Marketing Science Institute Working Paper Series 2022

Report No. 22-121

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

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

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

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

September 21, 2022

Abstract

We show that websites' vendor use falls after the European Union's General Data Protection Regulation (GDPR), but that market concentration also increases 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 are relatively more likely to retain top vendors, which increases the 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. Though the aggregate changes in vendor use and vendor concentration dissipate by the end of 2018, we find that the GDPR impact persists in the advertising vendor category most scrutinized by regulators. Our findings shed light on potential explanations for the sudden drop and subsequent rebound in vendor usage.

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, Ana Martinovici, 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 may 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 work provides novel, empirical evidence that is consistent with a tradeoff between privacy and competition policy.

Both privacy and competition concerns are acute in digital markets. These concerns are often at odds: a landmark joint statement between British competition and privacy regulators formally noted the tension between their two mandates in digital markets and pledged greater future cooperation (CMA & ICO, 2021). We study 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 (CNIL, 2019, ICO, 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 possible impact of privacy regulation 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 American federal level. Despite this policy momentum, a growing literature reveals unintended consequences of privacy policy. Privacy regulation 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., 2020, 2021), and reduced web traffic and revenue (Goldberg et al., 2021; Aridor et al., 2020; Schmitt et al., 2021).

To design privacy regulation, policymakers must balance consumer privacy concerns with the benefits of the data economy. If privacy regulation harms competition, for instance, this compounds concerns about market power in the economy (Council of Economic Advisors, 2016; Berry et al., 2019). To inform policy-maker's design choices, we consider the GDPR's consequences more broadly. Policymakers trade off the size of fines with the probability of levying a fine to ensure compliance (Polinsky and Shavell, 2000). The GDPR

emphasizes large fines—up to 4% of a firm's global revenue. This penalty design may influence the relative compliance efforts by domestic EU firms and by foreign firms that target EU customers, which are both subject to the GDPR. In the 2018 period that we consider, EU regulators issued few fines and the European Commission (2019) acknowledged that the regulation was under-enforced. In the industry we study, EU regulators did not issue fines until the end of 2020, though regulators repeatedly criticized the industry's compliance efforts (e.g., AP, 2019; DPC, 2020). Firms may also adjust their compliance strategies and efforts in response to this lack of enforcement. We provide descriptive evidence of consequences of these design and enforcement decisions for websites and their vendors.

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 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.

We find that websites restrict their use of web technology vendors after the GDPR, but that vendor market concentration also increases. We observe a steep 15% drop in website-vendor relationships one week post-GDPR, which return to pre-GDPR levels by the end of 2018. Three quarters of this drop occurs within a two-day window of the enforcement deadline. All vendor categories decline except for the privacy compliance category. Advertising-related vendors see the largest short-run decline (-24%), and remain below pre-GDPR levels at the end of 2018 (-6%). We emphasize the short-run estimates because they are robust to any confounding exogenous trend in vendor usage.

We find that relative market concentration increases 17% in aggregate a week after the GDPR, which implies that websites prefer to retain top web technology vendors. Note that we do not observe market conduct (e.g. pricing) of web technology vendors; we instead document changes to the industry's market structure. 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. As with usage, aggregate market concentration returns to pre-GDPR levels by the end of 2018. The advertising category again exhibits the greatest short run change—concentration increases 25.3%— and remains 6.3% more concentrated at the end of 2018. Concentration is pronounced among web technology vendors that process personal information, suggesting 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.

We consider differences in vendor usage by website characteristics, which illuminate the policy's incidence and help explain the post-GPDR rebound in vendor usage. We find that sites with ads and larger audiences use more vendors before the GDPR and make deeper cuts in the short run. We note that sites with a low, but positive share of EU users cut more vendors than either sites with a high share of EU users or sites without EU users (which are exempt from the policy). We conjecture that this may be a consequence of the GDPR's penalty design, which is calculated as a share of global rather than EU-specific revenue. We focus on two mechanisms that help explain changes in vendor usage over time. First, we show that sites increase their vendor usage when they adopt a privacy-related technology, though this explains less than 2% of the rebound. Second, we note that this is a dynamic sector where site vendor use has grown for over a decade, including in the year leading up the GDPR. Though we lack a clean control group to quantify this counterfactual growth, we note that usage in some vendor categories exceeded pre-GDPR levels by the end of 2018. In particular, the GDPR appears to benefit privacy-related vendors, whose usage more than doubles.

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.

We contribute to the growing literature on the economic consequences of the GDPR surveyed by Prasad and Perez (2020). Several more researchers note that website use of vendors fell after the GDPR (e.g., Lefrere et al. 2020; Lukic et al. 2021). Nevertheless, Zhuo et al. (2021) find no GDPR impact on Internet connectivity at the network level, reflecting the modest contribution of vendors to aggregate Internet data flows. 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. (2022) 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. (2022) rely on spillovers of the GDPR in

how websites treat US users to evaluate the GDPR. In addition to the above studies, the GDPR literature considers the impact on firm financial performance (Koski and Valmari, 2020; Chen et al., 2022), marketing communications (Godinho de Matos and Adjerid, 2022), mobile apps (Janssen et al., 2022), consumer online search (Zhao et al., 2021), as well as theoretical implications (Ke and Sudhir, 2022; Sharma et al., 2021).

Our analysis of website vendor choices under the GDPR also contributes to the literature on regulatory design. Privacy compliance is challenging to regulate because—unlike vehicle emission standards for instance—compliance can be multi-dimensional, subjective, and/or difficult to observe. Most compliance-related empirical studies examine the impact of enforcement actions on firm behaviors (Johnson, 2020; Kang and Silveira, 2021). We consider compliance in a data-intensive industry where strict compliance with the GDPR would impose high costs even relative to the GDPR's maximum fines. Relatedly, Gowrisankaran et al. (2022) show that policy uncertainty can have important consequences for firm decisions. Since EU regulators did not levy penalties on website vendors in 2018, websites were likely uncertain about how and how much to comply with the GDPR. We describe website compliance activities on two observable dimensions—websites use of vendors and privacy extensions—and show how these respond to the regulation's enforcement deadline and later evolve in the absence of enforcement.

The rest of the paper proceeds as follows. In Section 2, we review the web technology industry and the GDPR policy. Section 3 then describes our data. Section 4 discusses our empirical results for vendor usage and Section 5 presents our market concentration analysis. Section 6 considers website-level heterogeneity and potential mechanisms. Section 7 concludes.

2 Background: Web technology industry & the GDPR

As with other firms, websites depend on upstream vendors. Sites rely on technology vendors to monetize user traffic with ads, to load code and content elements like videos, and to measure and optimize user site visits. Sites vary in category, traffic, and audience, which affects their demand for vendor services. Further, the web technology sector is dynamic: aggregate vendor usage grew steadily since 1996, though individual vendor's market shares fluctuated greatly (Lerner et al., 2016).

The ties between sites and the vendors they engage are notable in two respects. First, sites instruct a user's browser to interact with the vendor's web domain in order for the vendor to perform its services. Website-vendor ties are therefore observable using specialized software to detect these third-party domain interactions. This provides the unusual opportunity to observe firm-vendor networks across many firms over time. Second, web tech vendors must process user personal data at some level to function. Vendors require user IP addresses to interact with a user's browser. Many vendors require a user identifier (e.g., stored via

browser cookie) to distinguish between a site's users. Some vendors further want a persistent, cross-site user identifier—e.g., to measure and target advertising.

The European Union's General Data Protection Regulation (GDPR) regulates the processing of personal data of EU residents. The GDPR Article 4(1) defines personal data broadly to include "online identifiers" like IP addresses and cookie identifiers. Under the GDPR, data processing refers to the collection, sharing, and use of data. 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 than 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. Firms must 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.

Though the GDPR lays out six legal bases for processing personal data, EU regulators maintain that consent is the most appropriate basis for websites and web technology vendors. Under the GDPR, valid consent 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. Prior EU regulator guidance indicated that websites should obtain user consent (Article 29 Data Protection Working Party, 2012) and EU regulators have maintained this stance after the GDPR's enforcement deadline (ICO, 2019; CNIL 2019).[1]

Websites face a dilemma in complying with the GDPR. Site compliance strategies involve some combination of limiting vendor ties and obtaining user consent. Nonetheless, sites must balance compliance against the benefits that vendors and personal data processing provide. Sites that generate revenue with advertising are particularly reliant on these benefits, as extant research shows that ad prices fall by half without cookie identifiers (Ravichandran and Korula, 2019; Johnson et al., 2020). To minimize data processing, sites may

[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 (ICO, 2019).

drop vendors. If sites prefer to drop smaller vendors—e.g., because they are lower quality—competition in the vendor market could be weakened. Sites also want to obtain high user consent rates in order to maximize the benefits that vendors deliver. However, sites may worry that obtaining user consent could introduce friction for site users—particularly given the GDPR's stringent consent requirements. Finally, a site's consent and vendor compliance strategies may interact in important ways. If a site lists all its vendors when obtaining consent, sites may choose fewer and more recognizable vendors to lift consent rates. If instead sites hide or omit their vendor lists, consent management may instead facilitate vendor adoption.

With the benefit of hindsight, our own and contemporary research reveal features of the industry's GDPR compliance during our 2018 data period. Sites reduced their vendor use post-GDPR (Libert et al., 2018; Peukert et al., 2022). Sites increased their use of privacy extensions to request consent and/or provide cookie notice (Hils et al., 2020). However, most websites seeking consent used a *de facto* opt-out approach: e.g., asking users to either click "OK" to accept data processing, or click "More Options" to opt out of specific vendors or data processing purposes (Utz et al., 2019). Moreover, sites interacted with third-party vendors and loaded cookies prior to obtaining consent (Sanchez-Rola et al., 2019; LINC, 2021). The present study provides additional evidence for the above (incomplete) compliance patterns.

Despite the GDPR's goal of harmonizing data protection in the EU, the GDPR is enforced by country-level regulators that vary in resources. Any EU country's data protection agency can open a case at their discretion or in response to a resident's complaint. Such cases are based on data processing that "substantially affects" (Article 4(23)) that country's residents. In cross-border cases, the GDPR's one-stop shop mechanism designates one country's regulator as the lead regulator based on the relevant firm's main establishment. Nevertheless, the other affected regulators retain considerable power under the GDPR's Article 60. Though the GDPR's Recital 9 expresses the need to standardize EU data protection law, important differences between country regulators persist. A survey by the central EU privacy regulator found that 21 of 30 country-level data protection authorities report they have insufficient human, financial, and technical resources (European Data Protection Board, 2020).

During our 2018 sample period, GDPR enforcement was limited. The European Commission (2019) one-year GDPR status report emphasized the need for greater enforcement for the GDPR to "become fully operational." Though EU regulators released multiple reports criticizing this industry's aforementioned practices (e.g. AP, 2019; DPC, 2020), regulators did not levy precedent-setting fines until the end of 2020. In the absence of enforcement, the industry established its own norms and practices for GDPR compliance. Two and a half years after the GDPR enforcement deadline, the French data regulator CNIL levied the first GDPR penalties on sites for non-compliant cookie practices. CNIL fined Google €100 million and Amazon €35 million in December 2020, the French website `lefigaro.fr` €50,000 in July 2021, and Google (again)

€150 million and Facebook €60 million in January 2022.

