Does Privacy Regulation Harm Content Providers? A Longitudinal Analysis of the Impact of the GDPR

V. Lefrere,[*] L. Warberg,[†] C. Cheyre,[‡] V. Marotta,[§] and A. Acquisti[¶]

October 14, 2022

Abstract

While the European General Data Protection Regulation (GDPR) has received significant attention in the information systems literature, concerns that it would adversely affect websites' ability to provide quality content to their visitors have not been thoroughly investigated. We construct a longitudinal data-set of news and media websites to study how online content providers adapted their responses to the GDPR over time, and whether restrictions on online tracking enforced by the regulation affected downstream outcomes such as the quantity of content those websites offer to their visitors and visitors' engagement with such content. We find robust evidence of websites' reactions to the GDPR in both the US and the EU, including an initial reduction in the number of third-party cookies and intensity of visitor tracking. However, reactions differ between US and EU websites, and several months after the enactment of the regulation the initial reduction in tracking is reversed, as tracking among EU websites bounces back. We use difference-in-differences, LATE, and look-ahead matching models to assess downstream effects of the regulation, capturing both ecosystem effects and website-level effects. We find a small reduction in average page views per visitor on EU websites relative to US websites near the end of the period of observation, but no statistically significant impact of the regulation on EU websites' provision of new content, social media engagement with new content, and ranking in both the short-term and the long-term. We also find no evidence of differences in survival rates across EU and US content providers, and no evidence that monetization strategies change at a higher rate for EU websites relative to US websites. While industry predictions forebode dire consequences arising from the GDPR for content providers, we find that websites that responded more strongly to the GDPR were those less likely to be affected by such a response; in contrast, websites that relied in great part on EU visitors found, over time, ways to avoid being negatively affected by the regulation.

[*] Institut Mines Telecom, Business School. Email: vincent.lefrere@imt-bs.eu

[†] Engineering and Public Policy, Carnegie Mellon University. Email: warberg@cmu.edu

[‡] Cornell Bowers CIS, Cornell University. Email: ccheyre@infosci.cornell.edu

[§] University of Minnesota Twin Cities Carlson School of Management. Email:vmarotta@umn.edu

[¶] Heinz College, Carnegie Mellon University. Email:acquisti@cmu.edu.

1 Introduction

Because of growing consumer concerns over data privacy and governments' expanding policy efforts to address them, information systems, marketing, and economics scholars have become interested in estimating the impact and potential costs of privacy regulation (Goldfarb and Tucker, 2012). In May 2018, the European Union (EU) implemented the General Data Protection Regulation (GDPR)—a major new component of EU privacy law. With stiff fines for non-compliance, the GDPR could significantly impact the collection and use of personal data both within and outside EU markets. Since its implementation, much empirical attention has been devoted to capturing the GDPR's direct impacts - such as degrees of firms' compliance, changes in online consent mechanisms, and variations in flows of personal data across the advertising ecosystem (Degeling *et al.*, 2019; Johnson and Shriver, 2019; Aridor *et al.*, 2020; Peukert *et al.*, 2020; Jia *et al.*, 2021; Goldberg *et al.*, 2021; Godinho de Matos and Adjerid, 2022; Lukic *et al.*, 2022). Relatively less attention has been devoted to understanding downstream consequences the regulation might have on economically relevant metrics, such as the ability of online publishers to produce new, quality content. Content providers such as news and media websites frequently rely on online advertising for revenue. Hence, fears were raised that they could be uniquely and adversely affected by regulation of the flows of data used in programmatic ads and restrictions on online tracking (IHS Technology, 2015). We construct a longitudinal data-set of nearly one thousand EU and US content providers (news and media websites), mining information from their websites at regular intervals before and after the implementation of the regulation. We study how those content providers adapted their response to the GDPR, and whether the regulation affected the quantity and quality of their content, as well as their survival, and thus the ability of those websites to continue to engage audiences. In contrast to prior work, we capture website-level data that allows us to tie websites' responses to downstream outcomes, accounting for ecosystem effects that impact websites' outcomes because of their location (EU vs US), and website-level effects that are function of website-specific responses to the GDPR. Our longitudinal panel also allows us to account for how content providers reacted to the regulation in both the short- and the long-term.

