vendor market concentration. We observe a steep but short-lived 15% drop in website-vendor relationships, which return to pre-GDPR levels by the end of 2018. Though we do not observe market conduct (e.g. pricing) of web technology vendors, we document changes to the market structure of the industry. We find that relative market concentration increases 17% in aggregate in the short run, implying that the GDPR most reduces the market shares of small web technology vendors. We emphasize the short-run estimates, where the drop in website-vendor relationships is largest, as our model suggests this is when website beliefs about the probability of enforcement are highest, and because our short-run effect estimates are robust to confounds from independent trends in vendor usage. We further find that concentration increases in the top four web technology categories that comprise 94% of categorized vendor ties: advertising, web hosting, audience measurement, and social media. Concentration is pronounced among web technology vendors that process personal information, so that personal data collection also becomes more concentrated after the GDPR. We find that concentration does not depend on whether websites elicit consumer consent for data processing, implying that website rather than user choices drive increases in concentration. Finally, we show that website choices entrench Google and Facebook, whose web technology offerings drive increased concentration.

Our study contributes to several streams of academic literature. By viewing the data through an economic lens, we complement computer science research documenting the GDPR's impact on web technology (e.g. Libert et al., 2018; Sørensen and Kosta, 2019; Urban et al., 2020). Our study adds to a broader economic literature documenting the unintended consequences of legislation designed to promote consumer health or welfare. Prior studies have documented how restrictions on advertising led to increased market concentration

in markets for cigarettes (Eckard JR., 1991; Gallet, 1999; Clark, 2007) and alcohol (Sass and Saurman, 1995). As with studies of restrictions on information flows from firms to consumers, we find anti-competitive effects from restrictions on information flows from consumers to firms. A blog post by WhoTracks.me (2018) first noted that Google and Facebook fared relatively better than smaller ad vendors after the GDPR. Subsequent work by Peukert et al. (2020) supports our key findings: the GDPR increased concentration in this industry and Google plays a dominant role. A key difference is that we collect data from the vantage point of an EU user whereas they use public data from the vantage point of a US user. Since the GDPR only applies to EU users, Peukert et al. (2020) rely on spillovers of the GDPR in how websites treat US users to identify the effect of the GDPR. Lefrere et al. (2020) show that online content providers also cut vendor use after the GDPR, but do not find that these sites reduced content creation. In theoretical work, Sharma et al. (2019) show that privacy regulation can hurt competition for both ad vendors and online publishers.

Our analysis of website vendor choices under the GDPR also contributes to the literature on regulatory design. To ensure compliance, a regulator trades off the size of fines with the probability of levying a fine (Polinsky and Shavell, 2000). Most related empirical studies examine the impact of enforcement actions on firm behaviors (Kang and Silveira, 2018; Johnson, 2020). EU regulators did not levy penalties on website vendors in 2018 but have since criticized the industry for contravening the GDPR (e.g. Data Protection Commission, 2020). As such, firm beliefs about the chance of enforcement actions play an important role here. Our model suggests that the rise in vendor use after the GDPR can in part be explained by declining firm beliefs about the probability of enforcement over time. We connect website vendor choices to beliefs about local regulatory strictness, following Sheffrin and Triest (1991), who relate self-reported beliefs about tax enforcement to tax evasion. The GDPR shares enforcement powers between an EU-wide regulator and country-level regulators which may vary in regulatory strictness and resources. We show that a survey measure of regulatory strictness in data protection predicts the degree to which a website's reduction in vendors persists through 2018 or reverses.

The rest of the paper proceeds as follows. In Section 2, we briefly review the GDPR policy. Section 3 presents a theoretical model of the web technology industry. Section 4 then describes our data. Section 5 discusses our empirical results for vendor usage, and in Section 6 we present our formal analysis of web technology vendor concentration. Section 7 concludes.

2 GDPR Background

The European Union's General Data Protection Regulation (GDPR) regulates the processing of personal data of EU residents. Though passed in April 2016, enforcement of the GDPR was delayed until May 25,

2018 to allow stakeholders time to adjust. The regulation acknowledges the global nature of information flows: the GDPR applies to both EU firms and non-EU firms that target EU residents. GDPR fines can reach 4% of a firm's annual global revenue. Given that fines apply to global revenue rather revenue from the EU, the GDPR incentivizes global firms that serve EU residents to abide by its principles.

Though the GDPR is a multifaceted regulation, many of its elements support the GDPR's key principle of data minimization: firms must limit the personal data that they process (i.e., collect and use). Firms are explicitly required to audit internal data processes, encrypt and anonymize personal data, and notify affected individuals and the regulator in the event of a data breach. Firms are also responsible for respecting the new data rights of EU residents under the GDPR, including the rights to: access personal data, correct data, erase data, transfer data and object to data processing. In sum, the GDPR incentivizes firms to limit personal data processing by increasing both its associated operational cost and legal liability.

The GDPR has significant implications for the web technology sector and its methods of operation. In particular, the definition of personal data under the GDPR includes browser cookies and IP addresses, which are used extensively by web technology vendors to identify individual consumers. Post-GDPR, websites are restricted from sharing such personal user data with third parties, except under explicitly defined legal bases. EU regulators further clarified in 2019 that "consent" is the most appropriate of the GDPR's legal bases for technology vendors to process personal data for affiliated websites (Information Commissioner's Office, 2019).[1] Valid consent under the GDPR requires that individuals opt-in to data processing, and that consent notices must list both the purposes of data processing (e.g., for advertising or audience measurement) and all third parties processing the data. In 2018, most websites that sought consent did so by asking users to either click "OK" to accept all disclosed data processing, or by clicking "More Options" to opt-out of specific usage categories or vendors (Utz et al., 2019).

In the two years since the GDPR's implementation, EU regulators have released multiple reports critical of industry practices but have not levied precedent-setting fines (e.g. Autoriteit Persoonsgegevens (2019); Data Protection Commission (2020)). The European Commission (2019) one-year GDPR status report emphasized the need for greater enforcement for the GDPR to "become fully operational." In addition to the overall lack of enforcement actions, additional concerns have been raised about unequal enforcement across national jurisdictions, which arise because EU member states are individually responsible for funding and managing GDPR enforcement activities. A European Data Protection Board (2020) survey found that 21

[1] Other than individual consent, valid bases for data processing include: compliance with a legal obligation or contractual performance, protecting "vital interests" (life, safety) of a data subject, acting under official public authority, and "legitimate interests." Legitimate interest may potentially be claimed when "processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject" (GDPR (6)(1)(f)). Given the need to establish a balance of interests among stakeholders, "legitimate interest" is considered a risky compliance strategy and is discouraged by regulators (Information Commissioner's Office, 2019).

of 30 country-level data protection authorities report they have insufficient human, financial, and technical resources. We find patterns consistent with these observations in our analysis of the GDPR's effects over time and across EU countries.

3 Theory

Before turning to our empirical work, we first devise a concise theory model to guide intuition for the market response to GDPR enforcement. Our primary interest lies with interactions between *websites*, who utilize technology services to monetize their web traffic, and web technology *vendors*, who provide the related technology services. We formulate an oligopoly model with asymmetric vendor costs in the spirit of Ledvina and Sircar (2012), where websites choose technology vendor service levels and vendors strategically compete in service output. The perceived probability of GDPR enforcement enters the website profit function like a tax and shifts website expected profits accordingly. The equilibrium solution to the model delivers predictions for the number of active vendors in the market and the market shares associated with each vendor. We summarize key comparative statics (principally with respect to GDPR enforcement) in a series of propositions.

3.1 Vendor service demand

Ex-ante, a total of N vendors offer technology services to websites. Websites choose continuous service levels for each vendor, q_i, where $i \in \{1, ..., N\}$ and $0 \leq q_i \leq 1$. The per-unit market price of vendor $i's$ service is given by p_i.

Website revenue We posit a reduced form revenue function for websites, wherein vendor service levels shift baseline website revenues. The representative website revenue function takes the following form:

$$R(\vec{q}) = \beta \left[1 + \sum_{i=1}^{N} q_i - \frac{1}{2} \sum_{i=1}^{N} q_i^2 - \frac{\gamma}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} q_i q_j I(i \neq j) \right] \tag{1}$$

In equation (1), β (where $\beta > 0$) captures baseline revenue (when vendor service levels are zero) and incorporates exogenous shifters of website revenue. Revenues increase at a decreasing rate in vendor service levels q_i, which implies that at sufficiently low prices, websites will contract with multiple vendors. In addition to individual vendor satiation, the model flexibly captures the extent of cross-vendor substitutability through the γ parameter, where $0 \leq \gamma \leq 1$ ($\gamma = 0$ implies all vendor services are independent, while $\gamma = 1$ implies all vendor services are identical).

Website profit Absent GDPR incentives, websites choose vendor service levels to maximize net profit, which incorporates the cost of web technology vendor services, as well as any costs of goods sold. For simplicity, we assume that the cost of products sold on websites is a fixed proportion of revenue, where ϕ ($0 \leq \phi \leq 1$) captures the constant product cost to price ratio. For websites that act as publishers, we expect $\phi = 0$ because the product being sold is display ad inventory, which entails high fixed (content creation) costs but effectively zero marginal costs.

If websites violate GDPR privacy restrictions for EU citizens, a penalty of 4% of global revenue may be imposed. GDPR enforcement therefore changes the objective function for websites. Hence, we include a term in the profit function to capture the expected enforcement penalty of 4% total revenue times α, where α ($0 \leq \alpha \leq 1$) captures the website's perceived probability of GDPR violation and enforcement. The resulting website expected profit incorporating GDPR incentives is thus:

$$\begin{aligned}\Pi^W(\vec{q}) =& R(\vec{q})(1-\phi) - \vec{p}\cdot\vec{q} - 0.04\alpha R(\vec{q}) = (1-\phi-0.04\alpha)\,R(\vec{q}) - \vec{p}\cdot\vec{q} \\ =& \tilde{\beta}\left[1 + \sum_{i=1}^{N} q_i - \frac{1}{2}\sum_{i=1}^{N} q_i^2 - \frac{\gamma}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} q_i q_j I(i \neq j)\right] - \sum_{i=1}^{N} p_i q_i\end{aligned} \tag{2}$$

where we have defined $\tilde{\beta} \equiv (1-\phi-0.04\alpha)\,\beta$ for notational convenience. $\tilde{\beta}$ may be interpreted as website baseline expected profit (post-GDPR), absent the use of technology vendor services ($\vec{q} = 0$).

3.2 Vendor service supply

We assume vendors are heterogeneous with respect to service costs, c_i. Without loss of generality, we assume vendors are ordered by service costs so that $0 < c_1 < c_2 ... < c_N$. For tractability, we assume a convenient cost distribution such that $c_i = \delta i$. The supply of vendor services is then driven by the vendor profit function:

$$\Pi_i^V(\vec{q}) = q_i(p_i - c_i) = q_i\,(p_i - \delta i) \tag{3}$$

3.3 Equilibrium

In Appendix A.1 we determine the equilibrium in two stages. We first solve for aggregate market outcomes, which we summarize in Proposition 1 below.

Proposition 1. *Ignoring integer constraints, the number of active vendors in equilibrium is:* $n^* = \frac{\sqrt{2\delta(2-\gamma)[(1-\gamma)\delta+\gamma\beta(1-\phi-0.04\alpha)]}-(2-\gamma)\delta}{\gamma\delta}$. *Aggregate vendor output with n^* firms is then given by:* $Q(n^*) = \frac{2\gamma\beta(1-\phi-0.04\alpha)+\delta(4-3\gamma)-2\sqrt{2\delta(2-\gamma)[(1-\gamma)\delta+\gamma\beta(1-\phi-0.04\alpha)]}}{2\gamma^2\beta(1-\phi-0.04\alpha)}$.

Proof. See Appendix A.1. □

We then solve for individual vendor market shares, which we summarize in the following corollary:

Corollary 2. *In equilibrium, the market share for vendor i is given by:*

$$s_i^* = \frac{q_i^*}{Q(n^*)} = \begin{cases} \frac{1}{(2-\gamma)Q(n^*)} \left(1 - \gamma Q(n^*) - \frac{\delta i}{(1-\phi-0.04\alpha)\beta}\right) & i \leq n^* \\ 0 & i > n^* \end{cases}$$

Proof. See Appendix A.1. □

Vendor market shares may then be used to compute concentration measures, such as the n-firm concentration ratio: $CR(n) = \sum_{i=1}^{n} s_i$.

3.4 Comparative statics

Finally, we analyze how the equilibrium model solution changes with respect to the model primitives. We begin with the primary parameter of interest, α, which captures website's perceived risk of GDPR-related fines. We summarize our findings in the following Proposition:

Proposition 3. *As the perceived probability of GDPR enforcement (α) increases:*

(a) the number of active vendors in equilibrium (n^*) falls, and

(b) vendor market concentration increases.

Proof. See Appendix A.2. □

Proposition 3 supports the notion that privacy regulation and related incentives can lead to anti-competitive effects, including increased market concentration among web technology vendors. A further consequence of Proposition 3 is that any time dependence in the perceived risk of GDPR-related fines (α) will be reflected in changes to vendor market concentration over time. We provide empirical support for Proposition 3 in Sections 5.1 and 6.1.

Another parameter of interest is the website baseline revenue (β), which can be related to observed website revenue shifters. We summarize the comparative statics of the model solution with respect to β in Proposition 4:

Proposition 4. *The number of active vendors in equilibrium (n^*) increases in baseline website revenue (β). For higher revenue websites, increasing the perceived probability of GDPR enforcement (α) leads to a larger reduction in the number of active vendors than for smaller revenue websites.*

Proof. See Appendix A.3. □

In Section 5.2.2, we validate the predictions of Proposition 4 by constructing a dataset of website characteristics and vendor usage and then regressing the observed number of vendors (post-GDPR) on website characteristics.

Finally, we summarize the effects of shifting the other model parameters, which are consistent with economic intuition:

Proposition 5. *The number of active vendors in equilibrium (n^*) decreases in website product costs (ϕ), vendor service costs (δ) and vendor service substitutability (γ).*

Proof. See Appendix A.4. □

Proposition 5 illuminates important mechanisms that may drive observed levels of vendor usage, independent of the GDPR. First, Proposition 5 helps to explain observed cross-sectional variation in vendor usage by vendor service category. Further, Proposition 5 speaks to expected trends in vendor usage over time. The web technology industry is characterized by steady technological innovation, with vendor innovation arising from falling data processing costs (lower δ) and increasing service differentiation (lower substitutability, γ). To the extent our sample coincides with a period of technological innovation, the model predicts an increasing trend in vendor usage, similar to the predicted effect of declining beliefs about GDPR enforcement over time.

4 Data

4.1 Data description

To study the GDPR's impact on the web technology industry, we collect panel data on the web technology vendors employed by thousands of top websites. Websites rely on inputs from specialized vendors to provide various services. For instance, many websites engage "audience measurement" vendors to record user site visits and generate statistics on user characteristics, onsite activities, and referral channels. Websites can choose the category's dominant vendor—Google Analytics—and/or competitors like Adobe Audience Manager. When a user visits a website, their browser interacts with third-party domains owned by those vendors: `google-analytics.com` (Google) and/or `demdex.net` (Adobe). By recording these third-party domain interactions, we can observe if the website employs Google Analytics and/or Adobe Audience Manager. In practice, some vendors use multiple domains to support their operations, so our analyses aggregate observations to the more economically substantive website-vendor level.

We collect third-party domain data using the "webxray" tool developed by Tim Libert (webxray.org) and first used in Libert (2015). For each website in our panel, webxray opens an instance of the Chrome browser and records all interactions with third-party domains. Third-party cookies are the best-known form of third-party domain interactions, but webxray also records third-party domain interactions arising from HTTP and Javascript requests. We use a VPN service to represent the browser as originating from France. This ensures that our data captures the site-vendor ties that are regulated by the GDPR. Note that webxray does not interact with the website in any way, so we measure vendor interactions arising *without* explicit user consent.

To construct our sample, we use Amazon's Alexa service to identify the top 2,000 websites in each of the 28 EU countries as well as in each of the US, Canada, and globally. These lists overlap such that our initial sample includes 28,227 unique sites. For our baseline, we collect data in the two days leading up to the GPDR's enforcement deadline. Beginning on the May 25 deadline, we collect data weekly for six weeks, bi-weekly for the next six weeks, then every four weeks through the end of 2018. In total, we scan websites 14 times over the 28 week period from May 23, 2018 to December 3, 2018 (1 week pre-GDPR to 27 weeks post-GDPR). As Libert (2015) explains, webxray sometimes fails to scan a site. When this happens, we make at least three attempts to scan the site. During our data collection, 3.27% of sites never scan, perhaps because the sites block VPN users or potential bots. Our panel dataset (summarized in Table 1) is therefore based on the 27,303 sites that successfully scan at some point during our observation window.