In sum, our setting carries four challenges for inferring the GDPR's impact. First, site vendor use is dynamic and has grown in aggregate over time, which may confound a long-run evaluation of the GDPR's impact. Second, the GDPR's broad geographic scope complicates finding a suitable control group of websites. Third, lack of enforcement may weaken sites' compliance incentives and may lead sites—rather than regulators—to determine industry compliance norms. Fourth, website compliance is multi-dimensional and includes both vendor use and consent practices. These practices may interact to limit or even facilitate expanding a site's vendor use. Our empirical analysis endeavors to make progress with these background challenges.

3 Data

3.1 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. In this vertical market, downstream websites rely on inputs from upstream 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 its homepage using 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 terms of

user traffic for 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.

Our data collection procedure resembles that of EU data regulators. The French regulator CNIL audits websites by recording their third-party domain interactions from the perspective of a French user. CNIL periodically collects panel data for the top 1,000 websites in France according to Alexa since at least January 2021 with a scan success rate of 88.9% (LINC, 2021). Similarly, the Irish regulator DPC audited 38 popular Irish sites by examining their third party domain interactions and consent practices (DPC, 2020).

3.1.1 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] The top categories by unique vendors in our data 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

[2]We re-label some categories. We treat "hosting" as a single category: i.e., our category analyses ignore its "general", "code", "font", and "video" subcategories. 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).

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." Web Appendix A.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.

3.1.2 Website characteristic variables

We gather data on website characteristics from several sources. We first create several static variables using pre-GDPR data whenever possible. We use Alexa Web Information Services to obtain data on both site traffic rank and share of users by country. We use the site traffic rank as a measure of site popularity, since traffic rank is inversely related to the volume of site traffic. We use data on the share of traffic originating from each country to construct the site's total share of EU users and to use as weights for other site-level variables. For instance, we construct a user income variable by combining these weights with country-level income per capita data from the World Bank. We also measure each site's reliance on advertising by counting their number of ads. Following Shiller et al. (2018), we visit each website homepage in August 2018 while using an ad blocker to count the number of blocked ads.

We construct a regulatory strictness variable that leverages the country-level nature of GDPR enforcement as discussed in Section 2. 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 site-level measure using the average of the four-point country-level survey measure, which we weight by the share of traffic by EU country. Our weighting procedure may sometimes misidentify the

relevant (expected) regulator, though jurisdiction is clear in domestic cases.[3] Websites are often located in the same country as their main audience, due to language and the local nature of content demand. For instance, we find that 87.2% of traffic with from sites with ".fr" domains originates from French users. More generally, 91.5% of traffic is domestic for similar sites whose top-level domain specifies an EU country (54.0% of our site sample). Finally, our analysis on this dimension only considers majority EU user sites as we expect these sites will have a clearer sense of their (expected) EU regulator and its relative strictness.

We collect two dynamic site-level variables in addition to the static variables discussed above. To complement our regulatory strictness measure, we collect a measure of the GDPR's saliency. Specifically, we collect search volume on the topic of the GDPR by week and by EU country using Google Trends. To create the site-level search variable, we again weight this data using the EU country traffic shares. To examine websites' consent-related compliance activities, we also collect data on website use of a privacy extension over time. Websites work with privacy extension providers to add boilerplate site pop-ups that either provide cookie notice or manage user consent. We obtain this data from BuiltWith, which regularly scans websites to identify the technologies they use. Note that some privacy extension providers load content from their own or other third-party domains, so are recorded as vendors in our data. However, the BuiltWith data provides a wider view by including privacy extensions that do not load third-party content.

3.2 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 (baseline) cross-section* corresponds to the first time observation of websites and serves as our only pre-GDPR scan. We use this baseline to evaluate changes in vendor use after the GDPR. We compare the baseline cross-section to the *1 week post-GDPR (short run) cross-section* as well as the *27 weeks post-GDPR (long run) 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. We then report the associated number of 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

[3]In cross-border cases, one concern is that some firms strategically locate their headquarters (e.g., in Ireland) to ensure a more favorable lead regulator. In particular, Google and Facebook are two such firms, though they have negligible shares in our website-level analysis. Furthermore, major tech firms have not avoided enforcement by locating in Ireland: France has leveled large fines against Amazon, Facebook, and Google and Ireland leveled large fines against Facebook.

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 traffic 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
Ad count	25,889	4.1	7.6	0	1.0	150.0
User income ($ thousands)	26,331	33.0	17.9	1.6	27.8	116.6
Regulatory strictness (4-point index)	24,942	2.6	0.4	1.7	2.6	3.1
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

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%. Sites have a median of 1 ad on the page and an average of 4.1 ads. Average user income is $33 thousand. Average regulatory strictness is 2.6, which is near the middle of its 1 to 4 survey scale. Note that we summarize the time-varying site characteristics—privacy extension use and GDPR-related search volume—later in Figures 5 and 6.

The remaining 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 post-GDPR, which mitigates concerns that GDPR compliance interferes with our measurement procedure. 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 a short-

run decline 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.

4 Vendor usage analysis

We empirically analyze this vertical market by first examining how downstream websites' use of technology vendors changed after the GDPR. We outline our empirical approach in Section 4.1. We then examine the post-GDPR changes in website vendor usage in aggregate (Section 4.2.1) and by vendor category (Section 4.2.2). We proceed to analyze concentration in the upstream vendor market in Section 5. We examine usage differences by websites characteristics and discuss potential mechanisms in Section 6.

4.1 Empirical approach

We analyze how the number of vendors engaged by a website evolves after enforcement of the GDPR using the following fixed effects regression:

$$y_{it} = \mu + \theta_i + \sum_{t>0} \lambda_t \cdot GDPR_t + \varepsilon_{it} \tag{1}$$

where y_{it} is website i's number of technology vendors during week t, which ranges from 0 (week pre-GDPR) to 27 (weeks post-GDPR). $GDPR_t$ is an indicator for the post-GDPR enforceability date, θ_i is a site fixed effect, and ε_{it} is the error term. We use website fixed effects to model unobserved, time-stationary characteristics that are potentially correlated with the site's response to the GDPR, where these fixed effects are normalized to be mean zero across websites ($E_i[\theta_i] = 0$). The coefficients λ_t therefore capture the average difference in the number of web technology vendor usage (t weeks post-GDPR) relative to our pre-GDPR baseline (μ) conditional on static, unobserved website characteristics. Our estimates of λ_t therefore provide a descriptive summary of the average change in website vendor usage post-GDPR.

4.1.1 Discussion

The conditions required to interpret the estimates of λ_t as causal effects of the GDPR ($E[GDPR_t \cdot \varepsilon_{it}] = 0$) are strong. We lack a valid control group, so we instead rely on before-after comparisons like other GDPR studies of this industry (Libert et al., 2018; Sørensen and Kosta, 2019; Peukert et al., 2022). We must therefore grapple with potential pre-trends or post-trends and how these affect our interpretation of λ_t. We are most concerned about a probable exogenous post-GDPR trend in vendor usage that is unrelated to the GDPR. If such trends are present, the bias in causally interpreting the λ_t coefficients would increase over

the post-GDPR period. For this reason, we anticipate our λ_t estimates to most credibly reflect the causal effect of the GDPR in the short run.

Lack of a valid control group 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 (in addition to EU users) to reduce administrative costs or enforcement scrutiny. We show that non-EU websites implement GDPR measures for EU users (Section 6.1) and we show that websites expose EU users to fewer vendors than non-EU users (Appendix A.4). We conclude that neither non-EU sites nor non-EU users represent a clean control group. We therefore favor pre- vs. post-GDPR differences to evaluate the GDPR in our setting.

Pre-trend in vendor usage 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. We begin by differentiating between exogenous (unrelated to the GDPR) and endogenous pre-trends. Aggregate vendor usage has grown over time (Lerner et al., 2016) and grew 6% in the year leading up to the GDPR enforcement deadline (Peukert et al., 2022). An exogenous pre-trend is not a serious problem because we collect data right before the enforcement deadline. Note that this timing is ideal: it represents the end of that exogenous pre-trend, so that the pre-trend's confounding influence is minimal. However, an endogenous pre-trend would confound our estimates. In particular, we are concerned that sites may have changed their web technology vendors before the GDPR deadline. Such anticipatory compliance would generate a downward, endogenous pre-trend in vendor use immediately prior to the enforcement deadline. In this case, our short-run estimates λ_t would *understate* the effect of the GDPR and would therefore provide a conservative estimate of the policy's impact. In Appendix A.2, we leverage external data and consider related research to discuss vendor usage prior to the enforcement deadline. This evidence suggests an increasing or stable pre-trend, which is inconsistent with material anticipatory compliance.

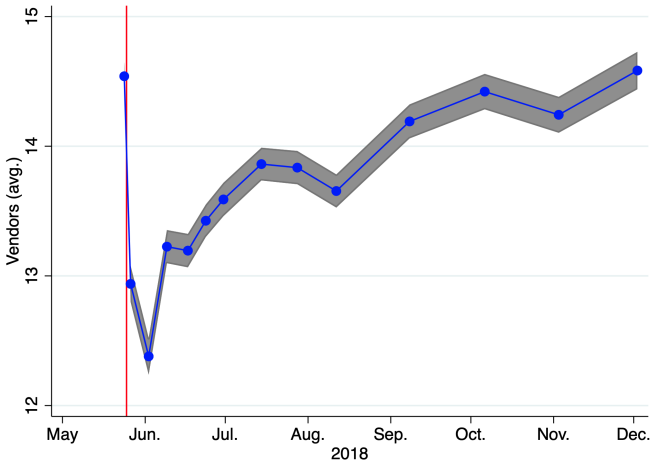
Post-trend in vendor usage We note above that aggregate web technology vendor usage grew over time including in the year that led up to the GDPR. In the absence of a clean control group, we do not observe the counterfactual growth rate post-GDPR. However, industry history suggests a continued positive trend in vendor usage is probable. An exogenous growth trend would mean that our λ_t estimates underestimate the effect of the GDPR, and the extent of this bias would increase in the time elapsed after the enforcement

deadline. Thus, the short-run λ_t estimates best represent (conservative) estimates of the effect of the GDPR, as independent usage trends have the least potential to conflate the GDPR's impact.

4.2 Changes in vendor usage

4.2.1 Aggregate usage

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



We use our complete dataset to estimate equation (1) 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 became enforceable 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 *short-run* estimate. The short-run estimate shows that sites reduce web technology vendors 14.9%, from an average of 14.5 to 12.4 vendors.[5] A model with site-level fixed effects therefore leads to a slightly higher short-run estimate than the simple means comparison in Section 3.2 (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. The GDPR applies to vendors who use

[4]Figure 1 plots the average web technology vendors pre-GDPR ($\hat{\mu}$) and for subsequent weeks ($\hat{\mu} + \hat{\lambda}_t$), where confidence intervals use the robust standard errors for μ and λ_t respectively. As our $\hat{\mu}$ estimate, we use the mean of vendor use for the set of sites that we scan pre-GDPR.

[5]If sites wait for consent before instructing the browser to interact with a third-party domain, consent compliance mechanisms would censor our data. Nonetheless, this does not appear to be material in our data. We estimate a small and statistically insignificant short-run decrease (-0.002 sites) in sites that load no vendors among the 14.4% of this site sample that use a privacy extension (as detected by BuiltWith five weeks post-GDPR). This result 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 (AP, 2019; DPC, 2020).

personal data (e.g., cookies) and EU regulators have scrutinized ad technology vendors in particular. In Appendix A.3, we show that the number of advertising technology vendors and the number of vendors that use third party cookies both resemble Figure 1 over time.