Given the widespread collection of (and reliance upon) personal data across different sectors of the economy, the GDPR was predicted in early industry reports to produce substantial negative economic effects.[1] The online advertising industry was expected to be especially affected by the GDPR, since its growth is driven by the ability to track users' online behavior to deliver personalized advertising. Limitations on data collection (such as those imposed by the GDPR—see Section 3) may decrease the effectiveness of online advertising campaigns (Goldfarb and Tucker, 2011), depress ad spending,[2] and result in market concentration – favoring dominant players (Johnson and Shriver, 2019). In turn, reductions in advertising effectiveness, spending, and competition could ultimately harm online content providers, as advertising is a major revenue component for producers of digital goods (Lambrecht *et al.*, 2014). Industry predictions on the impact of privacy regulation on the ad-supported publishing ecosystem were particular dire: in 2015, the CEO of the Interactive Advertising Bureau of Europe suggested that burdensome privacy regulations may "limit digital advertising's ability to continue to deliver a wide range of online content to users at little or no cost at the point of consumption" (IHS Technology, 2015); an earlier report by the Information Technology and Innovation Foundation stated: "[t]he evidence clearly suggests that the tradeoffs of stronger privacy laws result in less free and low-cost content and more spam (i.e., unwanted ads)" (Castro, 2010). Such an impact of data regulation on the availability of free online content would raise legitimate questions over the appropriate balance between the regulatory goal of privacy protection and other societal interests.

Despite industry claims and theoretical predictions, available evidence on the effects of privacy regulation in general, and the GDPR in particular, on ad-supported content providers is limited and contradictory, with some anecdotal reports even suggesting that online publishers could reduce reliance on behaviorally targeted advertising, post-GDPR, while continuing to enjoy stable advertising revenues (Davies, 2019). To date, no empirical study has tested the relationship between the enactment of privacy regulation such as the GDPR, content

[1] A 2013 Deloitte impact assessment report suggested that the GDPR (Deloitte, 2013) could cause a loss of around 2.8 million jobs and a reduction of European GDP by around 1.34% (corresponding to around €173 billion).

[2] Johnson *et al.* (2020) find that reductions in the ability to target advertising through "opt-out" industry self-regulatory initiatives resulted in a decrease of around $8.58 of ad spending for each consumer who chose to opt-out, borne by publishers and ad exchanges.

providers' individual responses to it, and the their ability to provide content to visitors. We construct a longitudinal data-set of 909 content providers—both large and small news and media websites located in the EU and the United States (US) and which, due to their frequent reliance on ads for traffic and revenue generation, may be particularly affected by the regulation (Competition and Markets Authority (CMA), 2020).[3] We collect two sets of data: *technical variables* and *downstream outcomes*, spanning a period of time of at least 19 months for some metrics (from April 2018 to November 2019), and longer for other metrics (April 2017 to November 2019). Technical variables are mined at regular intervals, before and after the enactment of the GDPR, by visiting each website from both EU and US IP addresses. These variables include the number of first and third party cookies used by each website, the provision of consent mechanisms, changes in privacy policies, measures of advertising intensity on the website, and so forth. These variables give us a measure of how websites responded to the GDPR over time, including how they interacted with visitors and how they managed the collection of visitors' information. Downstream outcomes are collected from multiple publicly available sources and measure the quantity and the quality of the content those websites produce. Following related work, quantity of content is measured as the number of new URLs of content generated over time; as proxies for quality, we use various metrics of user engagement previously adopted in the literature, including traffic metrics (Page Views Per User, Page Views Per Million, Reach, and Rank) and visitors' engagement (social media reactions on Facebook). Finally, we capture data on websites' market exit and changes in their revenue- and data-generating strategies that may signal regulatory-induced distress.