Although our panel provides extensive cross-sectional and post-GDPR temporal coverage, our single pre-GDPR observation does not identify trends in vendor usage prior to the enforcement deadline. Comparably-sourced data from other studies suggest that the trend in pre-GDPR vendor usage was relatively flat (WhoTracks.me, 2018; Sørensen and Kosta, 2019) to slightly increasing (Peukert et al., 2020). Non-decreasing vendor usage leading up to GDPR implies that a drop in vendors May 25, 2018 would be credibly associated with the GDPR (rather than an ongoing exogenous trend). If the pre-trend is increasing, simple pre/post GDPR outcome comparisons would *understate* the GDPR's effect. We return to the pre-trend issue in our discussion of robustness in Section 5.3.

4.2 Vendor classification

A central challenge in measuring concentration is to appropriately define the market. We define markets by classifying web technology vendors into broad purposes, like advertising and audience measurement. Measuring market concentration requires not only classifying vendors by purpose, but also linking third-party domains to vendors. To solve both challenges, we use a third-party domain database by Libert

(2019). This database clarifies when vendors use a third-party domain without the vendor's name and when vendors use multiple third-party domains. For instance, Google's advertising category offering uses both the `doubleclick.net` domain (a past acquisition) and the `2mdn.net` domain.

The Libert (2019) database groups vendors into nine categorizes.[2] By unique vendors in our data, the top categories are advertising (165 vendors), hosting (25), audience measurement (24), and social media (11). The "advertising" category includes full service ad vendors (e.g. Google Marketing Platform/Ad Manager, Xandr) and different ad intermediaries. These include ad exchanges (e.g. OpenX, Index Exchange); demand side platforms (The Trade Desk, AdForm); supply side platforms (Rubicon, PubMatic); and data management platforms (Oracle Bluekai, Lotame). "Social media" includes social platforms like Facebook and Twitter as well as social sharing tools like AddThis and ShareThis. "Hosting" is a broad category for vendors that host websites or site content elements. The category contains webhosts (Amazon Web Services, Cloudflare); tag management (Google Tag Manager only); website code (Google APIs, jQuery Foundation); video serving (Google Video/YouTube, Vimeo); and fonts (Typekit, Fonts.com). "Audience measurement" includes vendors that focus on reporting for the site's internal purposes (Google Analytics and Adobe Audience Manager) as well as vendors that focus on external reporting (Comscore and Alexa). Smaller categories include website security and bot detection ("security"), customer service chat widgets ("CRM": customer relationship management), platforms for "native ads," and "privacy compliance." Appendix B.1 lists the top five vendors in each category.

We use the Libert (2019) database because it provides an independent and reasonable categorization that covers the majority of our data. Categorization is challenging even for broad categories because vendors can offer multiple services that straddle multiple categories. For instance, the "audience measurement" and "design optimization" categories differ by whether they offer advanced services like experiments and user session recording. Vendors such as Hotjar, which offers both types of services, appear in both categories under the Libert (2019) classification.

4.3 Descriptive statistics

Our *complete dataset* is an unbalanced panel of website observations that captures website use of technology vendors over time. The *1 week pre-GDPR cross-section* corresponds to the first time observation of websites and serves as our only pre-GDPR scan. We use this baseline to evaluate the GDPR's effects over time. We compare the baseline cross-section to the *1 week post-GDPR cross-section* in order to generate our short-run

[2]We re-label some categories and omit the subcategories of the "hosting" category: "general", "code", "font", and "video." We combine Google's ad vendor offering ("Google Marketing Platform/Ad Manager") as well as its video offering ("Google Video/YouTube"). We combine "TrustArc" & "TRUSTe" (a rebranding) and "Are You a Human" & "Distill Network" (a 2017 merger).

GDPR effect estimates. Our long-run GDPR effect estimates emphasize the comparison of the baseline cross-section with the *27 weeks post-GDPR cross-section,* our final observation in 2018. Table 1 provides key descriptive statistics for each of these samples.

The first (horizontal) section of Table 1 provides a detailed view of the *baseline cross-section,* which comprises the 26,368 websites (96.6%) that we successfully scan the week prior to the GDPR. We first report the number of third-party domain interactions, which captures interactions of all types (cookies, HTTP, Javascript). We then report the associated number vendor interactions, where we use the Libert (2019) database to map third-party domains to vendors. Before the GDPR, websites interact with an average of 16.4 third-party domains and 14.5 web technology vendors per site. Websites have a median of 9, a minimum of 0, and a maximum of 199 vendors. On average, 7.3 (50.5%) of these vendors place a third-party cookie on the browser. The Libert (2019) database categorizes 8.5 vendors per site on average, representing 58.1% of the 383,384 website-vendor ties in the baseline cross-section.

In terms of sample selection, the median global Alexa rank for these sites is #57,714. The data contain the top ranked site (`google.com`) and the lowest ranked site is #6,589,497. Our site selection emphasizes top sites in the EU, with an average of 74.6% of all site traffic generated by EU users. Across sites, the EU user share of site traffic covers the full range of 0% to 100%, with a median of 97.9%.

The second, third and fourth sections of Table 1 report vendor usage outcomes for the *1 week post-GDPR cross-section, 27 weeks post-GDPR cross-section* and the *complete (panel) dataset,* respectively. In the post-GDPR period, 98.1% (26,781/27,303) of sites scan successfully 1 week post-GDPR and 96.7% scan successfully 27 weeks post-GDPR. For the complete dataset, the panel is 96.4% (368,487/14*27,303) balanced, reflecting the overall scan success rate. Scan rates improve slightly in those key post-GDPR scans, which mitigates concerns that GDPR compliance interferes with our measurement procedure.[3] The mean number of vendors employed one week post-GDPR (12.6) is -13.5% smaller than the pre-GDPR baseline (14.5) and -9.5% smaller than the sample average vendor usage (13.9). However, by 27 weeks post-GDPR, average vendor usage (14.9) rebounds to be 2.3% higher than baseline levels. This simple comparison of conditional means suggests a post-GDPR pattern of short-run declines in vendor usage followed by a long-run recovery in vendor usage. In the next section, we use regression models to investigate this pattern with greater econometric rigor.

[3]Across our 14 time observations, the average scan success rate is 96.4% with standard deviation of 2.33%.

Table 1: Descriptive statistics of website observations

1 week pre-GDPR (baseline) cross-section	Obs.	Mean	St. Dev.	Min.	Med.	Max.
Third-party domains	26,368	16.4	18.2	0	11	211
Vendors	26,368	14.5	16.8	0	9	199
Vendors using third-party cookie	26,368	7.3	12.3	0	3	144
Categorized vendors	26,368	8.5	10.4	0	5	92
First party cookie	26,368	0.9	0.3	0	1	1
Alexa rank	26,345	155,548	291,605	1	57,714	6,589,497
EU user share (%)	26,331	74.6	35.8	0	97.9	100.0
1 week post-GDPR (short run) cross-section						
Vendors	26,781	12.6	14.9	0	8	204
27 weeks post-GDPR (long run) cross-section						
Vendors	26,414	14.9	17.1	0	10	162
Complete dataset (panel)						
Vendors	368,487	13.9	16.1	0	9	220

5 Vendor usage analysis

We begin our empirical analysis by examining how the GDPR impacted websites' use of technology vendors. To estimate the effect of the GDPR on website use of technology vendors, our empirical strategy relies on a simple before and after comparison. By examining the GDPR's effect on website vendor usage (Section 5.1) and heterogeneity in the GDPR effect by website characteristics (Section 5.2), we test the comparative static relationships predicted in Section 3. We discuss measurement issues and the robustness of our analysis in Section 5.3. Section 6 then analyzes vendor market concentration.

5.1 GDPR impact on vendor usage

5.1.1 Aggregate usage

Our empirical strategy relies on before-after comparisons to measure the effect of the GDPR. Identifying a control group poses a key challenge for studying the GDPR. First, non-EU websites may not be representative and are still subject to the GDPR if they target EU users. Given confusion of the interpretation of "targeting"—which the European Data Protection Board clarified in November 2018—non-EU sites may treat EU traffic with caution. GDPR therefore can affect how non-EU websites treat EU users. Second, non-EU users may experience some spillover effects of the GDPR: that is, websites may implement GDPR measures for non-EU users to reduce administrative costs or enforcement scrutiny. We show that non-EU websites implement GDPR measures for EU users (Section 5.2.2) and we show that websites expose EU users to fewer vendors than non-EU users (Section 5.3.6). We conclude that neither non-EU sites nor non-EU

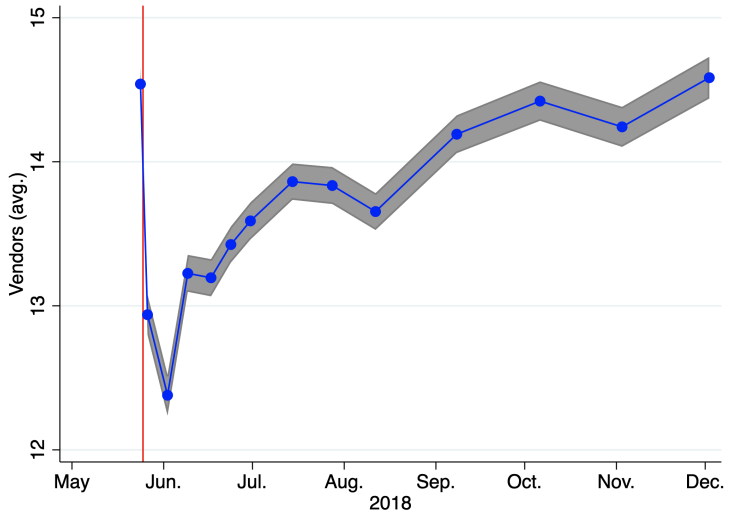
users represent clean controls.

We therefore favor pre- vs. post-GDPR comparisons for a set of sites to quantify the effect of the GDPR on web technology vendors. Specifically, we analyze how the number of vendors engaged by a website evolves after enforcement of the GDPR using the following fixed effects regression:

$$y_{it} = \mu + \lambda_t \cdot GDPR_t + \theta_i + \varepsilon_{it} \tag{4}$$

where y_{it} is website i's number of technology vendors at time t, $GDPR_t$ is an indicator for the post-GDPR enforcement date t, θ_i is a site fixed effect, and ε_{it} is the error term. The coefficients λ_t therefore capture the difference in the average number of web technology vendors relative to our pre-GDPR baseline (μ), after conditioning on website fixed effects. We note that our econometric specification attributes all common temporal changes in outcomes to the GDPR. For this reason, we anticipate our GDPR effect estimates to be most accurate in the short-run, when independent usage trends have the least potential to be conflated with changes due to the GDPR.

Figure 1: Evolution of average web technology vendor usage per website



We use our complete dataset to estimate equation (4) and present the results graphically in Figure 1. The figure shows how website use of web technology vendors evolves over 2018 by plotting our regression estimates (λ_t) as a function of time.[4] We see that average web technology vendor use drops sharply after GDPR enforcement on May 25, 2018 (denoted by the vertical red line). Vendor use reaches its minimum one week later. We refer to this comparison between the pre-GDPR baseline and one week post-GDPR as the

[4]Figure 1 plots the pre-GDPR average web technology vendors ($E[y_{i0}]$), and subsequent weeks ($E[y_{i0}]+\lambda_t$) where confidence intervals use the standard errors for μ and λ_t respectively.

short-run GDPR effect estimate. The short-run estimate shows that sites reduce web technology vendors 14.9%, from an average of 14.5 to 12.4 vendors. Site-level fixed effects therefore leads to a slightly higher GDPR effect measurement than the simple means comparison in Section 4.3 (13.5% vendor reduction). Three quarters of this reduction happens right after the enforcement deadline as the number of vendors falls 11.0% between the initial scan on May 23-24 and the second scan on May 25-28. This finding suggests that most publishers waited until the last minute to adjust the vendors on their site.

One of our starkest findings is that the short-run GDPR effect appears to erode over time: by the end of 2018, the average number web technology vendors effectively returns to its pre-GDPR level. This result is still consistent with Zhuo et al. (2019) who do not find that GDPR affected the Internet's network level connectivity: due to the modest contribution of vendors to aggregate data flows as well as the magnitude and duration of the GDPR effect. In interpreting the pattern in Figure 1, we emphasize the descriptive nature of our results. Lacking a control group, we caution against attribution the full series of λ_t estimates in Figure 1 to the GDPR alone—particularly for the later estimates. With these limitations in mind, we posit that the post-GDPR growth in vendor use arises from a combination of: (1) declining site beliefs about GDPR enforcement, and (2) natural industry growth. We address alternate explanations in Section 5.3 including: sites blocking EU users, sites blocking third-party domain loads, vendor entry & exit, and transient compliance issues. We show these explanations play little to no role in the evolution of vendor use as seen in Figure 1.

These two explanations may be understood within the context of our theory model. Explanation (1) corresponds to enforcement beliefs declining over time following the GDPR implementation ($\alpha \to 0$ as $t \to \infty$), a corollary to Proposition 3. We suggest that beliefs about GDPR enforcement risk were most heightened around the May 25 enforcement deadline and subsequently declined as regulators did not levy fines and signaled they would first study the industry (e.g. Information Commissioner's Office, 2019; Data Protection Commission, 2020). In Section 5.2.1, we provide evidence that variation in enforcement expectations plays a contributory role in the observed pattern of GDPR effects. The enforcement expectations explanation is also consistent with the European Commission (2019) review of GDPR, which cited lack of enforcement as an obstacle to the regulation's full realization.

Explanation (2) posits that the GDPR induced a downward shift in vendor usage after May 25, which eroded over time due to innovation among web technology vendors. As discussed in Section 3.4, Proposition 5 implies that vendor innovation could materialize in the form of declining data processing costs (δ) or reduced service substitutability (γ). Growth and innovation are typical for the vendor industry: Lerner et al. (2016) show a steady increase in website use of third-party domains between 1996 and 2016. Moreover, Peukert et al. (2020) document an increasing trend in vendor use through 2017 and 2018 (excepting the enforcement

deadline).

Explanations (1) and (2) highlight the rationale for using the short-run effect estimate as the relevant benchmark for the GDPR's impact on the web technology industry. To the extent that GDPR effects diminish over time due to flagging enforcement expectations, the short-run estimate captures the regulation's effect when enforcement beliefs were highest in 2018. This reference frame strikes us as the most policy-relevant since it captures what the effect could be under a persistent enforcement regime. The short-run estimate is also preferred on econometric grounds, because our long-run estimates become increasingly confounded by industry growth trends. For these reasons, our subsequent analyses emphasize the short-run impact of the GDPR on web technology.

5.1.2 Usage by vendor category

Next we characterize vendor usage by service category. Category-level analysis lends insight into the GDPR's effects on different types of web technology vendors and links directly to our discussion of market concentration, which relies upon the same categorization scheme. With nine categories, we simplify our presentation of results to a comparison of the short-run GDPR effect (1 week pre- vs. 1 week post-GDPR) and a single long-run effect (1 week pre- vs. 27 weeks post-GDPR) for each category, which we report in Table 2.

Table 2: GDPR impact on average vendor use by category

	Pre-GDPR†	Short run (SR)‡			Long run (LR)*		
Category	Average	Estimate	St. Err.	Diff. (%)	Estimate	St. Err.	Diff. (%)
All vendors	14.54	-2.09	0.063	-14.4%	0.05	0.081	0.3%
All categorized vendors	8.45	-1.49	0.040	-17.6%	-0.24	0.051	-2.8%
Advertising	4.39	-1.06	0.033	-24.1%	-0.28	0.044	-6.3%
Hosting	1.78	-0.17	0.005	-9.7%	0.09	0.006	5.0%
Audience measurement	1.25	-0.14	0.004	-10.9%	-0.02	0.004	-1.6%
Social media	0.79	-0.09	0.003	-11.5%	-0.03	0.004	-3.2%
Design optimization	0.22	-0.02	0.001	-10.5%	-0.01	0.002	-2.7%
Security	0.15	-0.03	0.001	-17.7%	0.00	0.002	0.1%
Native ads	0.08	-0.01	0.001	-14.6%	-0.01	0.002	-13.2%
CRM	0.02	-0.002	0.0004	-9.7%	-0.001	0.001	-3.7%
Privacy compliance	0.02	0.004	0.001	22.9%	0.02	0.001	123.6%

Notes: Coefficient estimates from separate fixed effect regressions (see equation 4) for each category. †Pre-GDPR baseline given by means of scanned sites1 week pre. ‡1 week pre vs. 1 week post. *1 week pre vs. 27 weeks post.