One of our starkest findings is that the short-run GDPR impact appears to erode over time. By the end of 2018, the average vendor usage effectively returns to its pre-GDPR level, though with three important caveats. First, if vendor use continued to grow by about 6% as it had in the year leading up to the GDPR (Peukert et al., 2022), average vendor use at the end of 2018 would instead be about 15 vendors—in the absence of the post-GDPR drop. Second, a small component of the post-GDPR growth vendor use is in the privacy compliance category, and is therefore a mechanical consequence of the GDPR. Since our pre-GDPR vendor classification (see Table 2) excludes entrants in the privacy category (see also Hils et al., 2020), we manually identify 17 additional privacy-related vendor domains among sites that use a privacy extension. Once we remove the growth of 0.081 in our expanded set of privacy-related vendors, non-privacy-related vendors usage is actually 0.036 *lower* at the end of 2018 than the pre-GDPR baseline. Third, in interpreting the pattern in Figure 1, we again caution against attributing the λ_t estimates to the GDPR alone—particularly for the later estimates.

With these factors in mind, we posit that the post-GDPR growth in vendor use arises from a combination of (at least) two mechanisms: (1) evolving GDPR compliance norms (including consent management), and (2) exogenous vendor innovation and market dynamics. We discuss supporting evidence for each of these mechanisms in Section 6.2.

4.2.2 Usage by vendor service category

We characterize vendor usage by service category, lending insight into the GDPR's impact on different types of web technology vendors. Our analysis here links directly to our discussion of market concentration (Section 5.2), which relies upon the same categorization scheme. Table 2 reports the short-run (1 week post) and long-run (27 weeks post) post-GDPR change in vendor use for each category relative to the pre-GDPR baseline.

Table 2: Post-GDPR change in average vendor use by category

	Pre-GDPR†	Short run (1 week post)			Long run (27 weeks post)		
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***	0.000	-9.7%	-0.001	0.001	-3.7%
Privacy compliance	0.02	0.004***	0.001	22.9%	0.020***	0.001	123.6%

Notes: Coefficient estimates from separate fixed effect regressions (see equation 1) for each category. †Pre-GDPR baseline given by means of scanned sites. Robust standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The second 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 (1)) corresponding to each vendor category outcome. We see that web technology vendors overall again fall 14.4% and the subset of categorized vendors falls 17.6% from 8.45 to 6.97. 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. Nevertheless, 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 (24.1%), from 4.39 to 3.33 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 by the end of 2018, compared to the measured short-run differences. As with the short-run estimates, the change in vendor use varies substantially by category. In particular, the advertising and native advertising categories remain below pre-GDPR levels while hosting and especially privacy compliance categories exceed pre-GDPR levels.

In sum, both our aggregate and category-level estimates present a pattern of results that is broadly consistent with a GDPR impact. The short-run drop is sharp, with 74% of the 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 prior to the GDPR either (Peukert et al., 2022;

Sørensen and Kosta, 2019). Vendor usage falls in all categories but privacy compliance. This category-level pattern is consistent with a GDPR impact rather than a technology change or some other transitory shock. Although our long-run estimates have greater bias concerns for causal interpretation, we note the persistence of reduced vendor usage among advertising-related categories and increased usage in privacy. Thus, our long-run heterogeneity results reveal the aggregate rebound in vendor usage is not uniform, but is correlated with vendor categories that either invite regulatory scrutiny or stand to benefit from compliance requirements.

5 Vendor concentration analysis

We lay out our empirical approach to measuring vendor market concentration in Section 5.1. We present our aggregate and vendor-category level concentration results in Section 5.2. We extend our concentration analysis in Section 5.3.

5.1 Empirical approach

We now turn to our analysis of concentration in the upstream vendor market, and therefore shift our unit of analysis from websites to vendors. 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.[6] With regard to market definitions, we consider the industry in aggregate as well as category-level markets derived from the vendor classification discussed in Section 3.1.1.

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

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 concentration measures only respond to changes in market share, not in the overall market size. As was demonstrated in Section 4.2.2, 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.

5.2 Changes in vendor market concentration

5.2.1 Aggregate concentration

We first consider the web technology industry as a single market by including all vendors. Figure 2 plots the evolution in market concentration over 2018, as measured by aggregate HHI.[7] Aggregate HHI is 146 points

[7] To account for changes in scanned sites over time, Figure 2 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 3 as the baseline. Note too that these sample differences explain the small differences in baseline HHI and CR2 in Tables 3 and 4.

Figure 2: Evolution of web technology vendor concentration (HHI)

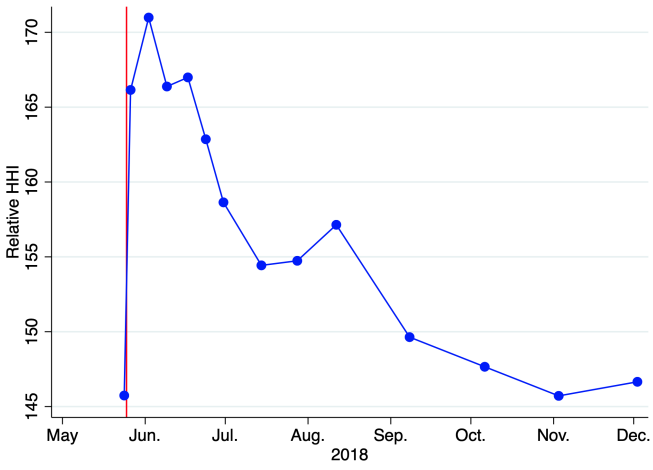


Table 3: Short-run post-GDPR change in concentration (1 week post)

	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.

before the GDPR and HHI reaches its maximum of 171 points one week post-GDPR, a 17.1% increase. We note that the evolution of concentration is effectively a mirror image of the evolution of vendor usage in Figure 1. As such, Figure 2 provides empirical support for market concentration increasing in response to the GDPR's implementation. As with average vendor use, Figure 2 reveals that web technology concentration returns to its pre-GDPR level by the end of 2018, for the same potential reasons discussed in Sections 4.2 and 6.2.

5.2.2 Concentration by vendor category

We next examine the post-GDPR changes in market concentration by vendor category. As with our vendor usage analysis, we report category-level point estimates for the short run (1 week post) and long run (27 weeks

post). For the reasons outlined in Section 4.1.1, we emphasize the short-run results, but also summarize our findings over the long-run horizon.

Table 3 reports 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, 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. Both multi-homing and defining a category broadly may lead our absolute concentration measures to overstate the competitiveness of some categories. However, we emphasize that our analysis investigates *changes* in concentration metrics after the GDPR rather than the levels themselves.

Short-run change in concentration Turning to the short-run changes in market structure, Table 3 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.

CR2—the combined market share of the top two firms—can be a more intuitive metric than HHI. In Table 3, 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 Web Appendix B.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 3 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.

We consider the robustness of our short-run concentration findings to alternative definitions of both markets and market shares. In Web Appendix B.2, we consider a second third-party domain classification scheme by Karaj et al. (2018). Among semantically similar categories, we find that our concentration results are broadly robust. In Web Appendix B.3, we consider two alternative definitions of market shares. 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 $1/N$. 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. Our concentration findings are again broadly robust to these alternative market share definitions.

Long-run change in concentration For completeness, we next examine a longer time horizon by comparing concentration levels at the end of 2018 with the pre-GDPR period. Table 4 explores the change in concentration by category 27 weeks after the GDPR by replicating the calculations in Table 3 for the later time period. While aggregate HHI returns to baseline levels (0.6% higher), the aggregate HHI among vendors

Table 4: Long-run post-GDPR change in 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 3 are explained by small differences in the set of scanned sites. †Libert (2019) classification. ††Google ad platform include Google Marketing Platform & Google Ad Manager.

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 difference reverses to +10.3% from -4.8% in the short run.

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. As with vendor usage, we do not wish to over-interpret the long-run changes in concentration in light of the industry's dynamism and related potential for confounding trends.

5.3 Concentration extensions

In this section, we extend the concentration analysis. We focus on short-run changes throughout, though Web Appendix B.4 examines long-run changes. We investigate how vendor use of personal data moderates concentration (Section 5.3.1) and whether user consent plays a visible role in concentration results (Section 5.3.2). In Section 5.3.3, we examine the contributions of Google and Facebook, as the leading web technology vendors, to market concentration post-GDPR.

Table 5: GDPR & aggregate concentration: Three extensions

	HHI			
Data samples	Pre	Post	Diff.	Diff. (%)
A) Role of vendor personal data use				
Likely use personal data	185.9	221.7	35.8	19.3%
Unlikely use personal data	487.4	515.0	27.5	5.6%
B) Role of consent dialogs				
Sites with privacy extension	126.2	151.2	25.0	19.8%
Sites without privacy extension	151.6	176.7	25.1	16.5%
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%

5.3.1 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 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 3. We report the results of this exercise in extension (A) in Table 5. 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.[8]

Our results suggest that the pool of online personal data thus became more concentrated in the hands of the largest vendors. This is a second apparent 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.

[8]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 by 12.0%.

5.3.2 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. Some sites adopt privacy extensions as part of their GDPR compliance efforts. As described in Section 3.1.2, we split the sample by whether sites implemented a known privacy extension (14.4% of sites) according to BuiltWith. As BuiltWith scans longer tail sites less often, we sort sites by whether BuiltWith detects a privacy extension five weeks after the enforcement deadline.[9] In extension (B) of Table 5, we report that the relative increase in concentration is similar whether the site implements a privacy extension (19.8%) or not (16.5%) and a 25-point increase in both cases. We therefore conclude that 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.

5.3.3 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%. These companies top the four largest vendor categories and Table 3 shows that sites usually keep the dominant vendor over a competitor post-GDPR.

Extension (C) of Table 5 compares the short-run change in aggregate HHI with and without the big two companies. The "all vendors" baseline replicates the all vendor aggregate HHI measures in Table 3, 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]

[9] Our results are robust to earlier or later cut-off dates.

[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.

Despite relative market share gains, the absolute positions of the two companies are 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 consent signals to each non-participating vendor. Google and Facebook's incompatibility decisions should make them less appealing to websites. Despite this, Table 3 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's consent 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 Website heterogeneity & potential mechanisms

Next, we explore heterogeneity in website vendor usage after the GDPR in Section 6.1. This illustrates the policy's incidence as well as potential consequences of its penalty design. Our analysis also indicates potential mechanisms (Section 6.2) that explain the post-GDPR evolution of vendor usage and the rebound in particular.

6.1 Usage heterogeneity by website characteristics

Here, we explore the relationship between website characteristics and variation in vendor choice. As noted in Section 3.1.1, we observe six static characteristics (Z_{ik} for $k \in \{1, ..., 6\}$) and two time-varying site characteristics (X_{itj} for $j \in \{1, 2\}$). We first explore these relationships using a sequence of semi-parametric regressions that characterize expected vendor use conditional upon each characteristic, $E\left[y_{it}|Z_{ik}, \theta_i\right]$ or $E\left[y_{it}|X_{itj}, \theta_i\right]$, and conditional on site fixed effects θ_i. We then use a parametric model to explore expected vendor use conditional upon all characteristics, $E\left[y_{it}|Z_i, X_{it}, \theta_i\right]$.

6.1.1 Individual website characteristics: Semi-parametric analysis

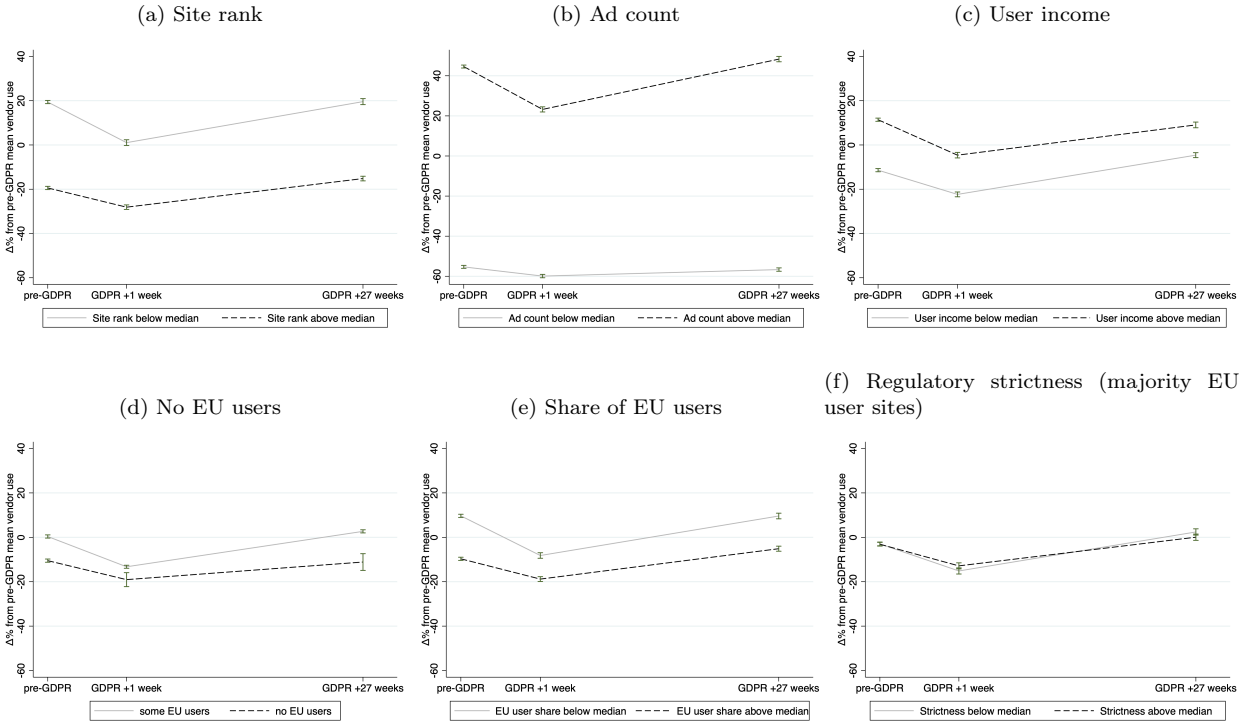
Vendor use conditional upon static site characteristics, $E[y_{it}|Z_{ik}, \theta_i]$. Our static, pre-GDPR site characteristics are measures of: (a) the site's (log) traffic rank (inversely related to the volume of site traffic), (b) the (log) number of ads on the page, (c) the average user income, (d, e) the site's EU user share and an indicator for no EU users at all, and (f) regulatory strictness. We conjecture that these characteristics may shift website vendor usage through their influence on the site's expected revenue. We include regulatory strictness for descriptive purposes as contemporary GDPR research suggests a correlation with the regulation's impact (Goldberg et al., 2021; Jia et al., 2020, 2021). We explore these relationships using a sequence of semi-parametric regressions similar in form to equation (1):

$$y_{it} = \mu + \theta_i + \sum_{t>0} GDPR_t \cdot \left(\lambda_{kt}^L \cdot \left(Z_{ik} < Z_k^{cut}\right) + \lambda_{kt}^H \cdot \left(Z_{ik} \geq Z_k^{cut}\right)\right) + \varepsilon_{it} \tag{2}$$

Equation (2) delivers estimates (λ_{kt}^L, λ_{kt}^H) of differential vendor use, conditional upon post-GDPR week t and low-vs-high values of site characteristic k, Z_{ik}. In equation (2), Z_k^{cut} captures the cutoff used for low/high characteristic classification: continuous regressors are median-split while binary regressors are split by level. Rather than plotting the full time evolution of differential vendor use, we compare predicted conditional means from equation (2) for the pre-GDPR period, 1 week post-GDPR (short run) and 27 weeks post-GDPR (long run). We report percentage changes in expected vendor use relative to the pre-GDPR global average. Figure 3 plots the results for each of the six static site characteristics.

Figure 3 reveals several key insights. First, the primary drivers of variation from the unconditional pre-GDPR mean can be assessed informally by comparing the range of percentage changes across characteristics. For example, variation in ad count is the largest driver of vendor usage variation relative to the pre-GDPR mean: below median ad count sites use 55-60 percentage points (p.p.) fewer vendors while above median ad count sites use 23-48 p.p. more vendors, depending on the time period. As in Table 1, the median site has 1 ad per page, so Figure 3b effectively splits sites by whether or not they run ads. Site rank has the second largest range of percentage changes and is next biggest driver of usage variation. Second, comparing changes in levels between the pre-GDPR and 1 week post-GDPR periods across characteristics and their levels reveals sources of the short-run drop in aggregate vendor usage noted in Section 4.2.1. The largest short-run drops are among high ad count sites, followed by sites with low rank (more popular), then low EU user share, high income, no EU users, and high regulatory strictness. Third, comparing changes in levels for 1 week post-GDPR and 27 weeks post-GDPR across characteristics and their levels reveals sources of the long-run rebound in vendor usage noted in Section 4.2.1; similarly, comparing change levels for pre-GDPR

Figure 3: Comparison of vendor usage (over time) for low/high values of static site characteristics



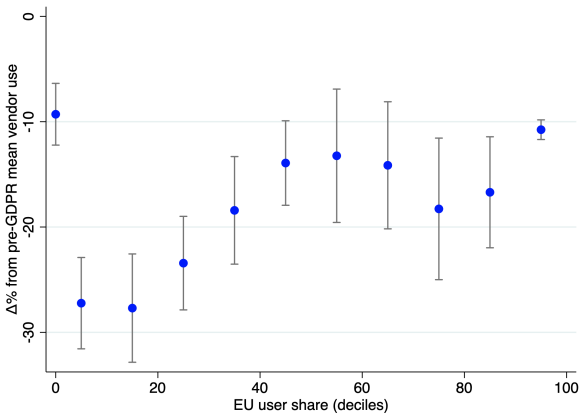
Notes: (i) Levels represent the Δ% from the model-predicted pre-GDPR unconditional mean vendor usage.
(ii) Continuous regressors split by below/above median; binary regressors split by 0/1 values.
(iii) Error bars indicate 95% confidence intervals using robust standard errors.

and 27 weeks post-GDPR captures deviations from a complete rebound for that sub-population. We note the largest short-to-long rebounding in usage among high ad count sites, followed by sites with low rank (more popular), then low EU user share, low income, low regulatory strictness, and some EU user traffic. Sites with high income, low ad count, and no EU users do not regain pre-GDPR usage levels, while all other groups rebound or exceed pre-GDPR levels.

The pattern of results largely conforms to our expectations. GDPR sensitivity appears positively associated with site advertising intensity and popularity—both shift the revenue that sites derive from vendors. Though the differences are modest, sites in stricter regulatory regimes add fewer vendors back in the long run, but also cut fewer vendors in the short run: we discuss a possible explanation in Section 6.2.3. Sites with higher income users exhibit this same pattern. The relationship to EU traffic share is more nuanced, in that low (but positive) EU share sites appear more responsive to the GDPR than either high EU share or no EU share sites.

Figure 4 illustrates this discontinuous relationship between EU traffic share and the short-run drop in vendor usage. The figure plots the results of an analogous regression to equation (2) that breaks the short-run change in vendor use into eleven EU user share bins: 0%, (0%,10%], (10%,20%], ..., (90%,100%]. We

Figure 4: Short-run change in vendor use by EU user share



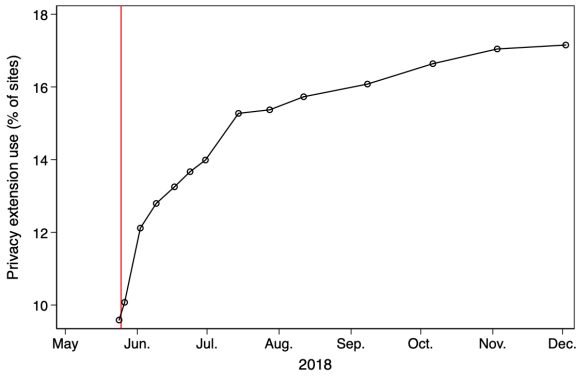
see that sites with a high share of EU users cut about half as many vendors as their counterparts with low, but positive share of EU users. In particular, we note a discontinuous relationship at 0%: sites without EU users cut fewer vendors than sites with low, but positive traffic from the EU. Note that these results are robust to including more covariates in the full model (Table 6). Moreover, Figure 9a of Appendix A.4 reveals that this pattern persists three months after the GDPR. We speculate that this pattern arises from the GDPR's the penalty design. All else equal, a low EU user share site derives lower value from EU users than a high EU user share site, but both face the same potential 4% fine on their global revenue. Thus, the low-EU user share site face relatively greater incentives to limit vendor use, despite processing less EU user data. Finally, sites that do not target EU users are exempted from the EU. Sites without EU users can therefore maintain their vendor usage at pre-GDPR levels, though some may opt to limit their vendor use as a precaution (particularly for sites in EU languages).

Vendor use conditional upon time-varying site characteristics, $E\left[y_{it}|X_{itj},\theta_i\right]$. We observe two key time-varying site characteristics: an indicator for website use of a privacy extension (PE_{it}) and a site-specific index for GDPR-related web search activity ($SEARCH_{it}$). Privacy extension use indicates site adoption of a consent management platform or related privacy-enhancing technology. Observing privacy extension use lends insight into the differential roles of the consent-related and the data minimization-related aspects of the GDPR, with the latter being our primary empirical focus. Our search variable combines EU country-level Google Trends data for GDPR-related web search with site-level country traffic weights. We examine $SEARCH_{it}$ for descriptive purposes and propose it conceptually as a site-specific proxy for the salience of the GDPR reflecting, for instance, news coverage of the regulation.