The first column of Table 2 reports the pre-GDPR mean and the next three columns report the short-run GDPR coefficient estimate, standard error, and percentage difference relative to the pre-GDPR mean. Each coefficient represents a separate fixed effect regression (equation 4) corresponding to each vendor category outcome. We see that web technology vendors overall fall 14.5% and the subset of categorized vendors falls

17.7% from 8.4 to 6.9. The category-level results in Table 2 reveal that the average number of vendors falls for all but one category in the short run. The exception is the "privacy compliance" category, which we expect would benefit from the GDPR. However, few sites use vendors in the privacy compliance category, as these increase from only 0.017 to 0.021 vendors on average. Advertising is both the largest category and the category that falls the most (24.3%), from 4.35 to 3.29 average vendors. Hosting, audience measurement, and social media are the next largest categories and these categories fall by 9.7%, 10.9%, and 11.5% respectively. The remaining categories appear infrequently with means of at most 0.22 vendors per site.

The last three columns of Table 2 report the long-run change in web technology vendors by category. For all categories, vendor usage increases in the long run, compared to the measured short-run effects. The advertising category shows the largest attenuation of the short-run effect, which is consistent with either significantly revised GDPR enforcement expectations or greater innovation in this category.

In sum, our short-run estimates demonstrate clear evidence of a GDPR effect. This drop is sudden with 74% of the short-run reduction in vendors arising within a couple days of the enforcement deadline. We observe no comparable change in vendor usage for the rest of 2018, and other research suggests no such change in the year prior to the GDPR either (Peukert et al., 2020). Vendor usage falls in all categories but privacy compliance, where it increases. This pattern is consistent with a GDPR effect rather than technology change or some other transitory shock. We further discuss alternate explanations and robustness in Section 5.3.

5.2 Usage effect heterogeneity and GDPR mechanisms

In this section, we explore heterogeneity in website choice of vendors post-GDPR. We do so for several related reasons. We want to better understand how the design of the GDPR's incentives affect firm behavior. To that end, we introduced a model in Section 3 that generates testable predictions in the form of Propositions 3 and 4. Validating these predictions reveals the incidence of the regulation and suggests that our model captures important features of the regulation. Empirical verification of our predictions further confirms that the GDPR drives these changes rather than alternative explanations like a coincident change in technology, as discussed above.

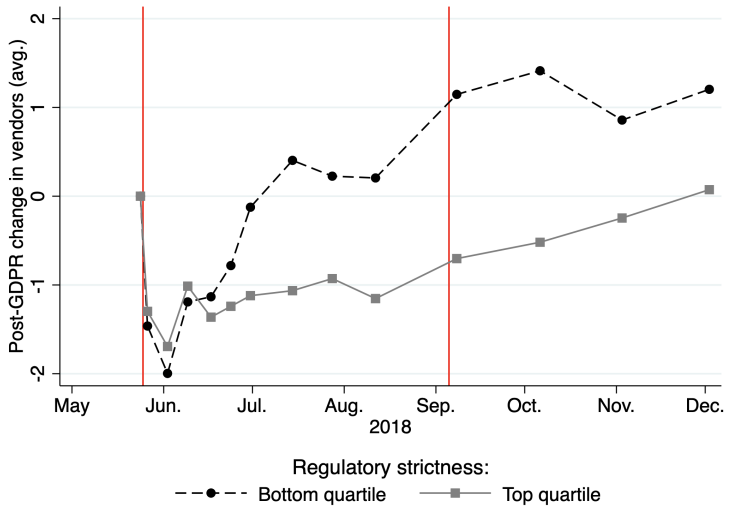
In support of Proposition 3, Section 5.2.1 provides evidence that GDPR effects are moderated by varied expectations of regulatory enforcement (α in the theory model). In support of Proposition 4, Section 5.2.2 provides evidence that GDPR effects are moderated by site characteristics that generate greater revenue (β in the theory model).

5.2.1 Enforcement expectations

The GDPR's Recital 9 expresses the need for the GDPR to standardize EU data protection law, though important differences between countries persist. The central EU privacy regulator recognizes that EU country-level regulators vary both in their resources and their enforcement intensity (European Data Protection Board, 2020). We exploit this cross-country variation to gain insight into GDPR's effects across regions with different expectations of regulatory strictness. Specifically, we use a European Commission (2008) survey of 4,835 data controllers that asks if the country's regulator is more or less strict than other countries in the EU. According to this survey, the strictest data regulators are Germany and Sweden whereas the laxest regulators are Bulgaria and Greece. We construct a regulatory strictness measure for each website as the weighted average of the four-point country strictness survey measures, with weights determined by the proportion of website users from each EU country (sourced from Alexa).

We analyze the role of enforcement expectations by comparing vendor usage among websites in high regulatory strictness regimes to those in low regulatory strictness regimes. To infer both short and long run effects, we replicate our analysis of Section 5.1, by estimating equation (4) on two data subsamples and plotting the model-predicted average usage over time. We focus on majority-EU sites (sites receiving more than 50% of their traffic from EU users) to ensure a tighter link between the website and a given EU country's regulator and to avoid any confounding relationship between foreign sites and traffic from certain (e.g. English-speaking) EU countries. We then split the sample by quartile of regulatory strictness.

Figure 2: GDPR effects moderated by expectations of regulatory enforcement



Note: To relate websites to EU country regulatory authorities, we restrict attention to sites with over 50% of traffic from the EU.

Figure 2 summarizes the analysis by plotting average vendor usage among websites in the top quartile of regulatory strictness and websites in the bottom quartile. We note that sites in stricter regimes return to pre-GDPR levels at the end of 2018, whereas sites in laxer regimes do so in July. These results strongly suggest that enforcement expectations materially contribute to the pattern of GDPR usage effects captured in Figure 1. Other GDPR studies and also find that regulatory strictness moderates the impact of the GDPR on technology venture capital (Jia et al., 2019b) and recorded site outcomes (Goldberg et al., 2020). In the next section, we further show that the correlation between regulatory strictness and the post-GDPR trend in vendor usage is robust to including covariates that are correlated with regulatory strictness, such as the site's traffic, share of EU users, and user income.

5.2.2 Website characteristics

We focus on website characteristics that potentially explain variation in vendor choice. Our theory model suggests an important role for exogenous website revenue shifters, denoted by β. Specifically, Proposition 4 predicts that higher revenue websites will reduce vendor usage more than their lower revenue counterparts. We suggest three observed variables shift website revenue. First, we use the traffic rank of the site—which is inversely related to the volume of site traffic—as measured by Alexa. Second, we use the average user income, which we expect correlates with advertising and e-commerce revenue. To construct this measure, we use each site's share of users by country (from Alexa) to weight income per capita at the country level. Third, we use the number of ads per page, which we expect will shift site revenue from advertising. We collect this variable in August 2018 by visiting each website homepage while using an ad blocker to count the number of blocked ads, following Shiller et al. (2018).

We further consider three website characteristics that are not related to revenue but are of interest. First, given the findings of the previous section, we include our normalized measure of regulatory strictness (for majority-EU sites) as a covariate in our analyses. Second, we consider the website's share of traffic from the EU, again using data from Alexa. The share of users from the EU potentially affects the relative benefit of sharing personal data given that GDPR fines are a share of global revenue. Third, we include an indicator for sites without traffic from the EU as sites that do not target EU users are not subject to the GDPR. Note that we avoid potentially endogenous changes in website characteristics over time, by fixing the associated variables using pre-GDPR levels wherever possible.

We analyze differences in both the short-run and long-run effects of the GDPR (on vendor usage) by website characteristics. To do this, we use our panel dataset to estimate a regression equation that interacts

the six website variables discussed above with a post-GDPR indicator and a post-GDPR time trend:

$$\begin{aligned} y_{it} &= \mu + \lambda \cdot GDPR_t + \sum_k \psi_k \cdot X_{ik} \cdot GDPR_t + \\ &\quad \xi \cdot GDPR_t \cdot t + \sum_k \nu_k \cdot X_{ik} \cdot GDPR_t \cdot t + \theta_i + \epsilon_{it} \end{aligned} \tag{5}$$

where y_{it} indicates the number of vendors used by website i at time t. We use $k \in \{1, ..., 6\}$ to index website characteristics so that X_{ik} is the k^{th} (time-stationary) characteristic of website i. The model includes site fixed effects θ_i, so we only include site characteristics X_{ik} as an interaction. This regression is designed to test whether website characteristics explain post-GDPR differences in usage levels (short-run effects) and usage trends (long-run effects). As such, we use the full panel but omit the May 25^{th} scan, so that the short-run GDPR interactions better capture the 1-week post-GDPR effect in keeping with the rest of the paper.[5]

Our estimates of the short-run GDPR interaction effects, in the first column of Table 3, illuminate the role of website incentives. In relation to our model, the short-run effect of the GDPR is to increase the chance of a penalty from 0 to some $\alpha_{SR} > 0$. Though we do not directly observe website beliefs, our theory model predicts how different sites will cut vendors as α increases. From Proposition 4, sites with greater revenue shifters (β) engage more marginal vendors ex-ante and thus will cut more vendors post-GDPR. Noting that the sign on site rank is opposite those on ad count and user income because site rank is inversely related to site traffic, the corresponding coefficients in column (1) have the predicted sign and are each statistically significant with $p < 0.01$, lending empirical support to Proposition 4.

The interaction with the share of EU users is positive and marginally significant ($p < 0.1$). This result suggests that sites with the smallest share of EU users make the deepest cuts to their vendors. We expect this finding relates to the design of the GDPR penalties. Since GDPR penalties are 4% of global revenue, sites with a small share of EU users have comparatively little to gain but more to lose from violating GDPR provisions, leading them to cut more vendors. We estimate a positive and significant discontinuity in the interaction between the GDPR and an indicator for sites without EU users. This discontinuity is expected, since sites that do not serve EU users are excluded from the GDPR and should therefore respond little to the GDPR, if at all. The positive interaction between share of EU users and the GDPR indicator illuminates how the design of the GDPR's penalties shapes websites' choice of vendors. This indicates an unintended consequence of the GDPR penalty design: sites that serve more EU users are less privacy-protective in this

[5] Our results are generally robust to including the May 25^{th} scan.

Table 3: Heterogeneity in GDPR effect by website characteristics: short-run effect & long-run trend

Interactions time horizon	Short-run (pre vs. 1wk post)	Long-run post-trend
$GDPR_t$ (λ)	-4.543***	
	(0.371)	
$GDPR_t$ x $Week$ (ξ)		0.0270***
		(0.00988)
Interactions (ψ, ν)*:*	$GDPR_t$ *x*	$GDPR_t$ *x Week x*
log(Site rank)	0.337***	-0.0042***
	(0.041)	(0.0012)
log(Ad count + 1)	-0.973***	0.0561***
	(0.076)	(0.0021)
User income†	-0.601***	-0.0081***
	(0.066)	(0.0017)
Share of EU users (%)	0.0043*	0.0003***
	(0.0026)	(7.04e-05)
No EU users	1.813***	-0.0192**
	(0.273)	(0.0081)
Regulatory strictness† x >50% EU users	0.296***	-0.0043**
	(0.072)	(0.0020)
Constant (μ)	14.85***	
	(0.056)	
Site fixed effects	x	
Observations	329,158	
R-squared	0.835	

Note: Full panel except the May 25, 2018 data pull. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. †Denotes normalized variable.

sense. Appendix C.4 provides more evidence for the moderating role of the share of EU traffic.

Finally, the coefficient on regulatory strictness is positive and highly significant ($p < 0.01$), so that sites facing stricter data regulators cut vendors less in the short-run. However, the magnitude of this effect is small, as suggested by Figure 2. A one standard deviation increase in regulatory strictness implies 2.0% (0.296/14.54) more vendors one week after the GDPR. We suggest three explanations. First, websites facing stricter regulators may have been more careful with their vendor use prior to the GDPR in response to existing laws and the EU e-Privacy Directive.[6] Second, we could interpret this as weak evidence of anticipatory behavior by sites that face a strict regulator, so that our GDPR estimates in Section 5.1 may understate the full reduction in vendors. Third, if sites are risk averse and low-strictness regimes are also more uncertain, we would then observe sites in low-strictness regimes cutting vendors more.

Our estimates of the long-run GDPR interaction effects, which capture heterogeneity in post-GDPR usage trends, are reported in the third column of Table 3. As expected from Figure 1, the baseline trend after the GDPR in column (3) is a positive: the number of vendors increases by 0.027 per week. The post-GDPR

[6]We thank Avi Goldfarb for suggesting this possibility.

trend linked to regulatory strictness is negative (-0.0043) and significant ($p < 0.05$). This result provides evidence that the central insight from Figure 2 is robust to the inclusion of covariates: website beliefs about the enforcement probability play an important role in determining vendor usage, with sites in stricter regimes being slower to add vendors post-GDPR.

The estimated signs of the revenue shifters and post-GDPR time trend interactions are largely as expected. Suppose that, in the absence of enforcement, website beliefs about the likelihood of enforcement fall – that is, α falls from its short run value α_{SR} to its long run value $\alpha_{LR} < \alpha_{SR}$. Then, Proposition 4 predicts that the sign on interactions with the revenue shifters should flip. The results validate this prediction for the interactions with site rank and ad count: sites with more traffic and more ads lead the post-GDPR growth in vendor use. The sign on income is negative and highly significant, which represents the sole finding contrary to our theory-driven hypotheses. We speculate that a correlation between income and user preferences for privacy might discipline sites with high-income user traffic.

Finally, we note that the interaction of the post-GDPR time trend with the share of EU users is positive and significant while the interaction with the indicator for sites without EU users is negative and significant. Thus, GDPR-exempt sites, which experienced smaller short-run usage declines, actually tend to reduce vendor usage slightly in the long-run post-GDPR. However, majority-EU traffic sites increase vendor usage more rapidly post-GDPR than sites with small shares of EU traffic. This is consistent with the GDPR's incentive structure, which penalizes firms based on their global revenues.

5.3 Alternative mechanisms and effect robustness

In this section, we discuss alternative explanations for our measured GDPR effect on vendor usage. We summarize these concerns and evidence supporting the robustness of our effects, with analysis detail relegated to the appendices.

5.3.1 Transient compliance

Some readers may wonder whether the short run nature of the GDPR reflects a transient compliance issue: perhaps sites removed non-compliant vendors at the GDPR deadline, then added these vendors back once they learned to comply. First, we examine the set of site-vendor ties that break one week after the GDPR in Appendix C.1.2. We find that only one third of these ties reconnect by the end of 2018. In other words, most of the post-GDPR increase in vendor use can be explained by websites forming new connections with existing vendors that the sites were not using pre-GDPR. Second, we emphasize that sites did not just drop small vendors but also dropped top Google services in the short run including Google Analytics (dropped by

5.6% of sites) & Google Marketing Platform/Ad Manager (dropped by 6.3% of sites). Third, the industry's GDPR compliance framework—the Internet Advertising Bureau's Transparency & Consent framework (IAB TCF)—creates a standard for sharing user consent signals. This framework is straightforward to implement, so we not see this as a major impediment for vendors. As we elaborate below, multiple EU regulators maintain that the industry's standard practices post-GDPR contravene the regulation, so we should not conclude that the industry reached full compliance at end of 2018 either.

5.3.2 Blocking third party domains or EU users

We consider website compliance concerns in Appendix C.1.1. We first consider the role of consent management systems interfering with our measurement and show that sites that request consumer consent do not block third-party domains, since we see no short run change in sites loading at least one third-party domain. This results is consistent with Sanchez-Rola et al. (2019), who also find that third-party domain interactions precede consent. Both the Irish and Dutch data regulators explain that this practice contravenes the GDPR, but also note the practice remained pervasive in 2019 (Autoriteit Persoonsgegevens, 2019; Data Protection Commission, 2020). Moreover, Figure 6b of the appendix shows a clear post-GDPR *increasing* trend in website's use of vendors with third-party cookies. Next, we consider the role of sites blocking EU users that some websites implemented as a draconian GDPR compliance strategy. We identify only 42 websites that block EU users post-GDPR and only 12 of these sites reduce the number of vendors below 10 in the short run.