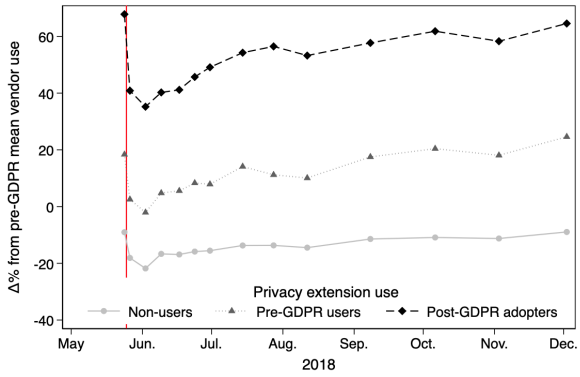
In the first panel of Figure 5, we plot the net adoption rate of privacy extensions in our sample ($E_i[PE_{it}]$)

Figure 5: Privacy extension (PE) use

(a) PE use over time



(b) Vendor use, split by PE adoption timing

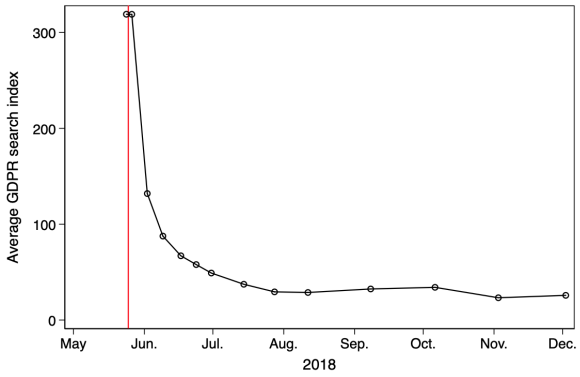


over time. We see that privacy extension adoption increases at a decreasing rate, from 9.5% pre GDPR to 17.1% 27 weeks post GDPR. We note that the industry's main compliance framework (IAB Europe, 2018a) was released right before the GDPR enforcement deadline, which contributed to low baseline privacy extension use. In the second panel of Figure 5, we use a regression analogous to equation (2) to plot the evolution of vendor use conditional upon privacy extension use. We categorize websites by privacy-extension use according to whether they are (1) non-users, (2) baseline pre-GDPR privacy-extension users, or (3) privacy-extension users at 27 weeks post-GDPR. We note that sites are selected such that baseline vendor use is highest among post-GDPR privacy-extension adopters, followed by pre-GPDR users, then non-users. Recall from Section 5.3.2 that privacy extension use had little influence on short-run concentration changes. Figure 5b shows that the evolution in vendor usage may be similar across our three privacy-extension adoption categories—at least on a relative basis. Two factors in combination—increasing privacy extension adoption and vendor usage rates increasing faster among adopters—suggest that privacy extension use contributes to the observed rebound in vendor usage post-GDPR. However, Section B.4 shows that this contribution is limited in comparison to the total observed rebound.

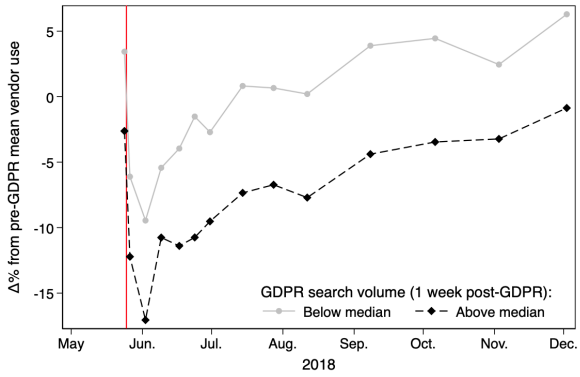
In the first panel of Figure 6, we plot the average value of $SEARCH_{it}$ over time. The plot shows that the salience of the GDPR, as proxied by $SEARCH_{it}$, peaks around the enforcement deadline, decreases rapidly over the summer of 2018, and falls at a far slower rate thereafter. The second panel of Figure 6 plots the evolution of vendor use, split by above/below median search one-week after the enforcement deadline. Higher search (salience) is correlated with lower vendor use. However, search alone is associated with a limited amount of variance about the pre-GDPR mean: usage differences between below/above median search sites remain between 5 and 9 percentage points.

Figure 6: GDPR search index (SEARCH)

(a) Average search index over time



(b) Vendor use, median split GDPR search volume



6.1.2 All site characteristics: Parametric analysis

Vendor use conditional upon all site characteristics, $E\left[y_{it}|Z_i, X_{it}, \theta_i\right]$. Finally, we use a parametric model to analyze vendor use conditional upon all observed site characteristics:

$$\begin{aligned} y_{it} \quad = \quad & \mu + \theta_i + \left(\lambda^1 + \lambda^2 \cdot t\right) \cdot GDPR_t + \sum_k \left(\nu_k^1 + \nu_k^2 \cdot t\right) \cdot Z_{ik} \cdot GDPR_t \\ & + \left(\beta^0 \cdot PE_{it} + \gamma^0 \cdot SEARCH_{it} + \psi^0 \cdot SEARCH_{it} \cdot STRICT_i\right) \cdot BASE_t \\ & + \left(\beta^1 \cdot PE_{it} + \gamma^1 \cdot SEARCH_{it} + \psi^1 \cdot SEARCH_{it} \cdot STRICT_i\right) \cdot GDPR_t + \epsilon_{it} \end{aligned} \tag{3}$$

where y_{it} indicates the number of vendors used by website i at time t. Equation (3) allows static site characteristics to influence post-GDPR usage levels and linear time trends. The post-GDPR level shifts are associated with the short-run GDPR impact, while the trends tie to the long-run post-GDPR evolution of vendor usage. Further, our model includes time-varying characteristics and allows them to have different impacts pre- versus post-GDPR. We include GDPR-related search and regulatory strictness for descriptive purposes. We further propose that GDPR salience—as reflected by our search index—may moderate vendor usage through its interaction with regulatory strictness. We use our full panel but omit the May 25^{th} scan, so that the short-run GDPR interactions better capture the pre- versus 1-week post-GDPR impact.[11]

Our estimates in Table 6 refine our findings from the analysis of vendor usage by individual characteristics. We begin with the time-varying characteristics, which are all highly statistically significant. We see that privacy extension use is associated with a 0.55 increase in vendor use post-GDPR. However, we obtain lower estimates when we instead employ a difference-in-differences approach[12] or remove privacy-related vendors

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

[12] Using an alternative difference-in-differences approach from Callaway and Sant'Anna (2021) as a robustness check, we

Table 6: GDPR impact heterogeneity by website characteristics

Time horizon	Post-GDPR (1)	Post-GDPR (2)
Interaction	$GDPR_t$ *x*	$GDPR_t$ *x Week x*
post-GDPR level, trend main effects (λ^1, λ^2)	-4.8230***	-0.0043
	(0.4284)	(0.0101)
Static site characteristic (Z_i) interactions (ν^1,ν^2):		
log(Site rank)	0.3217***	-0.0031***
	(0.0416)	(0.0011)
log(Ad count + 1)	-0.9799***	0.0558***
	(0.0762)	(0.0021)
User income†	-0.6512***	-0.0053***
	(0.0668)	(0.0017)
Share of EU users (%)	0.0048*	0.0003***
	(0.0026)	(0.0001)
No EU users	2.2372***	0.0011
	(0.3288)	(0.0082)
Regulatory strictness† *x* >50% EU users ($STRICT_i$)	-0.1787	0.0023
	(0.1493)	(0.0023)
Time-varying site characteristics & interactions (β,γ,ψ):		
Privacy extension use (PE_{it})	0.5501***	
	(0.1528)	
GDPR search volume ($SEARCH_{it}$)	-0.0084***	
	(0.0005)	
$SEARCH_{it}$ *x* $STRICT_i$	0.0038***	
	(0.0005)	
Constant (μ)	15.7983***	
	(0.1887)	
Site fixed effects	X	
Pre-GDPR interactions with time-varying variables	X	
Observations	329,158	
R-squared	0.836	

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

from the outcome variable.[13] As expected, GDPR search volume is negatively associated with vendor usage. The interaction of search and strictness is positive and significant in the post-GDPR period.[14] We speculate about a potential explanation in Section 6.2.3.

For the static characteristics, the estimates confirm that higher ad counts and lower site ranks are associated with larger short-run (level) decreases and larger long-run (trend) increases in vendor usage, even after conditioning on other site characteristics. The previous finding that higher income sites are associated with steeper short-run decreases and smaller long-run increases in vendor usage is similarly robust.[15] As

obtain a smaller but statistically consistent estimate of 0.16 (0.31 s.e.) vendors.

[13]More than half of the increase in vendor usage associated with privacy extension adoption appears to be a mechanical increase in privacy-related vendor use. Once we remove privacy-related vendors (using the expanded definition from Section 4.2.1), our revised equation (3) estimate falls from 0.55 to 0.25 (s.e. 0.15).

[14]Note that the search and strictness findings are robust to focusing on local EU sites defined either as: i) the 54.9% of our sites whose country-level domains corresponds to an EU country, or ii) the 58.9% of our site that get more than 80% of their traffic from a single EU country.

[15]Equation (3) wishes to disentangle the contributions of user income and other correlated site characteristics. Inasmuch as

with Figure 4, we again find that low (but positive) EU share sites are associated with steeper short-run decreases and smaller long-run increases in vendor use than both zero and high EU share sites.

6.2 Potential mechanisms

We conjecture that observed patterns of vendor use arise from a combination of (at least) two mechanisms: (1) evolving website compliance norms (e.g., consent management), and (2) exogenous vendor innovation and market dynamics. We summarize our evidence related to these potential mechanisms in the subsections that follow.

6.2.1 Evolving compliance norms

In Section 2, we outline how websites can comply with the GDPR along multiple dimensions. For instance, websites can limit vendor links and manage user consent in order to comply with the GDPR. At the extreme, websites can even block EU users. Below, we consider how the industry's evolving compliance norms may contribute to the vendor usage dynamics in Figure 1. We show that both privacy extension use and blocking EU users contribute little to the observed vendor usage dynamics.

Given regulatory complexity and the multi-dimensional nature of GDPR compliance, we speculate that industry norms help guide firm compliance activities. At the same time, regulators must be strategic in choosing which sites to target with enforcement, so that sites that violate both the law and industry norms become compelling targets. Given the GDPR's emphasis on consent and past regulatory guidance, the regulation pressures websites to obtain user consent for data processing. As industry adoption of privacy-related extensions grew, non-adopters may become greater enforcement targets, which reinforces further adoption. On the other hand, few websites waited to obtain consent before engaging vendor third-party domains (Sanchez-Rola et al., 2019) despite multiple regulator reports condemning the practice (e.g. AP, 2019; DPC, 2020). Given this compliance norm did not take root and vendor usage varies so much by site (see Table 1), regulators may find that vendor usage is less useful for selecting enforcement targets. Perhaps as a result, we see that average vendor use and privacy extension use both grew post-GDPR, despite the lack of enforcement in this industry. However, we note below that these (non-)compliance strategies were largely orthogonal to each other.

Privacy extension use We investigate the possibility that sites employed privacy extensions as a complement for increasing vendor usage. In other words, websites compliance norms shifted to obtaining user

user income shifts site income like site rank and ad count, the smaller long-run increase in usage among higher income sites is unexpected. Correlation between user income and privacy sensitivity may explain some of this pattern. The trend interaction on income becomes statistically insignificant when we add a survey metric for consumer privacy sensitivity to Equation (3).

(opt-out) consent, which allowed sites to add back vendors that they had initially dropped. In this case, we can directly measure privacy extension use and empirically evaluate its contribution to vendor usage.

Though several prior results indicate that privacy extension adoption aids the rebound in vendor use, we estimate a small contribution. Table 2 suggests rapid, relative growth in the privacy-related vendor category. However, we instead focus on privacy-extension usage as a wider measure of website consent-related compliance activities. Figure 5 shows: i) privacy extension use increases by 7.6% of sites over the course of our sample and 4.9% between weeks 1 and 27 post-GDPR, and ii) sites that self-select into privacy extension use have greater vendor use. Table 6 estimates a 0.55 increase in vendor usage associated with privacy extension adoption post-GDPR, which we noted above may overestimate the impact of privacy extension adoption on (non-privacy-related) vendor use. We combine this 0.55 incremental vendor estimate with the 4.9% net adoption of privacy extensions between weeks 1 and 27 post-GDPR to infer that privacy extension use accounts for an average per site increase of 0.03 vendors. Thus, the apparent complementarity between privacy extension adoption and vendor use represents only 1.4% of the rebound in vendor use from weeks 1 to 27 post-GDPR: a 1.95 vendor rebound as predicted by the equation (3) model estimates.