5.3.3 Vendor entry & exit

In Appendix C.1.2, we consider the role of vendor compliance. A particular concern here is that vendors may choose to exit the EU market in response to the GDPR. If such exit is common, our results would reflect supply-side rather than demand-side dynamics. To this point, we first show that the largest vendor to completely exit the EU only appeared on 101 sites (0.54% of majority EU sites). We further show that the post-GDPR reduction in average vendor usage is spread out among websites: 92.8% of vendors that appear on at least 100 sites see their website reach fall post GDPR. This pattern suggests that the websites rather than the vendors are the principal decision-makers in our observed measures of vendor usage. We further consider the role of vendor entry and exit between the beginning and end of our sample. We find that entrants contribute an increase of 2.03% relative to our initial website-vendor ties, whereas exiting firms contribute a reduction of 1.96%. Thus, net entry plays almost no role in explaining the post-GDPR growth in vendors.

5.3.4 Pre-GDPR trend in vendor usage

As previously discussed, collecting only one pre-GDPR observation limits our ability to identify trends in pre-GDPR vendor usage. Our key concern is anticipatory compliance—i.e. a downward pre-trend in vendor use—that would lead us to underestimate the GDPR effect. A secular pre-trend is less concerning because we scan websites right before the GDPR deadline as our baseline. If we instead used a scan several week prior to the enforcement deadline as our baseline, an increasing pre-trend in vendor use, for instance, would lead us to underestimate the GDPR effect. We mitigate this concern by emphasizing the short run effect of the GDPR within the span of little over a week. Anticipatory compliance should manifest as a decreasing trend in vendor use in the lead up to the deadline, but other GDPR studies show no such trend. Using data collection and sampling methods distinct from our own, Sørensen and Kosta (2019) suggest a relatively stable pre-trend in third-party domain usage, whereas Peukert et al. (2020) find a slightly increasing pre-GDPR trend that is replicated (after a level shift) in the post-GDPR period. These sources give some assurance that failing to account for pre-trends will entail limited bias, which if present would understate the true GDPR effects.

In Appendix C.2, we take a further step to establish the pre-GDPR trend line in our data. We use the publicly available WhoTracks.me database, which compiles third-party domain usage data for some sites in our sample beginning in March. We match 1,322 sites in both datasets and reconstruct the pre-GDPR trend. Figure 5 in Appendix C.2 shows a nearly flat pre-GDPR trend line. Third party vendor usage is steady between March and April 2018 (+0.04%) and declines only -0.67% in May 2018, an observation that includes almost a full week post-enforcement.

5.3.5 Robustness: Alternative outcome measures

In Appendix C.3, we discuss robustness of our vendor usage analysis to different vendor types. The GDPR applies to vendors who use personal data (e.g. using cookies) and EU regulators have scrutinized ad technology vendors in particular. We show that our results are robust to replacing our primary dependent variable (which includes all web technology vendors) with: (i) the number of advertising technology vendors, and (ii) the number of vendors that use third party cookies. Both outcomes follow the same time trend as all vendors in Figure 1. Our heterogeneous short run and post-GDPR trend results by website characteristics (Table 4) are also broadly robust to both outcome measures. In particular, the interaction between regulatory strictness and the post-GDPR trend is statistically significant at the 1% level for ad vendors and at the 10% level for vendors using third party cookies.

5.3.6 Differential treatment of EU & US traffic

Appendix C.4 examines how websites treat users differently by whether or not they are in the EU. Using a VPN service, we separately scan our sample of sites three months after the GDPR representing ourselves as a user in the United States then as a user in France. For the set of majority EU sites (over 50% traffic from EU users), we show that EU users see 1 fewer vendor on average than US users. Thus, even majority EU sites find treating non-EU users differently to be advantageous. This provides more evidence that our measured decline in vendors represents a GDPR effect. Moreover, US users on majority EU sites have essentially the same number of vendors on average as EU users face before the GDPR. Thus, any potential spillover from the GDPR to non-EU users appear to have dissipated within three months. For non-EU sites, the gap in average vendors between EU and non-EU users is stark. Relative to the pre-GDPR baseline, sites with between 0% and 10% EU users increase by about 3 vendors for US users and decrease by about 3 vendors for EU users. This gap of 6 vendors on average provides more evidence that the GDPR drives changes in vendor choices and that sites with low shares of EU users are particularly responsive to the threat of GDPR fines on global revenue.

6 Vendor concentration analysis

Analyzing vendor market concentration introduces two definitional requirements. First, we must define markets in terms of vendor membership. Second, we must define vendor market shares, from which concentration measures are derived, using some observable metric of demand for vendor services. The US Department of Justice and the Federal Trade Commission, 2010 (DoJ & FTC 2010) suggest best practices for such choices in their guidance on the analysis of horizontal mergers. With regard to demand metrics, the FTC guidelines note: "In cases where one unit of a low-priced product can substitute for one unit of a higher-priced product, unit sales may measure competitive significance better than revenues. For example, a new, much less expensive product may have great competitive significance if it substantially erodes the revenues earned by older, higher-priced products, even if it earns relatively few revenues." We believe the web technology industry generally conforms to this description, due to the high rates of service innovation and cost reduction. In our context, we conceptualize vendor "unit sales" as its *reach*—i.e., the number of websites with which it transacts. For example, in the case of the advertising technology market, vendor "unit sales" are measured by the number of websites that interact with a domain owned by the vendor, and market shares represent the fraction of website-vendor interactions attributable to the vendor.[7] With regard to market definitions, we consider the industry in aggregate as well as category-level markets derived from the vendor classification

[7] We thank an anonymous referee for this insight.

discussed in Section 4.2.

For a robust quantification of vendor concentration in a (N vendor) market, we examine three concentration metrics, two of which are defined in terms of market shares $s_j = \frac{reach_j}{\sum_{k=1}^{N} reach_k} * 100$:

- *Herfindahl–Hirschman Index (HHI)*: HHI summarizes market concentration as the sum of the squared market shares:

$$HHI = \sum_{j=1}^{N} s_j^2$$

- *Concentration ratios (CR)*: The total market shares of the top M firms:

$$CR\,(M) = \sum_{j=1}^{M} s_j$$

- *Head-to-head win rate*: We propose a simple metric to quantify which vendor sites are more likely to drop. In particular, conditioning on websites that drop one of two vendors they employed prior to the GDPR, we quantify how often the sites drop each vendor. We examine the win rate of each category's dominant vendor to provide an intuitive explanation for changes in concentration.

HHI is our primary concentration metric due to its simplicity and broad use by regulators, including the US DoJ and FTC. Market shares are on a 0 to 100 scale, so that HHI varies from 0 (perfectly competitive) to 10,000 points (monopoly). We complement HHI with concentration ratios and head-to-head win rates as the latter metrics can be more intuitive. Though some models of competition (e.g. Cournot) link market structure to market conduct, we do not observe conduct like pricing, so we restrict our analysis to market structure.

We emphasize that we measure *relative* concentration. As was demonstrated in Section 5.1.2, vendor usage falls on average in all but one web technology category. We seek to measure whether websites favor vendors with large or small ex-ante market shares when websites limit vendors. In other words, we measure whether the larger vendors get a bigger slice of the smaller pie. Note that both the HHI and CR metrics are invariant if all vendors fall by the same percentage.

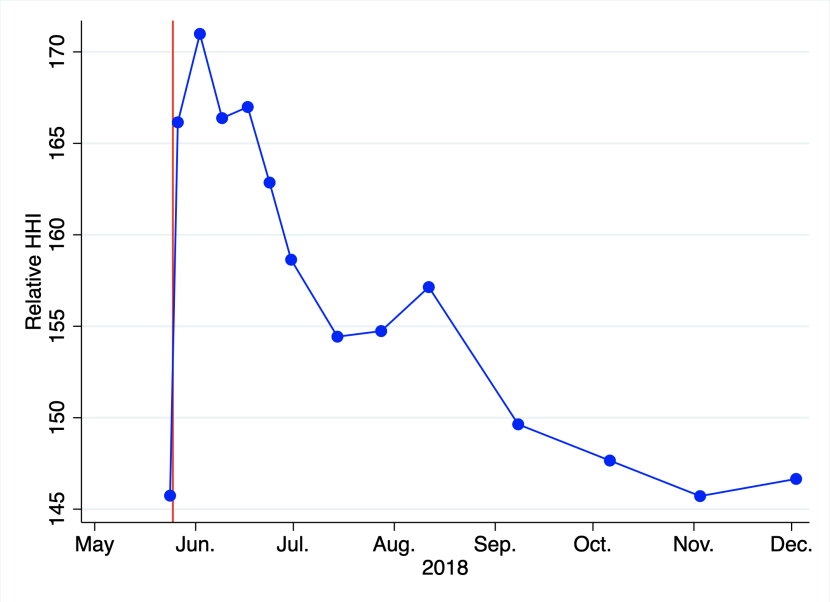
6.1 GDPR impact on vendor market concentration

6.1.1 Aggregate concentration

We first consider the web technology industry as a single market by including all vendors. Figure 3 plots the evolution in (relative) market concentration over 2018, as measured by aggregate HHI.[8] Aggregate HHI

[8]To account for changes in scanned sites over time, Figure 3 reports differences in HHI for the set of sites that are scanned both in pre-GDPR baseline and in a given post-GDPR scan. We use the pre-GDPR HHI of 146 from Table 4 as the baseline.

Figure 3: Evolution of web technology vendor concentration (HHI)



is 146 points before the GDPR and HHI reaches its maximum of 171 points one week post-GDPR, a 17.1% increase. We note that the GDPR's impact on concentration is effectively a mirror image of the impact on web technology vendors in Figure 1. As such, Figure 3 provides empirical support for part (b) of Proposition 3: market concentration increases in response to GDPR's implementation. As with average vendor use, Figure 3 reveals that web technology concentration returns to its pre-GDPR level by the end of 2018, for the same potential reasons discussed in Section 5.1.1.

6.1.2 Concentration by vendor category

We next examine the effect of the GDPR on market concentration by vendor category, with categories as described in Section 4.2. As with our vendor usage analysis, for category level results we report point estimates for the short run (1 week post) and long run (27 weeks post). For the reasons outlined in Section 5.1, we emphasize the short-run results, but also summarize our findings over the long-run horizon.

GDPR short-run concentration impact Table 4 reports short-run changes in market concentration by category. The columns labeled "Pre" show the baseline HHI and concentration ratios for the top two vendors (CR2) in every category. We see that all categories but advertising and hosting have HHI's above the 2,500 point threshold that American regulators define as a "highly concentrated market" (US DoJ & FTC 2010). Advertising has the lowest HHI (348 points) and CR2 (18.7), as ad-supported websites often employ several vendors to boost ad revenue. Advertising contains 165 vendors and sites use 4.35 ad vendors on average,

Note too that these sample differences explain the small differences in baseline HHI and CR2 in Table 4 and 5.

Table 4: Short-run GDPR impact on concentration (1 week post)

	HHI			Concentration ratio (CR2)			Head-to-head competition	
Category	Pre	Post	Diff. (%)	Pre	Post	Diff. (%)	Win (%)	Dominant firm
All vendors	146	171	17.3%	9.8	10.5	7.0%		
All categorized vendors†	308	363	17.8%	16.8	18.7	11.3%		
Advertising	348	436	25.3%	18.7	21.7	15.8%	98.9%	Google ad platform††
Hosting	1,892	1,936	2.3%	56.9	57.8	1.7%	74.3%	Google APIs
Audience measurement	4,116	4,355	5.8%	69.7	71.9	3.1%	93.5%	Google Analytics
Social media	4,251	4,412	3.8%	77.5	79.1	2.1%	87.2%	Facebook
Design optimization	2,874	2,861	-0.5%	72.0	71.6	-0.6%	50.0%	Hotjar
Security	8,926	9,722	8.9%	99.8	99.8	0.0%	94.7%	Cloudflare
Native ads	4,229	4,024	-4.8%	84.9	84.5	-0.5%	21.7%	Taboola
CRM	6,408	6,119	-4.5%	98.2	98.0	-0.2%	.	Zendesk Chat
Privacy compliance	3,925	4,116	4.9%	83.8	86.5	3.2%	25.0%	TrustArc

Notes: Includes 26,127 sites which are scanned both pre-GDPR and one week post-GDPR. †Libert (2019) classification. ††Google ad platform includes Google Marketing Platform & Google Ad Manager.

so that even the dominant vendor (Google Marketing Platform/Ad Manager) has a relative share of only 14.5 although it reaches 78.3% of sites. We emphasize that our analysis investigates vendor shares of all website-vendor interactions since websites frequently engage multiple vendors per category.

Turning to the GDPR's short run impact on market structure, Table 4 shows that aggregate HHI increases 17.3% among all vendors and 17.8% among all classified vendors. The top four vendor categories represent 94.3% of categorized vendors pre-GDPR, and HHI increases post-GPDR in each of these categories. The advertising category sees the largest increase in HHI, growing 25.3% from 348 to 436 points. The increases in HHI among the next three top categories are more moderate: 2.3% in hosting, 5.8% in audience measurement, and 3.8% in social media. Beyond the top 4 categories, we see mixed results. Design optimization changes little (-0.5%), whereas HHI in security increases 8.9%. The native ads and CRM categories become less concentrated: HHI falls -4.8% and -4.5% respectively. Both categories are highly concentrated and so small that they represent only 1.1% of total categorized vendor reach. The increase in HHI in the advertising category (25.3%) is proportional to the decrease in the average number of vendors (24.1%), though this relationship is less than proportional in the remaining categories. Several categories see HHI increases near or above the 100 point threshold that American regulators use to scrutinize mergers: advertising gains 88 points, audience measurement gains 239 points, social media gains 161 points, and security gains 796 points.

As the total share of the top two firms, CR2 can be a more intuitive metric than HHI. In Table 4, we see that the sign of the short-run change in CR2 reflects the change in HHI in all categories but security, where the baseline CR2 of 99.8 creates a ceiling effect. In Appendix D.1, we provide the change in concentration ratios for different numbers of top firms and we see that the change in concentration ratios generally reflects the change in HHI until the concentration ratio exceeds 95% of the market. As with HHI, the largest increase

in CR2 is the advertising category with a relative increase of 15.8% from a CR2 of 18.7 to 21.7. For the remaining top 4 categories, the relative increase in CR2 lies between 1.7% and 3.1%. The decreases in CR2 for design optimization, native ads, and CRM are small at -0.6%, -0.5% and -0.2% respectively.

Finally, Table 4 shows the head-to-head win rate of the dominant firm in each category. Recall that this metric reflects the probability that a website keeps the dominant category vendor and drops a competitor post-GDPR, conditional on employing both vendors pre-GDPR. The top 4 categories suggest that the increase in concentration is in part a story of Google and Facebook's dominance. In advertising, Google Ad Manager wins an exceptional 98.9% of these head-to-head battles. Google also wins in hosting (Google APIs) 74.3% of the time and in audience measurement (Google Analytics) 93.5% of the time. For its part, Facebook wins 87.2% of its head-to-head battles in social media. Below the top four categories, sites tend to use a single category vendor and we see fewer than 75 head-to-head battles per category. We see that the dominant firm's win rate also helps to explain the change in HHI for smaller categories. Hotjar wins only half of its head-to-head battles in the design optimization category, which helps explain why the category's HHI is flat. In the security category, Cloudfare wins 94.7% of the time, which helps to explain why that category sees the second largest increase in HHI. Taboola wins only 21.7% of the 23 head-to-head battles in the native ads category, which helps to explain why that category sees a 4.8% reduction in HHI. Our concentration ratio and win rate results thus suggest that sites prefer to keep the dominant firm over alternatives.

GDPR long-run concentration impact We next examine the long-run impact by comparing concentration levels at the end of 2018 with the pre-GDPR period. Table 5 explores the change in concentration by category 27 weeks after the GDPR by replicating the calculations in Table 4 for the later time period. While aggregate HHI returns to baseline levels (0.6% higher), the aggregate HHI among vendors whose purpose is classified is still 3.9% higher than the baseline. The largest category (advertising) remains 6.3% more concentrated than the pre-GDPR baseline, while the next three categories see small decreases. Average vendors in the native ads category (0.07 vendors) remains lower than the pre-GDPR baseline (0.08 vendors), though the sign of the long-run HHI impact has reversed to +10.3% from -4.8% one week post-GDPR.