We acknowledge that our privacy extension variable is measured with delay and may miss some custom website implementations of privacy notices or consent. However, the detection delay issue is mitigated by time delays between our website scans. We also expect that most sites employ a third-party privacy extension due to the engineering costs of managing consent. Finally, we note that BuiltWith detects the presence of the industry-standard approach (IAB Europe, 2018a) for sharing user consent choices even if the site designs a custom consent menu.

Blocking EU users 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 usage 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). We conclude that site blocking of EU users does not materially contribute to or interfere with observed site-vendor relationships.

6.2.2 Vendor market dynamics

Second, we consider supply-side explanations for the observed vendor usage patterns, including vendor entry/exit, and product innovation. Interpreted through this mechanism, the short-run drop in usage could

be explained by vendor choices to exit in the short run. The long-run rebound in vendor use could be explained by either vendor entry, vendor re-entry, or service innovation among incumbent vendors.

We find that entry and exit play a limited role in explaining the post-GDPR evolution of vendor usage. In Appendix A.1, we show that the largest vendor to completely exit the majority EU-user sites only appeared on 101 sites (0.54% of such sites). We next analyze the role of vendor entry and exit between the beginning and end of our sample. We find essentially no net entry: entrants contribute an increase of 2.03% relative to our initial website-vendor ties, whereas exiting firms contribute a reduction of 1.96%. Finally, we consider the role of vendor re-entry: vendors that exit in the short run, but return in the long run. We find that re-entry explains only 1.99% of the post-GDPR rebound in vendor use.

We show that vendor service innovation contributes to the post-GDPR growth in vendor usage. Vendor innovation could materialize in the form of new service features, declining data processing costs, or reduced service substitutability. To explore the potential contribution of vendor innovation, we examine the set of site-vendor ties that break one week after the GDPR in Appendix A.1. 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. This could result from either vendor service innovation or evolving website needs. In our data, the privacy compliance and hosting categories provide the clearest evidence of vendor innovation: Table 2 shows that usage grew 123.6% and 5.0% respectively over the pre-GDPR baseline by the end of 2018.

Since we do not observe vendor service features or prices and we lack a clean control group, we cannot further characterize the extent of vendor innovation without additional assumptions. For instance, if we project forward the pre-GDPR growth in vendor use—6% annual growth according to Peukert et al. (2022)—this continued exogenous growth would account for about a fifth of the post-GDPR rebound.[16]

6.2.3 Mechanism discussion summary

To varying degrees, we find evidence consistent with both mechanisms influencing vendor usage patterns: i) privacy extension adoption, and ii) vendor service innovation. We demonstrate that privacy extension adoption explains as little as 1.4% of the post-GDPR rebound, due to limited uptake over that period (4.9% of sites). Though our direct evidence for the vendor service mechanism is limited, our simple projection above suggests this explanation contributes as much as a fifth of the post-GDPR rebound. In assessing these findings, we again emphasize the descriptive nature of our results and the increasing severity of bias concerns for long-run GDPR impact estimates. We also acknowledge that other explanations are possible.

[16] Specifically, we compare 26 post-GDPR weeks of growth at a 6% annual rate on a baseline of 14.54 pre-GDPR vendors to the estimated 1.95 vendor rebound.

For instance, we speculate that the absence of regulatory enforcement in this industry in 2018 may have weakened website enforcement expectations. If websites felt less pressure to limit their vendor use, this may have contributed to the post-GDPR rebound in average vendor use. This enforcement explanation is consistent with the European Commission (2019) review of GDPR, which cited lack of enforcement as an obstacle to the regulation's full realization.

Though we do not observe enforcement expectations, we propose our regulatory strictness and search volume measures as rough proxies. Table 6 shows a highly statistically significant relationship between website's vendor use and both proxies. In particular, vendor use is negatively correlated with search volume, but the further interaction with regulatory strictness is negative. One explanation for this pattern is that sites behave as though they believe that enforcement is more likely when the policy's social salience is higher. For instance, regulators may be more likely to receive a citizen GDPR complaint when the policy's salience is higher. In addition, the negative interaction with search and strictness is consistent with sites believing that laxer regulators are more responsive to changes in policy salience. That is, low strictness regulators may be perceived as more reactionary to policy salience while high strictness regulators may be perceived as more consistent with their enforcement operations. Nevertheless, we stress that these variables are flawed proxies for enforcement expectations. Search volume data better reflects news coverage and social interest in the GDPR rather than enforcement expectations. Further, our regulatory strictness measure is cross-sectional rather than dynamic and arises from a 2008 survey. Nevertheless, other GDPR studies also find that our regulatory strictness measure moderates the impact of the GDPR on technology venture capital (Jia et al., 2021) and recorded site outcomes (Goldberg et al., 2021).

Though our short-run estimates are better suited to causal interpretation, this raises concerns about drawing generalizable insights from a short time period. For instance, we suggest that the GDPR would have had a greater impact if if the law had been enforced in this industry. Furthermore, the enforcement deadline was characterized by uncertainty—over firm compliance norms and regulator enforcement priorities—that may limit generalizability. Nevertheless, we note that the directional impact on both vendor usage and concentration persisted for months after the enforcement deadline. These also persisted at the end of 2018 in the advertising vendor category of most interest to regulators.

To explore the generalizability of our findings, we gathered and analyzed data on two supplemental event studies related to GDPR enforcement. In Web Appendix C, we analyze the vendor usage and market concentration responses to: i) the roll-out of IAB's revised Transparency and Consent Framework (TCF v2) in the Fall of 2020, and ii) the enforcement deadline by France's regulator CNIL in the Spring of 2021. These highly publicized events presumably shifted beliefs regarding compliance norms and the probability of non-compliance penalties. We show both events are associated with concurrent decreases in vendor use

and increases in market concentration, which is consistent with our main findings using the 2018 data. Several pieces of evidence therefore suggest that, when regulation pushes websites to limit their vendors, this disproportionately harms smaller vendors.

7 Conclusion

This paper provides novel empirical evidence suggesting a 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 line with the GDPR's data minimization mandate. Although these gains appear to erode over time, EU regulators have continued to criticize this industry's practices as non-compliant (ICO, 2019; AP, 2019) and issued substantial fines to websites at the end of 2020.

Our analysis reveals that the GDPR may have initially succeeded in its data minimization goal in the data-intensive web technology industry. We also find descriptive evidence of three unintended consequences of the GDPR. First, we see a short-run increase of 17% in aggregate concentration, which also attenuates over time. We show that the possible tradeoff between data minimization and concentration is not mechanical: some niche categories become less concentrated. However, relative concentration increases in the top four web technology categories that together represent 94% of website-vendor ties: advertising, hosting, audience measurement, and social media. Second, user personal data processing may also become more concentred as the increase in concentration is highest among the web technology vendors that process personal data. As policymakers wrestle with how to protect individual privacy, they may therefore seek to balance the risk of increasing the concentration of personal data ownership and increasing market power (Gal and Aviv, 2020; Geradin et al., 2020). Third, the global nature of GDPR fines may have most protected the privacy of EU users on foreign sites. We find that sites with the largest share of EU users reduce their vendor usage less than sites with a small share of EU users. More research is needed to determine how policymakers could mitigate these potential unintended policy consequences.

This 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 change in concentration. For instance, if ad vendors with greater reach are associated with greater ad revenue share, then our findings could understate the 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 a nuanced picture. 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 offering better compliance with the regulation, which in turn may deflect enforcement attention from websites to major technology vendors.

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.

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*.

Article 29 Data Protection Working Party (2012). Opinion 04/2012 on cookie consent exemption.

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

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.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

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

Chen, C., C. B. Frey, and G. Presidente (2022). Privacy regulation and firm performance: Estimating the GDPR effect globally. The Oxford Martin Working Paper Series on Technological and Economic Change.

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). Online targeted advertisement: what action plan for the CNIL?

Commission Nationale de l'Informatique et des Libertés (2021a). Cookies : sanction de 50 000 euros à l'encontre de la société du Figaro.

Commission Nationale de l'Informatique et des Libertés (2021b). Nouvelles règles pour les cookies et autres traceurs : bilan de l'accompagnement de la CNIL et actions à venir.

Competition and Markets Authority & Information Commisioner's Office (2021). Competition and data protection in digital markets: a joint statement between the CMA and the ICO.

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

Data Protection Commission (2020). Report by the Data Protection Commission on the use of cookies and other tracking technologies.

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

European Commission (2008). Flash eurobarometer 226: Data protection in the European Union : Data controllers' perceptions.

European Commission (2019). 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 Data Protection Board (2018). Guidelines 3/2018 on the territorial scope of the GDPR (Article 3).

European Data Protection Board (2020). Contribution of the EDPB to the evaluation of the GDPR under Article 97. Technical report.

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

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 (2022). Consumer consent and firm targeting after GDPR: The case of a large telecom provider. *Management Science 68*(5), 3330–3378.

Goldberg, S., G. Johnson, and S. Shriver (2021). Regulating privacy online: An economic evaluation of the GDPR. Available at SSRN 3421731.

Gowrisankaran, G., A. Langer, and W. Zhang (2022). Policy uncertainty in the market for coal electricity: The case of air toxics standards. Working paper, Columbia University, University of Arizona, and Renmin University.

Hils, M., D. W. Woods, and R. Böhme (2020). Measuring the emergence of consent management on the web. In *Proceedings of the ACM Internet Measurement Conference*, IMC '20, New York, NY, USA, pp. 317–332. Association for Computing Machinery.

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

IAB Europe (2018b). Header bidding and auction dynamics. White paper.

IAB Europe (2019). GDPR transparency and consent framework (v2).

IAB Europe (2020). The IAB Europe Transparency & Consent Framework (TCF) Steering Group votes to extend technical support for TCF v1.1. https://iabeurope.eu/all-news/the-iab-europe-transparency-consent-framework-tcf-steering-group-votes-to-extend-technical-support-for-tcf-v1-1/.

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

Janssen, R., R. Kesler, M. E. Kummer, and J. Waldfogel (2022). GDPR and the lost generation of innovative apps. NBER working paper.

Jia, J., G. Z. Jin, and L. Wagman (2020). Gdpr and the localness of venture investment. *Available at SSRN 3436535*.

Jia, J., G. Z. Jin, and L. Wagman (2021). The short-run effects of the General Data Protection Regulation on technology venture investment. *Marketing Science 40*(4), 661–684.

Johnson, G. A., S. K. Shriver, and S. Du (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science 39*(1), 33–51.

Johnson, M. S. (2020). 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. S. Silveira (2021). Understanding disparities in punishment: Regulator preferences and expertise. *Journal of Political Economy 129*(10), 2947–2992.

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

Ke, T. T. and K. Sudhir (2022). Privacy rights and data security: GDPR and personal data driven markets. *Available at SSRN 3643979*.

Koski, H. and N. Valmari (2020). Short-term impacts of the GDPR on firm performance. ETLA Working Papers.

Kostov, N. and S. Schechner (2019). GDPR has been a boon for Google and Facebook.

Laboratoire d'Innovation Numérique de la CNIL (2021). Observer les pratiques cachées du web. https://linc.cnil.fr/obs-cookies/.

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.

Lukic, K., K. M. Miller, and B. Skiera (2021). The impact of the general data protection regulation (gdpr) on the amount of online tracking. Working paper.

Matte, C., C. Santos, and N. Bielova (2020). Purposes in IAB Europe's TCF: which legal basis and how are they used by advertisers? In *APF 2020 - Annual Privacy Forum*, Lisbon, Portugal, pp. 1–24.

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 (2022). Regulatory spillovers and data governance: Evidence from the GDPR. *Marketing Science 41*(4), 318–340.

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.

Prasad, A. and D. R. Perez (2020). The effects of GDPR on the digital economy: Evidence from the literature. *Informatization Policy 27*(3), 3–18.

Ravichandran, D. and N. Korula (2019, August). Effect of disabling third-party cookies on publisher revenue. Technical report, Google Inc.

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.

Schmitt, J., K. M. Miller, and B. Skiera (2021). The impact of privacy laws on online user behavior. *arXiv preprint arXiv:2101.11366*.

Sharma, P., Y. Sun, and L. Wagman (2021). The differential effects of privacy protections and digital ad taxes on publisher and advertiser profitability. SSRN working paper.

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.

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). 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). Horizontal merger guidelines.

Utz, C., M. Degeling, S. Fahl, F. Schaub, and T. Holz (2019). (un) informed consent: Studying GDPR consent notices in the field. In *Proceedings of the 2019 acm sigsac conference on computer and communications security*, pp. 973–990.

WARC (2019). Internet in decline beyond Google and Facebook.

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

Zhao, Y., P. Yildirim, and P. K. Chintagunta (2021). Privacy regulations and online search friction: Evidence from GDPR. *Available at SSRN 3903599*.

Zhuo, R., B. Huffaker, kc claffy, and S. Greenstein (2021). The impact of the General Data Protection Regulation on internet interconnection. *Telecommunications Policy 45*(2), 102083.

Appendices

A Vendor usage robustness

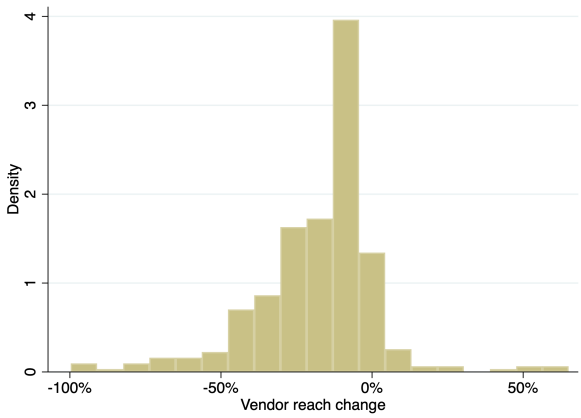
A.1 Vendor exit and entry

Our interpretation of the observed pattern of web technology vendor usage post-GDPR emphasizes website choice of vendors. However, the short-run drop in usage followed by resumed growth over the long run could be explained by vendor choices to exit in the short run and enter or re-enter the market in the long run. 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 explore these issues, we first establish that no vendor with high reach drops out of our data one week after the GDPR (short run). 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 7 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.

We also examine the entry, exit, and re-adoption of vendors at the end of 2018. To examine entry and exit, we focus on the complete panel of websites that we scan pre-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 further restrict the sample to sites that we also scan one week post-GPDR and then consider the 91,004 total vendor ties that these sites cut one week after the GDPR. We see that sites reestablish

Figure 7: Distribution of short run change in vendor reach



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

29,491 (32.4%) of these vendor ties by the end of 2018, so that 61,512 ties (67.6%) remain severed. Lastly, we consider the role of re-entering vendors that exit one week after the GDPR, but return by the end of 2018. Vendor re-entry accounts for 1,099 site-vendor ties, which represents 1.99% of the net increase of 55,200 ties post-GDPR. In sum, vendor re-adoption and re-entry play a minor role in the post-GDPR growth in vendor use whereas net entry plays almost no role.

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

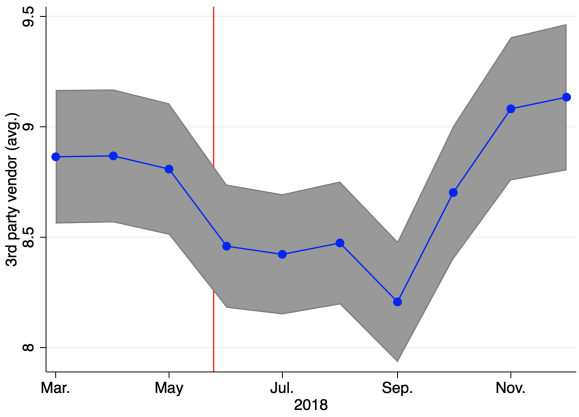
We use external data from WhoTracks.me (2018) to examine the trend in web technology use prior to GDPR enforcement. WhoTracks.me (2018) released public data on the monthly web technology use of the 1,500 top EU websites beginning in March 2018. Karaj et al. (2018) describe the data collection methodology, 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.[17] Karaj et al. (2018) also map third-party domains into vendors (referred to as "trackers") using their own database (see also Appendix B.2).

Figure 8 shows the evolution of average third party vendor usage for the 1,322 WhoTracksMe sites that

[17]Results using the full sample of 1,452 sites are essentially the same. See footnote 18.

Figure 8: 3rd party vendors pre-post GDPR: Sites in WhoTracks.me (2018) & our data



are also in our data. Figure 8 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.[18] Thus, the WhoTracks.me (2018) data suggests a flat pre-trend in third party usage leading up to enforcement. Post-GDPR, Figure 8 shows a similar trend to Figure 1 as the number of vendors increases to pre-GDPR levels by the end of 2018.

As we discuss in Section 4.1.1, we are concerned about the potential for anticipatory compliance, which would create a decreasing pre-trend in vendor usage. Our above analysis suggests a flat pre-trend. Other GDPR studies of this industry present evidence against material anticipatory compliance in that they do not show a decreasing trend in vendor use leading up to the enforcement deadline. 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. (2022) find a slightly increasing pre-GDPR trend. We conclude that sites did not significantly reduce vendors before the GDPR deadline, which would otherwise lead us to underestimate the impact of the GDPR on technology vendor usage (in both the short and long run).

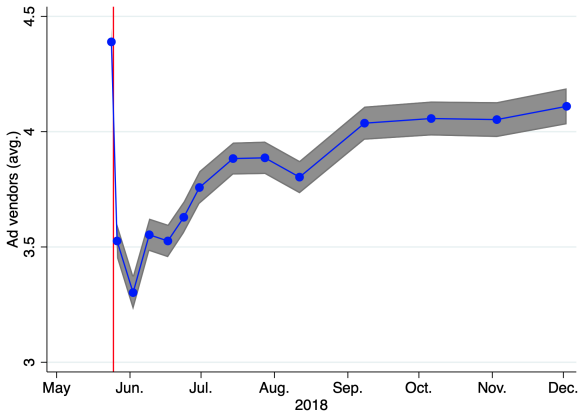
A.3 Alternative outcome measures

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

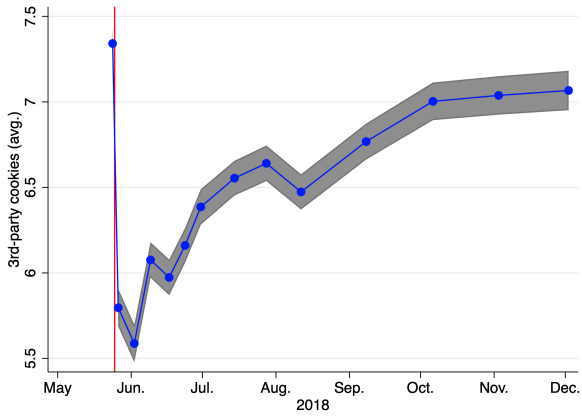
[18] For the full sample of 1,452 sites, the number of third party vendors rises by only 0.08% in April and falls -0.63% in May.

Figure 9: Evolution of average vendor usage per website: robustness to alternate dependent variables

(a) Ad vendors



(b) Third-party cookies



consider vendors in the advertising category because EU regulators have singled out this category for its use of personal data (ICO, 2019; CNIL, 2019). Further, ad category vendors directly generate revenue for websites, and our usage analysis (Table 2) and site heterogeneity analysis (Figure 3b and Table 6) suggest that website choice of these vendors are 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 9 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 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, and 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 9a contradicts this prediction both because it shows an overnight drop in vendor use on the GDPR enforcement deadline and because ad technology 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."[19]

Figure 9b 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 9b's pattern also contradicts a learning-to-comply story whereby websites dropped vendors until they learned to

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

comply with the GDPR: third-party cookie use increased post-GDPR even for users who did not provide consent.

A.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.

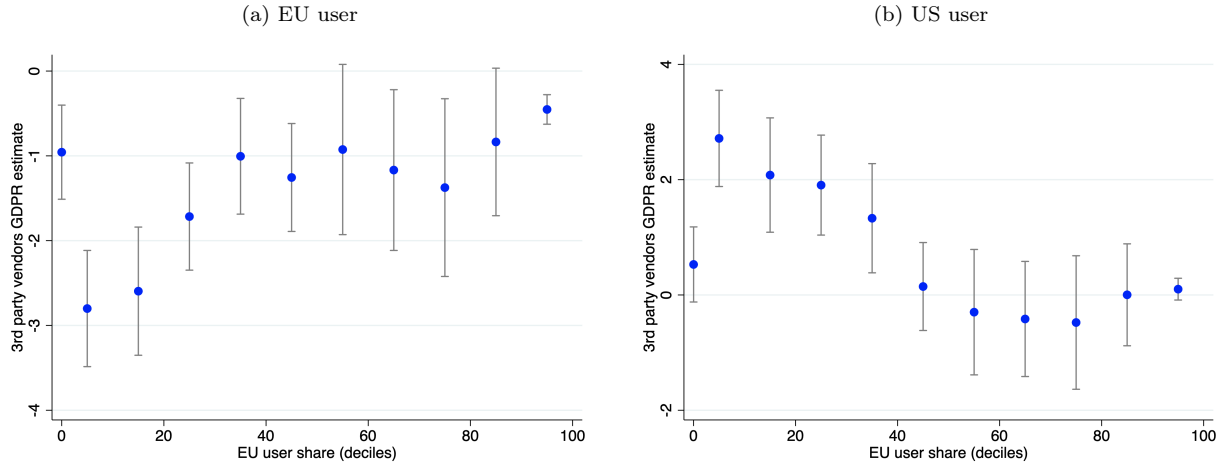
We find that websites in our sample employ fewer vendors on average for EU users than they do for US users after GDPR. To show this, we use regressions to compare each of the EU user and US user data collections to a common pre-GDPR baseline (which was collected using the VPN in France). As in equation (1), 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 we expect that the GDPR induces sites to treat users in the EU differently.

Figure 10 breaks down the GDPR effect estimates for an EU user (Figure 10a) and a US user (Figure 10b) 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. Analogous to equation (2), we then regress vendor count on a GDPR indicator interacted with these eleven indicators and include site fixed effects. We plot the resulting estimates in Figure 10.

Figure 10a highlights the deterrent effect of the GDPR penalties on non-EU sites discussed in Section 6.1. We see that the point estimates are increasing in the share of EU users. That is, sites with low, but positive EU user share 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 shows that the pattern observed in the short run (Figure 4) persists three months after the GDPR.