In sum, the GDPR coincided with a short-run increase in aggregate web technology concentration. While market concentration does not always follow a reduction in vendor use, the largest web technology categories become more concentrated. Many categories are highly concentrated initially and several categories exhibit significant increases in concentration, relative to both the underlying change in category use and the 100 point threshold that regulators use to scrutinize mergers. Three different concentration metrics paint a consistent picture of these results. In aggregate, short-run concentration effects appear to dissipate over the long run, though increased concentration in the advertising technology market persists to some extent. Our

Table 5: Long-run GDPR impact on concentration (27 weeks post)

	HHI			Concentration ratio (CR2)			Head-to-head competition	
Category	Pre	Post	Diff. (%)	Pre	Post	Diff. (%)	Win (%)	Dominant firm
All vendors	145	146	0.6%	9.8	9.4	-3.3%		
All categorized vendors†	307	319	3.9%	16.7	16.7	-0.2%		
Advertising	345	367	6.3%	18.7	19.1	2.3%	98.5%	Google ad platform††
Hosting	1,890	1,862	-1.5%	56.8	56.6	-0.3%	69.0%	Google APIs
Audience measurement	4,099	4,093	-0.2%	69.6	69.9	0.5%	93.0%	Google Analytics
Social media	4,258	4,103	-3.6%	77.4	75.4	-2.7%	86.1%	Facebook
Design optimization	2,880	3,009	4.5%	72.1	74.0	2.6%	65.8%	Hotjar
Security	8,936	9,426	5.5%	99.8	99.9	0.1%	90.2%	Cloudflare
Native ads	4,226	4,661	10.3%	85.1	87.9	3.3%	55.6%	Taboola
CRM	6,346	6,245	-1.6%	98.2	97.5	-0.6%	100.0%	Zendesk Chat
Privacy compliance	3,825	5,985	56.5%	82.9	92.4	11.4%	0.0%	TrustArc

Notes: Includes 25,561 sites which are scanned both pre-GDPR and 27 weeks post-GDPR. Small differences in pre-GDPR levels relative to Table 4 are explained by small differences in the set of scanned sites. †Libert (2019) classification. ††Google ad platform include Google Marketing Platform & Google Ad Manager.

model again suggests that this dissipation is consistent with publisher's declining beliefs about regulatory enforcement (α) in the absence of enforcement actions. Consistent with this explanation, Section 5.2 shows that the post-GDPR trend in vendor usage is negatively correlated with data regulation strictness. However, given the industry's dynamism and related potential for confounding trends, we do not wish to over-interpret the long-run changes in concentration.

6.2 Concentration extensions

In this section, we extend the concentration analysis. We focus on short-run effects throughout. We investigate how vendor use of personal data moderates concentration (Section 6.2.1) and whether user consent plays a visible role in concentration results (Section 6.2.2). In Section 6.2.3, we examine the contributions of Google and Facebook, as the leading web technology vendors, to market concentration post-GDPR.

6.2.1 Personal data concentration

We examine whether data minimization is accompanied by data concentration for personal data in particular. We wish to classify which vendor interactions contain personal information, though we do not directly observe this in our data. We classify web technology vendors as likely using personal data if they employ a cookie or if they are categorized as either "audience measurement" or "design optimization." While a cookie can contain non-personal information, cookies typically contain a user identifier, which is considered personal information under the GDPR. To maintain a constant classification, we classify vendors by whether they ever use a cookie pre-GDPR or one week post-GDPR. On the other hand, vendor interactions can still contain personal data without a cookie. For instance, Google Analytics relies on the unique user ID assigned

Table 6: GDPR & aggregate concentration: Three extensions

	HHI			
Data samples	Pre	Post	Diff.	Diff. (%)
A) Role of personal data				
Likely personal data	185.9	221.7	35.8	19.3%
Unlikely personal data	487.4	515.0	27.5	5.6%
B) Role of consent dialogs				
Sites with privacy extension	100.1	117.9	17.8	17.8%
Sites without privacy extension	153.6	179.4	25.8	16.8%
C) Role of top 2 companies (Google & Facebook)				
All vendors	145.7	171.0	25.2	17.3%
All but top 2 companies	46.0	43.2	-2.8	-6.2%

by the website, which is transmitted by a http request to `google-analytics.com`. Since the "audience measurement" and website "design optimization" categories require a user ID to function, we classify those vendors too as likely using personal data.

We find that data minimization led to increased concentration of personal data among top vendors. We split the vendor interactions in the data sample by whether they are likely to contain personal data or not and calculate the relative concentration using the aggregate HHI metric that ignores categories, as in Table 4. We report the results of this exercise in extension (A) in Table 6. We see that relative concentration increases 19.3% among vendor interactions that likely involve personal data, but only 5.6% among interactions that likely do not.[9]

The pool of online personal data thus became more concentrated in the hands of the largest vendors—another unintended consequence of the GDPR. To the extent that vendors can generate value from personal data, data concentration can further strengthen the market position of large vendors.

6.2.2 Consent dialogs

Past work has theorized a role for user privacy consent in increasing market concentration (Campbell et al., 2015). Under the GDPR, websites should obtain user consent for sharing personal data with vendors and list all these vendors. As a result, websites may reduce the number of listed vendors and favor large vendors familiar to consumers. Although sites may create their own systems to process user consent, many sites adopt third party consent management platforms (CMPs) in an effort to comply with the GDPR. Using data from `builtwith.com` described in Appendix C.1.1, we split the sample by whether sites implemented a known consent management platform (7.1% of sites) by the week after the GDPR. In extension (B) of Table 6, we report that the relative increase in concentration is similar whether the site implements a CMP

[9]Note that this result is not solely driven by ad vendors. When we recompute HHI for vendors that likely use personal data but exclude ad vendors, HHI still increases 12.0%.

(17.8%) or not (16.8%). We therefore conclude consent-related effects have limited influence on our market concentration results. In an August 2018 survey of 1,000 popular EU sites, Utz et al. (2019) show that no website lists vendors on the initial consent dialog and few list vendors on the secondary consent dialog. This practice all but eliminates the frictions associated with a user-facing vendor list. Thus, we show that increased concentration arises from business-to-business vendor choices rather than the consent dialog mechanism—a primary theorized anti-competitive mechanism.

6.2.3 Google & Facebook

We provide more evidence that Google-owned vendors and Facebook play an important role in increasing relative market concentration. These two companies dominate digital advertising, collecting 56% of global spending (WARC, 2019). In our pre-GDPR baseline, Google's many vendors represent 28.8% of all website-vendor pairs and Facebook represents 3.4%. The companies top the four largest web technology categories and Table 4 shows that sites keep the dominant vendor over a competitor the majority of the time post-GDPR.

Extension (C) of Table 6 compares the pre/post-GDPR change in aggregate HHI with and without the big two companies. The "all vendors" baseline replicates the all vendor aggregate HHI measures in Table 4, showing that relative vendor concentration among all vendors rises 17.3% from an HHI of 146 to 171. The final row excludes the big two companies from the HHI measures, revealing that relative concentration *falls* 6.2% in their absence (from 46 to 43). Note also that HHI is much lower without the big two companies, because the remaining vendors have smaller relative market shares even without the dominant companies. This difference arises because Google-owned vendors grow from 28.8% to 31.9% of site-vendor pairs in the short run and Facebook grows from 3.4% to 3.6%.[10]

Despite relative market share gains, the absolute position of the two companies is weaker after the GDPR. The fraction of all sample websites working with each top vendor falls one week post-GDPR: Google Marketing Platform/Ad Manager falls from 62.8% to 57.2% of sites, Google APIs falls from 55.6% to 50.9%, Google Analytics falls from 78.3% to 72.0%, and Facebook falls from 49.8% to 45.0%. Nonetheless, the Wall Street Journal used company fillings to suggest that both Google and Facebook's revenue grew faster than Europe's digital ad market in 2018 (Kostov and Schechner, 2019).

We note that sites were reluctant to drop Google or Facebook vendors despite the additional compliance cost that each entailed. Google and Facebook were among the companies that did not join the industry standard (IAB Europe, 2018a) for sharing user consent choices in 2018. Websites must transmit separate

[10]We also considered the role of each company separately. Excluding only Google still leads an HHI increase of 3.4%, highlighting that Facebook plays an important role. Excluding only Facebook reveals a 18.3% increase in HHI owing to Google's key role as a dominant vendor in multiple categories.

consent signals to each non-participating vendor. Google and Facebook's incompatibility decisions should make them less appealing to websites. Despite this, Table 4 shows that websites retained Google over another competing advertising vendor in 98.9% of such choices and websites retained Facebook over competing social vendors in 87.2% of the time. Google's ad offering may be less dependent on the IAB framework both because Google is the dominant ad vendor and because Google included several ad vendors under its aegis while ensuring GDPR compliance through contractual arrangements.

6.3 Concentration robustness

We consider the robustness of our short-run concentration findings to alternative definitions of both markets and market shares. Our ability to define markets relies on a mapping between domain names, vendors and vendor categories. To this point, we have relied on a domain database published by Libert (2019). In Appendix D.2, we consider a second third-party domain classification scheme by Karaj et al. (2018), which is associated with the "WhoTracksMe" project. WhoTracksMe defines the equivalent of the advertising and audience measurement/website analytics categories, but more broadly. The other defined categories have limited relationship to those defined by Libert (2019). Among semantically similar categories, we find that our concentration results are broadly robust to the WhoTracksMe classification.

We next consider two alternative definitions of market shares. Our reach-based definition follows the DoJ & FTC (2010) guidelines, but it treats all site-vendor links equally. We consider two alternatives that seek to treat site-vendor links more in line with their (unobserved) economic value and costs. First, we examine a "fractional-share" approach to market shares such that every website gets one "vote" and selected vendors (N total) each receive a share of $1/N$.[11] This definition offsets the tendency for multi-homing to depress the relative market shares of dominant vendors. Second, we examine a "traffic-weighted" approach that weighs site-vendor links by the site's traffic using data from Alexa. Reweighting indicates the economic importance of the vendors because sites with more traffic generally have both higher revenue and costs associated with vendors. The fractional-share and traffic-weighted approaches yield short-run increases of 19.2% and 17.0% respectively in HHI for all vendors, which resemble our original 17.3% estimate in Table 4. Appendix D.3 presents full category-level HHI results: we observe the same sign and similar magnitude short-run effect estimates in all cases but the traffic-weighted HHI for the smallest two categories.

[11] We thank Tesary Lin for suggesting this approach.

7 Conclusion

This paper provides novel empirical evidence of a potential tradeoff between privacy regulation and market concentration. We study the EU's GDPR, which serves as a global model for privacy policy. We examine the web technology industry, which attracts regulatory attention both for its permissive privacy practices and its high concentration. We examine over 27,000 top websites with a baseline of over 375,000 website-vendor ties. We show that websites reduce their web technology vendor use by 15% immediately after the GDPR enforcement deadline in response to the GDPR's data minimization mandate, but these compliance gains erode over time. In the short run, we see that concentration increases 17% in aggregate and in each of the top four web technology categories that together represent 94% of website-vendor ties: advertising, hosting, audience measurement, and social media.

In evaluating the GDPR, we see that the policy initially succeeded in its data minimization goal in the data-intensive web technology industry. Although these gains appear to erode over time, EU regulators criticize this industry's practices as non-compliant (Information Commissioner's Office, 2019; Autoriteit Persoonsgegevens, 2019) and have signaled heightened enforcement action in 2021 (Data Protection Commission, 2020; Commission Nationale de l'Informatique et des Libertés, 2020). Despite data minimization successes, the GDPR had the unintended consequence of increasing the industry's relative concentration. We show that this tradeoff between data minimization and concentration is not mechanical: some niche categories become less concentrated. However, relative concentration increases in the top web technology vendor categories that represent most of the industry. The increase in concentration is highest among the web technology vendors that process personal data, which the GDPR targets. Though requiring consumer consent could favor large vendors, the increase in concentration appears to be independent of consent. Instead, our evidence suggests that concentration increases because websites were more likely to drop smaller vendors. As policymakers wrestle with how to protect individual privacy, they should balance the risk of increasing concentration of personal data ownership and increasing market power (Gal and Aviv, 2020; Geradin et al., 2020). Further, the global nature of GDPR fines led to the unintended consequence that sites with the largest share of EU users do the least to reduce the vendors they use. Despite the GDPR's intent to harmonize regulation, country-level differences in compliance persist and are related to differences in local regulatory strictness. More research is needed to determine how policymakers could mitigate these unintended policy consequences.

The paper shows how market structure evolves post-GDPR, but ignores market conduct. Future research can further explore the consequences of greater relative concentration for market conduct such as vendor pricing. Vendor revenue and cost data would further elucidate the economic magnitude of the concentration effect. For instance, if ad vendors with greater reach are associated with greater ad revenue share, then

our findings could understate the increase in concentration. However, our theory model allows for sites to vary in terms of how much they use vendors, and the model still generates a post-enforcement increase in concentration. Our findings emphasize short-run changes in vendor use, which precludes any long-term competitive adjustment. Defining the market remains a central challenge for measuring market concentration. Our current classification scheme employs broad categories. For instance, the advertising category contains subcategories like ad exchanges, demand side platforms, and supply side platforms. From the perspective of evaluating the impact on competition, investigating these subcategories may provide further insight into the role of privacy regulation though the large vendors straddle multiple subcategories.

For regulators examining competition in technology industries, the GDPR presents nuanced effects. Most web technology vendors—including Google and Facebook—are worse off post-GDPR in that they lose website partners. However, the relative market shares of the largest vendors—particularly Google and Facebook—increase post-GDPR. This does not itself imply anti-competitive conduct by the large vendors. Rather, our evidence suggests that increased relative concentration results from website choices and not from vendor or user choices. We speculate that increased concentration could simply result from large vendors offering a better product or better compliance with the regulation.

References

Adjerid, I., A. Acquisti, R. Telang, R. Padman, and J. Adler-Milstein (2016). The impact of privacy regulation and technology incentives: The case of health information exchanges. *Management Science 62*(4), 1042–1063.

Allingham, M. G. and A. Sandmo (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics 1*(3-4), 323–338.

Aridor, G., Y.-K. Che, W. Nelson, and T. Salz (2020). The economic consequences of data privacy regulation: Empirical evidence from gdpr. *Available at SSRN*.

Autoriteit Persoonsgegevens (2019, December). Ap: veel websites vragen op onjuiste wijze toestemming voor plaatsen tracking cookies.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76*(2), 169–217.

Berry, S., M. Gaynor, and F. Scott Morton (2019). Do increasing markups matter? lessons from empirical industrial organization. *Journal of Economic Perspectives 33*(3), 44–68.

Brill, J. (2011). The intersection of consumer protection and competition in the new world of privacy. *Competition Policy International 7*, 7–313.

Campbell, J., A. Goldfarb, and C. Tucker (2015). Privacy regulation and market structure. *Journal of Economics & Management Strategy 24*(1), 47–73.

Clark, C. R. (2007). Advertising restrictions and competition in the children's breakfast cereal industry. *The Journal of Law and Economics 50*(4), 757–780.

Commission Nationale de l'Informatique et des Libertés (2019, June). Online targeted advertisement: what action plan for the CNIL?

Commission Nationale de l'Informatique et des Libertés (2020, October). Cookies et autres traceurs : la cnil publie des lignes directrices modificatives et sa recommandation.

Council of Economic Advisors (2016). Economic report of the president. Technical report.

Data Protection Commission (2020, April). Report by the data protection commission on the use of cookies and other tracking technologies. Technical report, Data Protection Commission.

Eckard JR., E. W. (1991). Competition and the cigarette TV advertising ban. *Economic Inquiry 29*(1), 119–133.

European Commission (2008, January). Flash eurobarometer 226: Data protection in the european union : Data controllers' perceptions. https://data.europa.eu.

European Commission (2019, July 24). Data protection rules as a trust-enabler in the eu and beyond – taking stock. Communication from the commission to the european parliament and the council, European Commission.

European Data Protection Board (2018, November). Guidelines 3/2018 on the territorial scope of the gdpr (article 3). Technical report, European Data Protection Board.

European Data Protection Board (2020, February). Contribution of the edpb to the evaluation of the gdpr under article 97. Technical report, European Data Protection Board.

Gal, M. S. and O. Aviv (2020, 05). The Competitive Effects of the GDPR. *Journal of Competition Law & Economics*. nhaa012.

Gallet, C. A. (1999). The effect of the 1971 advertising ban on behavior in the cigarette industry. *Managerial and Decision Economics 20*(6), 299–303.

Geradin, D., T. Karanikioti, and D. Katsifis (2020). Gdpr myopia: how a well-intended regulation ended up favouring large online platforms - the case of ad tech. *European Competition Journal*, 1–46.

Godinho de Matos, M. and I. Adjerid (2019). Consumer behavior and firm targeting after GDPR: The case of a telecom provider in europe. Working paper.

Goldberg, S., G. Johnson, and S. Shriver (2020). Regulating privacy online: An economic evaluation of the GDPR. Available at SSRN 3421731.

IAB Europe (2018a, March). GDPR transparency and consent framework.

IAB Europe (2018b, August). Header bidding and auction dynamics. White paper, Internet Advertising Bureau Europe.

Information Commissioner's Office (2019, June 20). Update report into adtech and real time bidding. Technical report.

Jia, J., G. Z. Jin, and L. Wagman (2019a). GDPR and the home bias of venture investment.

Jia, J., G. Z. Jin, and L. Wagman (2019b). The short-run effects of GDPR on technology venture investment. SSRN.

Johnson, M. S. (2020, June). Regulation by shaming: Deterrence effects of publicizing violations of workplace safety and health laws. *American Economic Review 110*(6), 1866–1904.

Kang, K. and B. Silveira (2018). Understanding disparities in punishment: Regulator preferences and expertise.

Karaj, A., S. Macbeth, R. Berson, and J. M. Pujol (2018). Whotracks.me: Monitoring the online tracking landscape at scale. *CoRR abs/1804.08959*.

Kostov, N. and S. Schechner (2019). Gdpr has been a boon for google and facebook.

Ledvina, A. and R. Sircar (2012). Oligopoly games under asymmetric costs and an application to energy production. *Mathematics and Financial Economics 6*(4), 261–293.

Lefrere, V., L. Warberg, C. Cheyre, V. Marotta, and A. Acquisti (2020). The impact of the gdpr on content providers.

Lerner, A., A. K. Simpson, T. Kohno, and F. Roesner (2016). Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In *25th {USENIX} Security Symposium ({USENIX} Security 16)*.

Libert, T. (2015). Exposing the hidden web: An analysis of third-party http requests on one million websites. *International Journal of Communication*.

Libert, T. (2019). https://webxray.org/.

Libert, T., L. Graves, and R. K. Nielsen (2018). Changes in third-party content on european news websites after GDPR.

Miller, A. R. and C. Tucker (2009). Privacy protection and technology diffusion: The case of electronic medical records. *Management Science 55*(7), 1077–1093.

Miller, A. R. and C. Tucker (2017). Privacy protection, personalized medicine, and genetic testing. *Management Science 64*(10), 4648–4668.

Miller, A. R. and C. E. Tucker (2011). Encryption and the loss of patient data. *Journal of Policy Analysis and Management 30*(3), 534–556.

O'Connor, J. (2019, March). https://verifiedjoseph.com/.

Peukert, C., S. Bechtold, M. Batikas, and T. Kretschmer (2020). European privacy law and global markets for data.

Phillips, N. (2019, July 27). Keep it: Maintaining competition in the privacy debate. Remarks for Internet Governance Forum.

Polinsky, A. M. and S. Shavell (2000, March). The economic theory of public enforcement of law. *Journal of Economic Literature 38*(1), 45–76.

Sanchez-Rola, I., M. Dell'Amico, P. Kotzias, D. Balzarotti, L. Bilge, P.-A. Verviker, and I. Santos (2019). Can I opt out yet? GDPR and the global illusion of cookie control. In *ACM ASIACCS 2019*.

Sass, T. R. and D. S. Saurman (1995). Advertising restrictions and concentration: The case of malt beverages. *The Review of Economics and Statistics 77*(1), 66–81.

Sharma, P., Y. Sun, and L. Wagman (2019). The differential effects of new privacy protections on publisher and advertiser profitability. *Available at SSRN 3503065*.

Sheffrin, S. and R. Triest (1991). Can brute deterrence backfire? perceptions and attitudes in taxpayer compliance. Working papers, California Davis - Institute of Governmental Affairs.

Shiller, B., J. Waldfogel, and J. Ryan (2018). The effect of ad blocking on website traffic and quality. *The RAND Journal of Economics 49*(1), 43–63.

Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature 57*(4), 904–54.

Sørensen, J. and S. Kosta (2019). Before and after GDPR: The changes in third party presence at public and private european websites. In *The World Wide Web Conference*, WWW '19, New York, NY, USA, pp. 1590–1600. ACM.

Urban, T., D. Tatang, M. Degeling, T. Holz, and N. Pohlmann (2020, June). Measuring the impact of the gdpr on data sharing in ad networks. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security (ASIA CCS '20)*. ACM.

US Department of Justice and the Federal Trade Commission (2010, August). Horizontal merger guidelines. http://www.ftc.gov/os/2010/08/100819hmg.pdf.

Utz, C., M. Degeling, S. Fahl, F. Schaub, and T. Holz (2019). (un) informed consent: Studying GDPR consent notices in the field. *arXiv preprint arXiv:1909.02638*.

WARC (2019). Internet in decline beyond google and facebook.

WhoTracks.me (2018). GDPR - what happened?

Yitzhaki, S. (1974). Income tax evasion: A theoretical analysis. *Journal of Public Economics 3*(2), 201–202.

Zhuo, R., B. Huffaker, k. claffy, and S. Greenstein (2019, October). The impact of the General Data Protection Regulation on internet interconnection. Working paper.

Appendices

A Theory model proofs

For all proofs, we assume $1 - \phi - 0.04 \geq 0$, or $\phi \leq 0.96$. This mild constraint on website profit margins ensures that $0 \leq 1 - \phi - 0.04\alpha \leq 1$.

A.1 Proposition 1

Websites maximize the objective function $\Pi^W(\vec{q})$ in equation (2), optimizing the cost/benefit trade-off of vendor services. Solving $\frac{\partial \Pi^W}{\partial q_i} = 0$ gives the inverse demand function for vendor i:

$$p_i(\vec{q}) = \tilde{\beta}\left[1 - q_i - \gamma \sum_{j=1}^{N} q_j I(i \neq j)\right] \tag{6}$$

Substituting the expression for prices into the vendor profit function (equation 3), we obtain:

$$\Pi_i^V(\vec{q}) = q_i(p_i - c_i) = q_i\left(\tilde{\beta}\left[1 - q_i - \gamma \sum_{j=1}^{N} q_j I(i \neq j)\right] - \delta i\right) \tag{7}$$

Vendors strategically compete in output à la Cournot. Vendor $i's$ best response function is then given by solving $\frac{\partial \Pi_i^V}{\partial q_i} = 0$ for q_i:

$$q_i^{BR}(\vec{q}) = \frac{1}{2}\left[1 - \gamma \sum_{j=1}^{N} q_j I(i \neq j) - \frac{\delta}{\tilde{\beta}} i\right] \tag{8}$$

Number of vendors and total vendor service demand Summing (8) over all vendors i gives an expression for aggregate vendor output (Q), holding the total number of vendors in the market (n) fixed:

$$Q(n) \equiv \sum_{i=1}^{n} q_i^{BR} = \frac{1}{2}\left[\sum_{i=1}^{n} 1 - \gamma \sum_{i=1}^{n}\sum_{j=1}^{n} q_j I(i \neq j) - \frac{\delta}{\tilde{\beta}} \sum_{i=1}^{n} i\right] = \frac{(1 - \frac{\delta}{2\tilde{\beta}})n - \frac{\delta}{2\tilde{\beta}} n^2}{2 + \gamma\,(n-1)} \tag{9}$$

Equation (9) should be interpreted as a family of candidate solutions for aggregate vendor output, where the number of firms n is allowed to vary. In reality, the n^{th} vendor will enter the market only if $\Pi_n > 0$ or equivalently if $q_n > 0$. The equilibrium number of active vendors n^* is obtained when aggregate vendor output is maximized, such that $\frac{\partial Q(n)}{\partial n}|_{n=n^*} = 0$. Ignoring integer constraints, the equilibrium number of

active vendors may be expressed as:

$$n^* = \frac{\sqrt{2\delta(2-\gamma)\left[(1-\gamma)\delta + \gamma\beta\left(1-\phi-0.04\alpha\right)\right]} - (2-\gamma)\delta}{\gamma\delta} \tag{10}$$

Aggregate vendor output with the equilibrium number of firms n^* is then given by:

$$Q(n^*) = \frac{2\gamma\beta\left(1-\phi-0.04\alpha\right) + \delta(4-3\gamma) - 2\sqrt{2\delta(2-\gamma)\left[(1-\gamma)\delta + \gamma\beta\left(1-\phi-0.04\alpha\right)\right]}}{2\gamma^2\beta\left(1-\phi-0.04\alpha\right)} \tag{11}$$

Vendor unit demand and market concentration measures Using equation (8) and noting $\sum_{j=1}^{N} q_j^* I(i \neq j) = Q(n^*) - q_i^*$, we can express vendor $i's$ equilibrium unit demand as:

$$q_i^* = \begin{cases} \frac{1}{2-\gamma}\left(1 - \gamma Q(n^*) - \frac{\delta i}{(1-\phi-0.04\alpha)\beta}\right) & i \leq n^* \\ 0 & i > n^* \end{cases} \tag{12}$$

With equilibrium vendor demand defined, vendor market shares s_i are then given by: $s_i = \frac{q_i^*}{Q(n^*)}$.

A.2 Proposition 3

The proof requires establishing that: (a) the number of active vendors is declining in GDPR enforcement probability (α), and (b) vendor market concentration increases in GDPR enforcement probability.

Part (a) requires that $\frac{\partial n^*}{\partial \alpha} = \frac{-0.04\beta(2-\gamma)}{\sqrt{2\delta(2-\gamma)((1-\gamma)\delta+\beta\gamma(1-\phi-0.04\alpha))}} \leq 0$, which holds for $0 \leq \gamma \leq 1$. We note that for $\phi \leq 0.96$, the radical expression in the denominator is positive since all terms in parentheses are positively signed.

To establish part (b), we show that the $CR(n)$ concentration ratio is increasing in α for $n < n^*$, i.e. that the cumulative market share of the top $n^* - 1$ (lowest cost, highest share) vendors increases in response to larger GDPR enforcement probability, or $\frac{\partial CR(n)}{\partial \alpha} > 0$ for $n < n^*$. We begin by noting that using equation (12) we can express the $n-$firm concentration ratio as:

$$\begin{aligned} CR(n) =& \frac{1}{Q(n^*)}\sum_{i=1}^{n} q_i^* = \frac{1}{(2-\gamma)\,Q(n^*)}\sum_{i=1}^{n}\left(1 - \gamma Q(n^*) - \frac{\delta i}{(1-\phi-0.04\alpha)\,\beta}\right) \\ =& \frac{n}{(2-\gamma)\,Q(n^*)}\left(1 - \gamma Q(n^*) - \frac{\delta(n+1)}{2\beta\gamma(1-\phi-0.04\alpha)}\right) \end{aligned}$$

Next we calculate $\frac{\partial CR(n)}{\partial \alpha}$, which is tedious due to the dependence of $Q(n^*)$ on α, but may be simplified to

the following expression:

$$\frac{\partial CR(n)}{\partial \alpha} = \left(0.08\beta\gamma^2\delta n\right)\frac{g(n)}{h(\cdot)} \quad where:$$

$$\begin{aligned} g(n) =& (2-\gamma)\left(2\beta\gamma(1-\phi-0.04\alpha)+\delta(\gamma(n-3)+4)\right) \\ & -(\gamma(n-2)+4)\sqrt{2\delta(2-\gamma)((1-\gamma)\delta+\beta\gamma(1-\phi-0.04\alpha)} \\ h(\cdot) =& (2-\gamma)\sqrt{2\delta(2-\gamma)((1-\gamma)\delta+\beta\gamma(1-\phi-0.04\alpha)} \\ & \cdot\left(2\sqrt{2\delta(2-\gamma)((1-\gamma)\delta+\beta\gamma(1-\phi-0.04\alpha)}-2\beta\gamma(1-\phi-0.04\alpha)+\delta\left(3\gamma-4\right)\right)^2 \end{aligned}$$

The constant term in the equation above is positively signed ($0.08\beta\gamma^2\delta n > 0$) as is the "denominator" function $h(\cdot) > 0$, given the model assumptions (i.e., $0 \leq \gamma \leq 1$, $\phi \leq 0.96$, $\delta > 0, n > 0$). Thus, we need to show $g(n) > 0$ for $n < n^*$. We can re-write the condition $g(n) > 0$ in terms of a condition on n, as follows:

$$g(n) > 0 \Longrightarrow n < \frac{\sqrt{2\delta(2-\gamma)\left((1-\gamma)\delta+\gamma\beta\left(1-\phi-0.04\alpha\right)\right)}-(2-\gamma)\delta}{\gamma\delta} = n^*$$

where the latter relation follows directly from equation (10). Since we have established $g(n) > 0$ for $n < n^*$ and that the sign of $\frac{\partial CR(n)}{\partial \alpha}$ is given by the sign of $g(n)$, we have established that $\frac{\partial CR(n)}{\partial \alpha} > 0$ for $n < n^*$.

A.3 Proposition 4

The proof requires establishing that: (a) the number of active vendors (n^*) increases in baseline website revenue (β), and (b) for higher revenue websites, increasing the perceived probability of GDPR enforcement (α) leads to a larger reduction in the number of active vendors than for smaller revenue websites.

Part (a) requires that $\frac{\partial n^*}{\partial \beta} = \frac{(2-\gamma)(1-\phi-0.04\alpha)}{\delta\sqrt{2\delta(2-\gamma)((1-\gamma)\delta+\gamma\beta(1-\phi-0.04\alpha))}} \geq 0$, which holds for $0 \leq \gamma \leq 1$ and $\phi \leq 0.96$.

To establish part (b), it suffices to show that the rate of change in the equilibrium number of firms (n^*) with respect to the GDPR enforcement probability (α) is decreasing in baseline website revenue (β). That is, $\frac{\partial}{\partial \beta}\left(\frac{\partial n^*}{\partial \alpha}\right) = \frac{\partial^2 n^*}{\partial \alpha \partial \beta} = \frac{-0.04(2-\gamma)^2\delta(\beta\gamma(1-\phi-0.04\alpha)+2(1-\gamma)\delta)}{2\sqrt{2}(\delta(2-\gamma)((1-\gamma)\delta+\beta\gamma(1-\phi-0.04\alpha)))^{3/2}} \leq 0$, which also holds under the model assumptions of $0 \leq \gamma \leq 1$ and $\phi \leq 0.96$.

A.4 Proposition 5

The proof requires establishing signs on the partial derivatives of the equilibrium number of vendors (n^*) with respect to γ, δ and ϕ are each decreasing.

We begin with the effect of increasing website product costs (ϕ):

$$\frac{\partial n^*}{\partial \phi} = \frac{-\beta(2-\gamma)}{\delta\sqrt{2\delta(2-\gamma)((1-\gamma)\delta+\gamma\beta(1-\phi-0.04\alpha))}} \leq 0$$

which holds under the model assumptions of $0 \leq \gamma \leq 1$ and $\phi \leq 0.96$. Next, we analyze the effect of increasing vendor service costs (δ):

$$\frac{\partial n^*}{\partial \delta} = \frac{-(2-\gamma)\beta\gamma(1-\phi-0.04\alpha)}{\delta\sqrt{2\delta(2-\gamma)((1-\gamma)\delta+\gamma\beta(1-\phi-0.04\alpha))}} \leq 0$$

which again holds under the model assumptions of $0 \leq \gamma \leq 1$ and $\phi \leq 0.96$. Finally, we analyze the effect of increasing vendor service substitutability (γ):

$$\frac{\partial n^*}{\partial \gamma} = \frac{4\sqrt{(2-\gamma)\delta\left((1-\gamma)\delta+\gamma\tilde{\beta}\right)} - 2\sqrt{2}\gamma\tilde{\beta} - \sqrt{2}(4-3\gamma)\delta}{2\gamma^2\sqrt{(2-\gamma)\delta\left((1-\gamma)\delta+\gamma\tilde{\beta}\right)}} \leq 0$$

This case is more involved, and requires obtaining a sign on the numerator in the equation above. Here, we leverage $\tilde{\beta} \equiv (1-\phi-0.04\alpha)\,\beta$ for notational convenience. Noting that $4\sqrt{(2-\gamma)\delta\left((1-\gamma)\delta+\gamma\tilde{\beta}\right)} - 2\sqrt{2}\gamma\tilde{\beta} - \sqrt{2}(4-3\gamma)\delta \leq 0 \implies 0 < 4\sqrt{2\delta(2-\gamma)\left((1-\gamma)\delta+\gamma\tilde{\beta}\right)} \leq \sqrt{2}\left(2\gamma\tilde{\beta} - (4-3\gamma)\delta\right)$. Squaring the latter relations and simplifying reduces to $0 \leq 2\gamma^2\left(\delta - 2\tilde{\beta}\right)^2$, which holds for all parameter values.