Figure 10b 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 10

Figure 10: Differential treatments: 3 month GDPR effect on vendor use by user country



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 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 10 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 three months post-GDPR 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.

Web Appendices

A Supplemental data description

A.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.

B Vendor concentration robustness

B.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 8, 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.

B.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 9 versus 14.4 in Table 3. This difference arises from defining uncategorized third-party domains as separate vendors and because WhoTracksMe categorizes more of the vendors in the data.

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.

Table 8: 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.

Our concentration results are broadly robust to the WhoTracksMe classification in Table 9. 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 3, 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%.

B.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.[20] 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

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

Table 9: Short run changes in 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.

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

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

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 10 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 3. Moreover, the signs of the category-level differences are the same as in Table 10 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.

B.4 Concentration extensions: long-run differences

To complement our concentration extensions in Section 5.3, Table 11 considers the changes in HHI at the end of 2018 for the same cuts of the data. As we have seen in Table 5, HHI returns to approximately its pre-GDPR level by the end of 2018. We largely see this pattern in Table 11: the long-run change in HHI is essentially zero for likely personal data, sites with and without privacy extensions, and the vendor market with or without Google and Facebook. However, we do see an increase in concentration for vendors that are unlikely to use personal data. This is perhaps surprising at face value, though again, we do not wish to

Table 10: Short-run changes in 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 analysis 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.

Table 11: GDPR & aggregate long-run concentration (27 weeks post): Three extensions

	HHI			
Data samples	Pre	Post	Diff.	Diff. (%)
A) Role of personal data				
Likely personal data	184.6	184.2	-0.4	-0.2%
Unlikely personal data	493.0	599.3	106.3	21.6%
B) Role of consent dialogs				
Sites with privacy extension	123.6	122.6	-1.0	-0.8%
Sites without privacy extension	152.0	153.7	1.7	1.1%
C) Role of top 2 companies (Google & Facebook)				
All vendors	145.3	146.2	0.9	0.6%
All but top 2 companies	46.0	46.3	0.3	0.7%

over-interpret differences that a) are conflated by exogenous industry trends, and b) reflect a small share of our data.

C Supplemental event studies

We consider two supplemental, GDPR-related event studies that provide additional opportunities to assess the GDPR's impact on vendor use and market concentration. Below, we describe these two event studies and then present the corresponding empirical evidence for both our key outcomes.

C.1 IAB TCF version 2 rollout

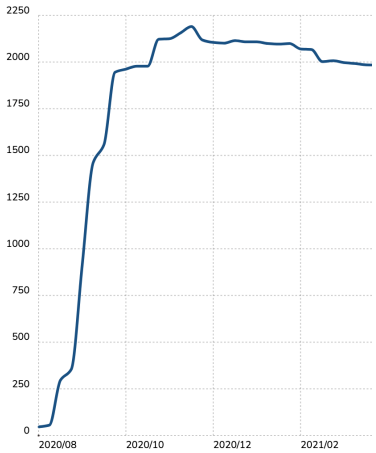
We first consider the rollout of a GDPR-related self-regulatory technology developed by Interactive Advertising Bureau (IAB) Europe, a leading industry association. IAB Europe developed the industry's main technical specification for transmitting user consent signals from websites to web technology vendors: the Transparency and Consent Framework (TCF). IAB Europe released its initial launch version of TCF (v1.1) right before the GDPR enforcement deadline. The initial version allowed users to provide consent by vendor and data-processing purpose (e.g., "ad selection, delivery, reporting" and "content selection, delivery, reporting"), but not to different data-processing purposes by vendor.

IAB Europe's Transparency and Consent Framework version 2 (TCF v2) remedied this short-coming and represented a significant upgrade on its predecessor (IAB Europe, 2019). TCF v2 also revised the set of codified data-processing purposes (Matte et al., 2020), and therefore is not backwards-compatible with its predecessor. Websites that adopted TCF v2 then needed to review their vendor lists and participating vendors' requested data-processing permissions. IAB Europe set an August 15, 2020 deadline for deprecating the TCF v2's predecessor—extending the original June 30, 2020 deadline (IAB Europe, 2020). Unlike TCF v1.1, publishers and vendors received a year of advanced notice, as the final version of the TCFv2 specification was released in August 2019. TCF v2 is also notable in that Google—the largest web technology vendor—agreed to eventually adopt the TCF standard for the first time. Figure 11 shows the adoption of TCF v 2 by the top 100,000 sites on the web globally, according to BuiltWith. Figure 11 shows that adoption grew rapidly after August 15, 2020 from near zero adoption and leveled off by November 2020. As a site adopts TCF v 2, the site performs a vendor-level review that may lead them to drop certain vendors. On the other hand, some sites may use TCF v2 as an opportunity to add vendors, as we saw modest evidence in Section 6.1.2 that sites that adopted a privacy extension use post-GDPR in 2018 added vendors. We examine the net effect empirically below.

C.2 CNIL enforcement deadline

The French data protection agency CNIL has made web technology and targeted advertising an area of focus since 2019. CNIL (2019) announced a plan to tackle the sector in June 2019 that began with releasing initial sectoral guidelines in July 2019 and included a yearlong transitional period for stakeholders. On October 1st, 2020, CNIL laid out its final sector-specific rules requiring explicit user consent before sites load cookies or similar. CNIL set an April 1, 2021 deadline for enforcing these rules (CNIL 2021b). On July 26, 2021, CNIL fined a french website (lefigaro.fr) 50,000 euros for using cookies without prior user consent (CNIL 2021a). In December 2020, CNIL also fined American tech giants Google and Amazon 100 million euros

Figure 11: Adoption of IAB Europe's TCF v2 among top 100,000 global websites



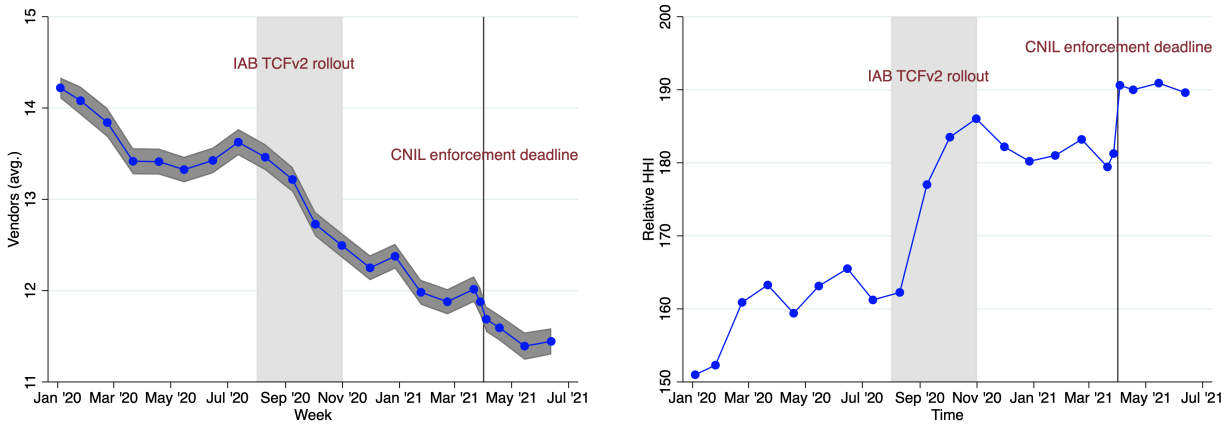
Source: https://trends.builtwith.com/ads/Consent-Management-Platform-API-v-2.0

and 35 million euros respectively for similar cookie violations on their websites. In all, CNIL (2021a) issued around 70 formal notices or sanctions regarding cookie violations between 2020 and to July 2021 (most sites were not named). Given CNIL's leading role in enforcing the GDPR for websites, we examine the impact of the CNIL enforcement deadline below. Our data are well suited for studying CNIL's deadline in particular, because we employ a VPN to appear as a user originating from France.

C.3 Change in vendor use and market concentration

To evaluate these supplemental GDPR event studies empirically, we continued collecting data every four weeks using the same procedure described in Section 3. We also collected weekly data within a four week window of CNIL's enforcement deadline. Figure 12 summarizes the evolution of both key outcome variables from January 2020 to June 2021.

Figure 12a graphs our estimates for the evolution of site vendor use. Like Figure 1, Figure 12a uses an analogous estimator to equation (1), where the initial January 3, 2020 scan serves as the baseline. Unlike Figure 1, Figure 12a demonstrates a general decline in observed vendor use, which represents a departure from the industry's past steady growth (Lerner et al., 2016). In particular, Figure 12a shows a reduction in vendor use until April 2020 followed by a plateau for about five months. IAB Europe's TCF v2 rollout coincides with a steep reduction in vendor use. To evaluate this, we compare our scans on August 10 and October 31, 2020. Using a fixed effects estimator (equation (1)), we see that vendor use falls -0.97 (s.e. 0.06) from a baseline of 13.68—representing a 7.11% decline. This reduction is half of our short-run GDPR estimate of -2.09 in average vendor use, but instead arises over a three month period. Figure 12a then shows continued decline until another plateau from January to March 2021. At the CNIL enforcement deadline, we



(a) Vendor use

(b) Aggregate market concentration

Figure 12: Impact of IAB TCF v2 rollout and CNIL enforcement deadline

observe a small but steep decline in vendor use. Within a four-week window of the CNIL's deadline (March 21 and April 17, 2021 scans), vendor use falls by -0.44 (s.e. 0.04) from a baseline of 12.26—a 3.55% decline. The majority of this drop occurs within a one-week window of CNIL's deadline (March 28 to April 3 scans): vendor use falls by -0.26 (s.e. 0.04).

Figure 12b graphs our estimates for the evolution of aggregate market concentration in 2020 and 2021. We construct this figure in the same way as its 2018 counterpart, Figure 2. As in 2018, the evolution of aggregate concentration is the mirror image of vendor use. Over this time period, the HHI increases by 38.6 points. Most of this increase coincides with the rollout of TCF v2 where HHI increases 23.8 points from 162.3. This increase is on par with the 25-point increase observed one week post-GDPR, though again over a longer time period. HHI exhibits a distinct discontinuity around CNIL's enforcement deadline, but is relatively flat for a least a couple months before and after that deadline. Within a four-week window of the CNIL's deadline, HHI rises 10.6 points from a baseline of 179.4—a 5.89% increase. Almost all of this drop occurs within a one-week window of CNIL's deadline: an increase of 9.3 points.

In sum, both natural experiments exhibit concurrent decreases in vendor use and increases in market concentration. Our short-run GDPR results exhibit the same pattern, which Peukert et al. (2022) also found in this industry using different data. Event studies can provide evidence of a policy's impact, but face inherent limitations as well. For example, concurrent changes may confound policy impact estimates. Particularly with the GDPR, a policy's impact may change as regulator enforcement and firm compliance decisions evolve. Here, we leverage two subsequent events that increase the salience of the GDPR for websites. Websites that adopt the IAB Europe's TCF v2 must re-evaluate each of their vendors, for instance, in light of the permissions they request for users. Figure 11 shows that most sites that adopted TCF v2 did so within

about six weeks. CNIL's enforcement deadline provides another discrete event study, where an EU privacy regulator threatens action after providing both months of notice and specific compliance prescriptions. In both cases, we replicate our short-run GDPR result that vendor use falls while market concentration rises. Together with our main GDPR results, this provides repeated evidence that the GDPR appears to reduce web technology vendor use, but that this relatively favors large vendors.