B Supplemental data description

B.1 Top vendors by category

We explore the top firms in the nine Libert (2019) vendor categories in Table 7 below, which lists the top five firms in each category as well as the number of vendors per category.

C Vendor usage robustness

C.1 Compliance concerns

C.1.1 Website compliance

The GDPR creates unusual challenges for empirical investigation because the regulation can potentially affect both the data collection process and the underlying data generating process (Goldberg et al., 2020).

Table 7: Web technology categories & top vendors

		Top Vendors				
Category†	Vendors	#1	#2	#3	#4	#5
Advertising	165	Google ad platform††	Xander	AdForm	The Trade Desk	Rubicon Project
Hosting	25	Google APIs	Google Tag Manager	Amazon Web Services	Cloudflare	Google Video/YouTube
Audience measurement	24	Google Analytics	Hotjar	ScorecardResearch	Adobe Audience Manager	Quantcast
Social media	11	Facebook	Twitter	AddThis	LinkedIn	Share This
Design optimization	8	Hotjar	New Relic	Optimizely	Visual Website Optimizer	Crazy Egg
Security	3	Cloudflare	Distil Networks	Knownsec		
Native ads	4	Taboola	Outbrain	nscontext.eu	ContentStream	
CRM	3	Zendesk Chat	liveperson.net	Salesforce		
Privacy compliance	3	TrustArc	Evidon	iubenda		

Notes: Vendor ranking based on pre-GDPR baseline. †Libert (2019) classification. ††Google ad platform includes Google Marketing Platform & Google Ad Manager.

Of particular interest is the role of user consent compliance and its potential to censor the automated collection of data. Since our data collection process does not interact with sites, all third-party domain (3PD) interactions in our data arise without consent, implying that consent compliance mechanisms could introduce empirical measurement issues.

We explore the role of consent compliance on our usage measures by augmenting our sample with data from `builtwith.com`, which tracks website adoption of known consent management platforms (CMPs). Using data from `builtwith.com` from August 2019, we find that only 24.4% of sites in our short-run sample at some point employ a third-party consent management platform, and only 7.1% of sites do so by July 4, 2018. However, regardless of CMP usage, sites in our sample do not appear to wait for consent before initiating interactions with third parties. We see no reduction in sites that load zero technology vendor content among sites using CMP's by July 4, 2018. Instead, we see no change (s.e. .0011) in sites with zero third parties – in other words, the change in vendors is all on the intensive margin. The surprising irrelevance of consent is supported by Sanchez-Rola et al. (2019), who found that 92% of the sites they scan in July 2018 set at least one identifier cookie without consent. Further, only 4% of sites provide a clear cookie opt-out option and only 2.5% of sites erase cookies after opting out. The Irish regulator found that all but one of the 38 sites it audited set cookies upon the user's arrival—even during the second half of 2019 (Data Protection Commission, 2020).

We further examine the incidence of sites blocking users as a compliance strategy, which could also have implications for measurement. Though some sites block EU users post-GDPR, this approach is rare in our data. We compare our sites to a list of 1,361 blocking sites compiled by O'Connor (2019), and only 13 of these sites appear in our list of 27,303 scanned sites. We found another 29 blocking sites in our data manually, by visiting sites with unusual post-GDPR changes. Notably, only 12 of the 42 blocking sites reduce vendors to fewer than 10 vendors by the week post-GDPR. Only 21 of these sites reduced their vendor usage at all in this period. The 12 blocking sites of interest include chicagotribune.com and latimes.com, which reduced vendors from 55 and 63 to a single vendor post-GDPR (their parent company's domain).

In sum, neither user consent compliance nor site blocking appears to be a significant concern for our data collection process in terms of observing site-vendor relationships.

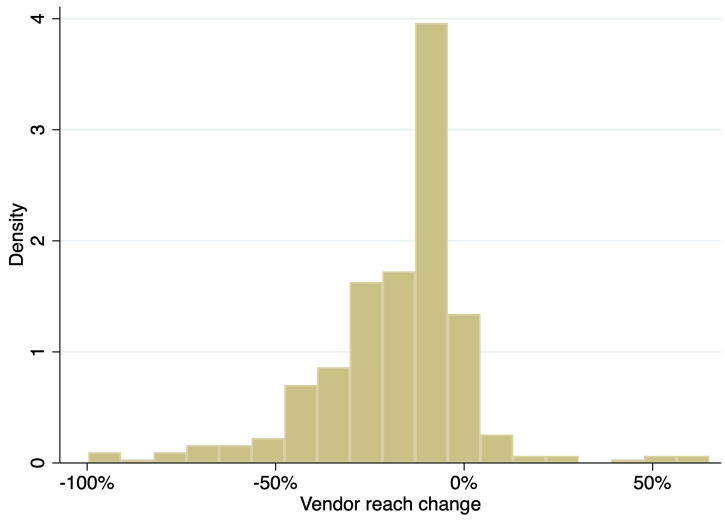
C.1.2 Vendor compliance

Our analysis assumes that the observed reduction in web technology vendors post-GDPR reflects website choice of vendors. Vendors experiencing their own technical issues with GDPR compliance or choosing to exit the EU market due to GDPR could also raise empirical measurement issues that would bias our results. For example, web technology vendors like Kargo, Verve, Drawbridge, and Factual—most of which provide

mobile ad services—indicated that they were exiting the EU as a result of the GDPR. If a sufficient number of such vendor exits actually occurred, it could produce usage outcomes observationally equivalent to websites actively choosing a smaller set of vendors.

To mitigate these concerns, we first establish that no vendor with high reach drops out of our data after the GDPR. With significant vendor exit ruled out, we present further evidence that our data reflects website choices. We do this by considering the results at the vendor level rather than the website level (as in the main results). Figure 4 presents a histogram for the short run change in reach by vendor, for the 361 vendors with baseline reach over 100 sites—0.38% of our site sample. Among these vendors, reach falls an average of -17.9% and a median of -12.7%. 92.8% of these vendors see their reach fall one week post-GPDR. The reduction in vendors is thus spread out among vendors, and no vendor of consequence removes itself from the site-vendor pair data. The largest vendor whose reach falls to 0 only has initial reach of 78 sites (0.30%). Focusing on sites with at least 50% of traffic from the EU, the largest vendor to exit these sites has reach of 101 sites or 0.54% (101/18,706) of majority EU sites in the short run panel. This pattern suggests that the websites rather than the vendors are the principal decision-makers in the market.

Figure 4: Distribution of short run change in vendor reach



Note: Restricts sample to vendors with initial reach of over 100 sites.

We also examine the entry, exit, and re-adoption of vendors at the end of 2018. To do so, we focus on the complete panel of websites that we scan pre-GDPR, one week post-GDPR, and 27 weeks post-GDPR. This panel includes a total of 371,167 initial site-vendor ties and 372,341 ties at the end of 2018. At the end of 2018, we find 3,283 vendor entrants with total reach of 7,526 sites (2.03% relative to initial ties) and 3,375 exiting vendors with total reach of only 7,272 sites (1.96 % of initial ties). In both cases, the majority of these vendors only reach a single site. Sørensen and Kosta (2019) show a similar pattern in third party

domains entry and exit, though they find a small net entry of 31 third-party domains. To examine vendor re-adoption, we consider the 91,004 total vendor ties that sites cut one week after the GDPR. We see that sites reestablish 29,491 (32.4%) of these vendor ties by the end of 2018, so that 61,512 ties (67.6%) remain severed. In sum, vendor re-adoption plays a minor role in the post-GDPR growth in vendor use whereas net entry plays almost no role.

C.2 Pre-GDPR trend in web technology vendor usage

We use external data from WhoTracks.me (2018) to examine the trend in web technology use prior to GDPR enforcement. WhoTracks.me (2018) released public data on the monthly web technology use of the 1,500 top EU websites beginning in March 2018. Karaj et al. (2018) describe the data collection methodology, which employs a large panel of consumers to measure third party domain usage on websites. Karaj et al. (2018) argue that their approach has several advantages, such as extensive sampling coverage of consumer browser and operating system set-ups, as well as the ability to view non-public websites (e.g., sites requiring user authentication). However, the WhoTracks.me (2018) data is dynamically selected because it evolves with the composition and preferences of its participating users, complicating site-level inference.

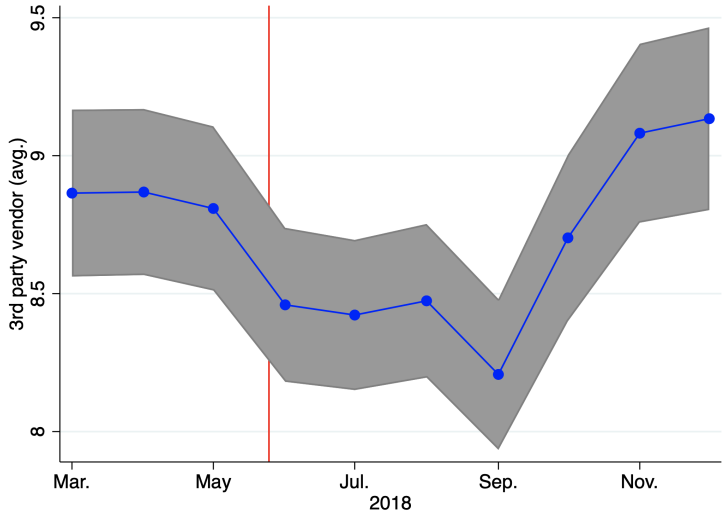
The WhoTracks.me (2018) data includes the third party vendors associated with the top 1,500 sites, as determined by their panel of users residing in the EU. Although these top sites vary over time, a complete panel is available for 1,452 sites between March 2018 and December 2018. We analyze the 1,322 sites (91.0%) that also appear in our sample data.[12] Karaj et al. (2018) also map third-party domains into vendors (referred to as "trackers") using their own database (see also Appendix D.2).

Figure 5 shows the evolution of average third party vendor usage for the 1,322 WhoTracksMe sites that are also in our data. Figure 5 plots the mean number of third party vendors by data collection month and 95% confidence intervals around each mean. The data reveal that third party vendor usage increased only 0.04% on average between March and April 2018. Third party usage on average fell -0.67% in May 2018, though this includes several days (May 25-31) after GDPR enforcement.[13] Thus, the WhoTracks.me (2018) data suggests a flat pre-trend in third party usage leading up to enforcement. Post-GDPR, Figure 5 shows a similar trend to Figure 1 as the number of vendors increases to pre-GDPR levels by the end of 2018.

To summarize, we do not find evidence of a pre-GDPR trend in web technology vendor use, which is consistent with sites delaying compliance until the May 25th enforcement deadline. This is supported by related work that finds a non-decreasing pre-trend Sørensen and Kosta (2019); Peukert et al. (2020). We conclude sites did not significantly reduce vendors before the GDPR deadline, which would underestimate

[12] Results using the full sample of 1,452 sites are essentially the same. See footnote 13.
[13] For the full sample of 1,452 sites, the number of third party vendors rises by only 0.08% in April and falls -0.63% in May.

Figure 5: 3rd party vendors pre-post GDPR: Sites in WhoTracks.me (2018) & our data



the effect of the GDPR on technology vendor usage. Further, to the extent that vendor usage exhibits an independent, increasing trend prior to GDPR (e.g., spurred by vendor innovation), we would again underestimate the effect of the GDPR: our model ignores vendor growth that would have occurred absent the GDPR.[14]

C.3 Alternative outcome measures

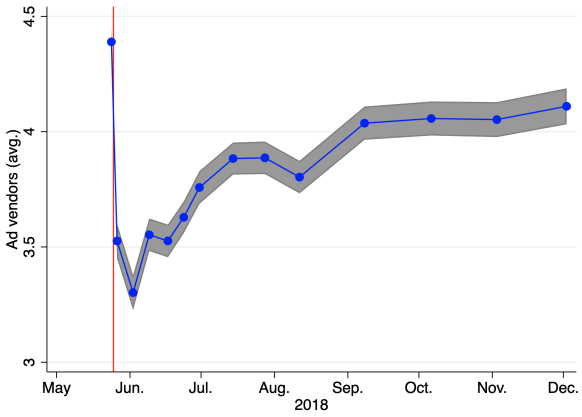
We consider alternative dependent variables for our analysis of vendor usage. The GDPR focuses on the use of personal information, and so we explore alternative measures of vendor use that are more likely to involve the exchange of personal information under GDPR. Here, we consider vendor use of third-party cookies as these likely contain a user identifier that the GDPR considers to be personal data. We also consider vendors in the advertising category because EU regulators have singled out this category for its use of personal data (Information Commissioner's Office, 2019; Commission Nationale de l'Informatique et des Libertés, 2019). Further, ad category vendors directly generate revenue for websites, so our model suggests that website choice of these vendors should be particularly responsive to the GDPR. Note these variables differ from our primary dependent variable, the total number of vendors: 50.3% of vendors use 3rd party cookies (Table 1) and 30.2% of website-vendor ties are classified as ad-related.

Figure 6 shows that both ad vendors and third party cookies follow the same sharp drop and subsequent recovery as exhibited by vendors in Figure 1. Sørensen and Kosta (2019) raise the possibility that the

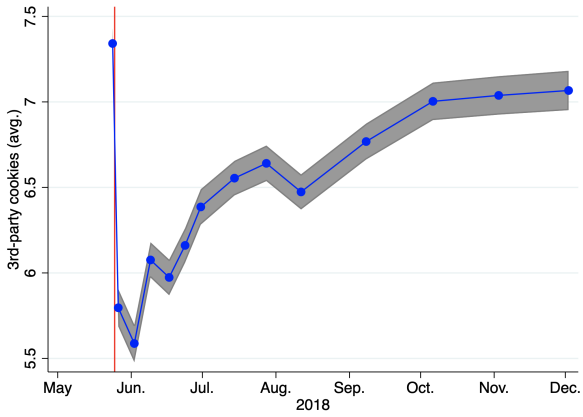
[14]Consistent with Equation (4), the GDPR's effect in some post-implementation period should be measured by comparing the observed post-GDPR outcomes to the model predicted outcomes for sites in the same period but absent GDPR. With an increasing usage trend, the counterfactual comparative baseline should be higher than the observed pre-GDPR usage level, leading GDPR effects to be underestimated.

Figure 6: Evolution of average vendor usage per website: robustness to alternate dependent variables

(a) Ad vendors



(b) Third-party cookies



post-GDPR reduction in ad vendor use could result from the adoption of server-side header bidding. This technology moves ad buying out of the browser, so would reduce the vendors observed by third party domain interactions. This explanation implies that we should observe a gradual reduction in ad vendors in our data. Figure 6a contradicts this prediction both because it shows an overnight drop in vendor use on the GDPR enforcement deadline and because adtech vendor use steadily increases post-GDPR. Though the macro trend could hide some growth in server-side header bidding (thereby reducing our visibility into website-vendor relationships), an IAB Europe report from August 2018 describes server-side header bidding adoption as in its "early stages."[15]

Figure 6b shows that vendor use of third-party cookies follows the sharp drop then recovery pattern. Third party cookies usually contain identifiers which the GDPR considers to be personal information. Thus, the GDPR apparently provided short-term and limited reduction in this personal information use. Figure 6b's pattern also contradicts a learning-to-comply story whereby websites dropped vendors until they learned to comply with the GDPR: third-party cookie use increased post-GDPR even for users who did not provide consent.

We explore robustness by replicating our analysis of vendor usage heterogeneity with respect to website characteristics (Section 5.2.2, as summarized in Table 3), but using the alternative dependent variables. Columns (1) and (2) present the results for the set of ad vendors and preserve the sign of the coefficients of Table 3. All coefficients are statistically significant at the 1% level with the exception of the interaction between the post-GDPR indicator and the share of EU users. This interaction coefficient is significant at the 5% level, though the coefficient was significant at the 10% level in Table 3. This result bolsters the increasing

[15] We thank an anonymous referee for alerting us to this issue.

relationship between EU shares and the post-GDPR effect on average vendor usage captured in Figure 7a as does the discontinuous relationship captured by the interaction between the post-GDPR indicator and the no EU user indicator. Columns (3) and (4) present the results for third-party cookies (at most one per vendor). Again the signs are preserved with the exception of the interaction between the post-GDPR indicator and the share of EU users, which is now statistically insignificant. Still, the interaction between the post-GDPR trend and the share of EU users is still positive and highly significant. The interactions between the post-GDPR trend and both the no-EU user indicator and the regulatory strictness measure are now only marginally significant, though the remainder remains significant at the 1% level.

Table 8: Heterogeneity in GDPR effect by website characteristics: robustness to alternate dependent variables

Dependent variable	Ad vendors		Third-party cookies	
Interactions time horizon	Short-run	Post-trend	Short-run	Post-trend
$GDPR_t$ (λ)	-2.589***		-4.104***	
	(0.208)		(0.302)	
$GDPR_t$ x $Week$ (ξ)		0.0036		0.0193**
		(0.0053)		(0.0076)
Interactions (ψ, ν):	$GDPR_t$ *x*	$GDPR_t$ *x Week x*	$GDPR_t$ *x*	$GDPR_t$ *x Week x*
log(Site rank)	0.196***	-0.0019***	0.347***	-0.0043***
	(0.023)	(0.0006)	(0.034)	(0.0009)
log(Ad count + 1)	-0.647***	0.0269***	-1.051***	0.0481***
	(0.041)	(0.0011)	(0.062)	(0.0016)
User income	-0.381***	-0.0073***	-0.542***	-0.0088***
	(0.037)	(0.0009)	(0.052)	(0.0013)
Share of EU users (%)	0.003**	0.0002***	-0.0003	0.0004***
	(0.001)	(3.70e-05)	(0.0021)	(5.45e-05)
No EU users	1.185***	-0.0128***	1.412***	-0.0118*
	(0.154)	(0.00430)	(0.212)	(0.0061)
Regulatory strictness	0.212***	-0.0030***	0.264***	-0.0027*
x >50% EU users	(0.041)	(0.0011)	(0.059)	(0.0016)
Constant (μ)	4.525***		7.549***	
	(0.031)		(0.0459)	
Site fixed effects	x		x	
Observations	329,158		329,158	
R-squared	0.798		0.796	

Note: Full panel except the data pull between the baseline & the 1 week post cross-section.
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

C.4 EU versus US user-facing vendor use

In this section, we leverage our sole data collection that visited sites both as a user outside of the EU (US) and as a user within the EU (France). We collected these data three months (11 weeks) after the GDPR enforcement deadline. We observe important differences in how sites treat EU versus non-EU users and show

these differences are largest for non-EU websites.

We find that websites in our sample employ fewer vendors on average for EU users than they do for US users after GDPR. To show this, we use regressions to compare each of the EU user and US user data collections to a common pre-GDPR baseline (which was collected using the VPN in France). As in equation (4), we augment this comparison of means with site fixed effects. The resulting GDPR estimates reveal a -0.893 (st. err. 0.073) average decrease in vendors for an EU user (replicating Figure 1) and a 0.475 (st. err. 0.083) average *increase* in vendors for a US user. Thus, a US user sees 1.37 more vendors than a EU user for sites in our data three months after the GDPR. This differential treatment provides evidence that the GDPR explains the post-May reduction in vendor use in Figure 1. In particular, sites prefer to engage more vendors, but the GDPR induces sites to treat users in the EU differently.

Figure 7 breaks down the GDPR effect estimates for a EU user (Figure 7a) and a US user (Figure 7b) by the website's share of traffic from the EU. We construct separate estimates as a function of a website's share of EU users, as measured by Alexa prior to the GDPR. To do so, we break sites up into ten even groups by their share of EU users and also group sites with no EU users—for a total of eleven groups. We then regress vendor count on a GDPR indicator with site fixed effects separately for each group again using the same pre-GDPR baseline. We plot the resulting estimates in Figure 7.

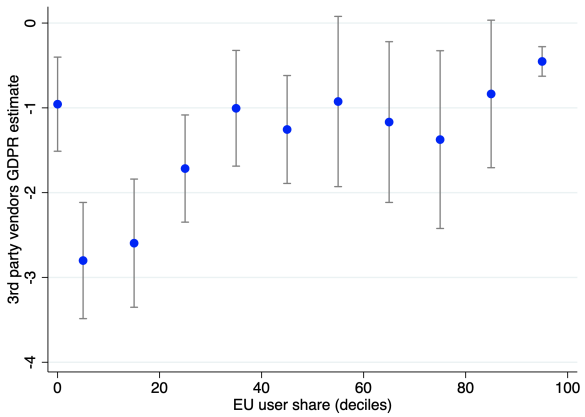
Figure 7a highlights the deterrent effect of the GDPR penalties on non-EU sites discussed in Section 5.2.2. We see that the point estimates are increasing in the share of EU users. That is, non-EU sites make the largest reduction in vendors. We also see a large discontinuity in this relationship for sites without users from the EU, which are outside the scope of the GDPR. Sites without EU users reduce vendors by -0.957 (s.e. 0.283) on average whereas as sites with over 0% but less than or equal to 10% EU users reduce vendors by -2.800 (s.e. 0.350). This vendor choice pattern non-parametrically illustrates the pattern in Table 3, whereas the latter shows that this relationship is robust to the inclusion of website covariates.

Figure 7b stands in contrast as the US user experience is the mirror image of the EU user experience. The US users see a large increase in vendors that is decreasing in the website's share of EU users. Figure 7 reveals stark differences in how minority EU traffic sites treat traffic from the US and the EU. Three months after the introduction of the GDPR, average vendor use by minority-EU sites rose by 1.678 (s.e. 0.188) for US users, but fell -1.892 (s.e. 0.146) for EU users. Minority-EU sites thus use 3.57 fewer vendors on average when serving EU versus non-EU users. This again highlights that sites can vary vendor choice by the user's location and that sites are willing and able to work with more vendors absent the GDPR.

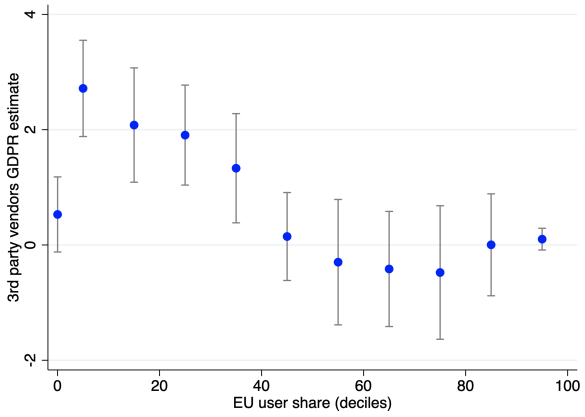
Our findings suggest any spillovers from the EU to US users dissipated by three months after the GDPR. Firms may opt to apply their GDPR changes to both EU and US users, which provides privacy benefits to US users outside the scope of the GDPR. Defining majority-EU sites as those sites with more than half

Figure 7: Differential treatments: 3 month GDPR effect on vendor use by user country

(a) EU user



(b) US user



of their traffic from the EU, we can see that majority-EU sites reduce vendors for EU users (-0.537, s.e. 0.084), but not for US users (0.041, s.e. 0.091). Consequently, we would understate the full effect of the GDPR if we only had data from users in the US—even on majority-EU sites. Figure 7 may lead some to conclude that US users on non-EU sites could serve as a control group. Nonetheless, those sites may not be representative of EU sites and may still experience GDPR spillovers. In particular, the null result for US users on majority-EU does not rule out any earlier GDPR spillovers to US users. Given the increasing trend in vendor use post-GDPR, US users could have seen a modest decrease in vendor use after the GDPR that eroded three months later.

Table 9: Short run change in concentration ratios

	Difference in Concentration Ratios						
Category†	CR1	CR2	CR3	CR4	CR5	CR8	CR10
Advertising	2.93	2.97	3.14	2.99	3.00	3.02	3.24
Hosting	0.43	0.97	0.96	0.58	0.28	0.26	0.18
Audience measurement	2.07	2.18	1.89	1.30	0.06	0.04	0.09
Social media	1.32	1.62	0.90	0.91	0.85	0.00	-0.01
Design optimization	0.42	-0.46	-0.39	-0.26	0.01		
Security	4.28	-0.04					
Native ads	-3.54	-0.47	-0.03				
CRM	-2.62	-0.19					
Privacy compliance	2.48	2.68					

Notes: †Libert (2019) classification.

D Vendor concentration robustness

D.1 Additional concentration ratios

To further unpack the changes in HHI, we examine the changes in concentration ratios. The concentration ratio is the total market share of the top M companies in the category. For instance, CR4 gives the total market share of the top 4 firms in the category. We see that the changes in CR(M) are typically the same sign as the change in HHI: this is by construction, but concentration ratios provide a more intuitive explanation.

In Table 9, we show the short run change in concentration ratios for the top {1, 2, 3, 4, 5, 8, 10} firms. Note that not all categories have sufficient firms to fill out the table.

D.2 Alternative categorization

We rely on an independent categorization of third-party vendors by Libert (2019). However, we also consider a second vendor categorization which we refer to as "WhoTracksMe", after the associated project developed by Karaj et al. (2018). The WhoTracksMe project is a large-scale monitoring initiative for online tracking. WhoTracksMe defines the equivalent of the advertising and audience measurement/website analytics categories more broadly so that these categories include an average of 6.9 and 1.9 vendors rather than 4.3 and 1.3 vendors respectively. The closest equivalent to the webxray hosting category is the content delivery network (CDN) and hosting categories in WhoTracksMe, which collectively have the same average size of 1.8 vendors per site. WhoTracksMe also has a social media category, but this excludes Facebook as WhoTracksMe instead classifies Facebook as advertising. The WhoTracksMe social media category thus has only 0.2 rather than the 0.8 vendors on average in Libert (2019). Note that the baseline average vendors is slightly lower under the WhoTracksMe classification: 14.1 vendors in Table 10 versus 14.4 in Table 4. This difference arises from defining uncategorized third-party domains as separate vendors and because WhoTracksMe categorizes

more of the vendors in the data.

Our concentration results are broadly robust to the WhoTracksMe classification in Table 10. The aggregate increase in relative HHI is higher for both all vendors (22.3%) and WhoTracksMe categorized vendors (23.5%). The WhoTracksMe classification results replicate the increase in HHI for advertising (29.0%), audience measurement/website analytics (6.7%) and hosting (5.1% for CDN and 0.1% for hosting). However, the social media HHI declines slightly (-0.5%), which confirms that Facebook plays a critical role in increasing HHI in the Libert (2019) social media category. As with the Table 4, the picture outside of these top categories is more mixed as several small vendor categories exhibit a decrease in HHI. Still, the top three categories here represent 84.6% of categorized vendors and the social media category represents an additional 1.8%.

D.3 Alternative market share definitions

We consider two alternative definitions of market shares that aim to capture the unobserved economic value and cost associated with site-vendor links. First, our fractional-share approach assigns credit for site-vendor links by vendor j's credit for site w as $credit_{jw} = 1/N_w$, where N_w is w's number of selected category vendors. For N total category vendors and W websites, the fractional share-based market share is formally:

$$s_j^{FS} = \frac{\sum_{w=1}^{W} 1/N_w \cdot I\left[link_{wj}\right]}{\sum_{k=1}^{K}\sum_{w=1}^{W} 1/N_w \cdot I\left[link_{wk}\right]}$$

where $I\left[link_{wj}\right]$ is an indicator function for the site-vendor link between w and j. For example, a site that engages two adtech vendors may split a share of total ad revenue between these vendors. As we do not observe the split, we assign equal shares to each vendor for a given site. Consequently, fractional-share-based market shares somewhat offset the reach-based market shares' tendency to depress the relative market shares of dominant vendors when sites multi-home. The fractional-shares approach also increases the relative share of vendors that are preferred by websites that use few vendors in the category. For instance, since Google Analytics receives a higher share under this definition because it it often appears alone or in combination with other audience measurement vendors.

Second, our "traffic-weighted" weighs site-vendor links by the site traffic metrics. Specifically, we use data from Alexa's Web Information Services on the site's estimated pageviews over the three months prior to the GDPR. We observe this traffic metric for 99.6% of sites in the short-run impact samples. The traffic-weighed market shares are then given by

$$s_j^{TW} = \frac{\sum_{w=1}^{W} views_w \cdot I\left[link_{wj}\right]}{\sum_{k=1}^{K}\sum_{w=1}^{W} views_w \cdot I\left[link_{wk}\right]}$$

Table 10: GDPR impact on concentration using WhoTracksMe classification

	Avg. vendors			HHI			Concentration ratio (CR2)			Head-to-head competition	
Category	Pre	Post	Diff. (%)	Pre	Post	Diff. (%)	Pre	Post	Diff. (%)	Win (%)	Dominant firm
All vendors	14.10	12.04	-14.6%	185	226	22.3%	13.7	15.8	15.6%		
All categorized vendors†	12.26	10.42	-15.0%	244	302	23.5%	15.7	18.3	16.2%		
Advertising	6.90	5.55	-19.6%	282	364	29.0%	18.2	21.9	20.7%	95.4%	Google ad platform††
Website analytics	1.85	1.64	-11.4%	1,903	2,030	6.7%	48.3	49.0	1.4%	94.1%	Google Analytics
Content delivery network	1.59	1.50	-5.6%	2,459	2,585	5.1%	67.6	69.8	3.2%	45.1%	Google APIs
Essential	0.65	0.62	-3.5%	5,041	4,611	-8.5%	76.4	72.6	-4.9%	67.9%	Google Tag Manager
Miscellaneous	0.28	0.24	-17.3%	485	403	-17.0%	22.6	19.0	-15.9%	27.6%	Walmart
Social media	0.22	0.20	-9.4%	3,106	3,091	-0.5%	62.5	62.4	-0.1%	75.9%	Twitter
Hosting	0.20	0.18	-9.9%	8,385	8,390	0.1%	94.9	94.9	0.0%	88.2%	Amazon Web Services
Customer interaction	0.20	0.17	-10.8%	366	361	-1.3%	18.2	16.8	-8.0%	20.0%	bidr.io
Audio-Visual player	0.19	0.16	-13.0%	4,070	3,998	-1.8%	68.9	67.3	-2.3%	64.9%	Google Video/YouTube
Comments	0.054	0.048	-10.1%	4702	4795	2.0%	92.0	91.7	-0.3%	0.0%	Yadro
Pornvertising	0.037	0.030	-17.9%	622	613	-1.4%	24.9	24.1	-3.3%	66.7%	exosrv.com

Notes: Includes 26,127 sites which are scanned both pre-GDPR and one week post-GDPR. †Karaj et al. (2018) classification. ††Google ad platform includes Google Marketing Platform & Google Ad Manager.

Table 11: GDPR impact on concentration using alternative market share definitions

Market share definition	Fractional-share			Traffic-weighted		
Metric	HHI			HHI		
Category	Pre	Post	Diff. (%)	Pre	Post	Diff. (%)
All vendors	334	398	19.2%	102	119	17.0%
All categorized vendors†	761	812	6.7%	212	250	17.9%
Advertising	2,371	2,717	14.6%	258	314	21.8%
Hosting	2,293	2,331	1.7%	1,870	1,911	2.2%
Audience measurement	6,001	6,194	3.2%	2,899	3,061	5.6%
Social media	5,505	5,675	3.1%	3,373	3,468	2.8%
Design optimization	2,997	2,981	-0.5%	2,521	2,462	-2.4%
Security	9,067	9,752	7.5%	8,687	9,255	6.5%
Native ads	4,559	4,280	-6.1%	4,689	4,421	-5.7%
CRM	6,413	6,124	-4.5%	3,974	4,038	1.6%
Privacy compliance	3,924	4,136	5.4%	4,821	4,685	-2.8%

Notes: Fractional-share anlysis includes 26,127 sites which are scanned both pre-GDPR and one week post-GDPR and traffic-weighted analysis includes 26,024 sites that also have Alexa pageview data. †Libert (2019) classification. ††Google ad platform includes Google Marketing Platform & Google Ad Manager.

where $views_w$ denotes site w's estimated pageviews. Reweighting approximates a vendor's economic importance because sites with more traffic generally have both higher revenue and costs associated with vendors. The distribution of pageviews is right-skewed such that the mean is about 10 times the median. As such, top websites dominate our traffic weighted market shares and ensuing concentration analysis.

Table 11 presents the HHI concentration metrics using both the fractional-share and traffic-weighted market shares. For all vendors, the relative HHI increases of 19.2% and 17.0% respectively are quite close to our preferred estimate of 17.3% in Table 4. Moreover, the signs of the category-level differences are the same as in Table 11 in all but two cases: CRM and privacy compliance for the traffic-weighted shares. The fractional-shares metric is notable because it creates greater baseline HHI than reach-based shares in all categories but privacy compliance. This metric better captures the importance of dominant firms in that the HHI is markedly higher in the advertising (versus 348 points for reach-based HHI), audience measurement (4,116), and social media (4,251) categories. In contrast, the HHI is lower for traffic-weighted than reached-based shares because large sites use more vendors, which decreases the traffic-weighted relative market shares.