Our longitudinal, website-level response data allows us to paint a rich picture of the evolution of the impact of the GDPR over time and to link website responses to it to downstream outcomes. The analysis first focuses on websites' responses (Section 5). We find robust evidence of websites' reactions to the GDPR in both the US and the EU, and of significant heterogeneity in response strategies both between EU and US websites, as well as within EU websites. We also find evidence of changes—especially among EU websites—in responses over time. In particular, we find evidence of an initial reduction in the number of third-party

[3] We determine the location of each website based, primarily, on the location of its headquarters. In the analysis, we also account for the country of origin of the traffic the website receives, which—as we show—tends to be highly correlated with the top-level domain country of the website. See Section 4.

cookies and visitor tracking among both EU and US websites following the enactment of the GDPR. Those initial reductions are followed, several months after the enactment of the regulation, by a trend reversal and an uptick in tracking among EU websites. In fact, we also find evidence that an initial increase in concentration in the EU market of third-party data trackers immediately after the enactment of the GDPR (an increase consistent with post-GDPR concentration dynamics documented by prior studies of the GDPR: Johnson and Shriver (2019); Peukert *et al.* (2020)) was followed by re-entry of numerous market players several months thereafter.

Next, we estimate the impact of the GDPR on downstream outcomes (Section 6). A known complication in GDPR analysis is its extraterritorial scope: organizations located outside the EU are subject to the requirements of the GDPR when interacting with EU data subjects, clouding the distinction between control and treatment groups. While the complication is real, and spillover effects can reach US-based websites, we argue in Section 3 that it is legitimate to expect content providers located inside vs. outside the EU to differ both in the mode of response to the GDPR and in the likelihood of their downstream outcomes being affected by the regulation. The first step in our empirical strategy consists of several difference-in-differences specifications that estimate an ecosystem effect—how the regulation impacts EU websites compared to US websites, accounting both for websites location and for the share of traffic they receive from the EU or the US. Next, we complement the difference-in-differences analysis by leveraging our data-set of individual websites' responses. In contrast to other GDPR studies, by virtue of capturing websites' data from both EU and US IP addresses, we are able to determine whether content providers respond differently to the regulation depending on the country of origin of the visitor, and therefore how they differ in mode and degree of response to the regulation. Thus we are able to account for website-level effects – the effect that the GDPR has on downstream outcomes as function of the specific responses by the websites to EU visitors (for example, a website may see a decrease in the value of the ads they display after they implement a consent mechanism that induces users not to consent to tracking). To account for endogeneity in response behavior, we use two specifications: an instrumental variable approach (local average treatment effect or LATE) to estimate the effect of the regulation on websites that do decide to respond; and a look-ahead matching analysis, which compares the outcomes experienced by websites that adopt the same response to GDPR, but at different points in time, allowing us to exploit the temporal variation in adoption to identify the effect of the response.

Our econometric specifications are consistent in failing to reject the null hypothesis of no significant differences in downstream outcomes for EU and US websites. While we find a small reduction in average number of page views per user in EU websites relative to US websites near the end of the period of observation, we find no statistically significant impact of the regulation on EU websites' ability to provide content or on various proxies of content quality, such as the amount of visitors' traffic they receive and visitors' social media engagement with new content in both the short-term and the long-term. Furthermore, we find that only a small number of websites exited the market following the enactment of the GDPR - and at rates no different for EU vs. US websites. In addition, we find no differences in the likelihood of adopting data- or revenue-increasing strategies such as cookie walls or subscription options highlighted on a website's frontpage; furthermore, we find no evidence that EU websites switched to placing content behind paywalls more frequently than US websites.

Contrasted to the pessimistic predictions on the impact that privacy regulation such as the GDPR would have had on content providers, our results may appear surprising. Prior literature has actually provided intuitions for why the impact of regulations such as the GDPR may be more nuanced than conventionally believed (Section 2). We discuss mechanisms that may explain our findings in Section 7. We are able to rule out a number of possible explanations (such as strategic changes in revenue models or advertising intensity to offset regulation-induced harm). We conclude that websites that received a significant proportion of traffic from EU visitors either did not implement measurable changes (for instance, they invoked "legitimate business interest" to keep collecting visitors' data) or adjusted their responses over time, conceivably revamping data collection efforts months after the enactment of the GDPR. In contrast, websites with the stronger and longer-lasting responses to the GDPR (such as curtailing tracking over an extended period of time) were those that received only a small fraction of traffic from EU visitors and thus did not rely on EU traffic for economic success.

2 Prior Literature

The economics of privacy literature investigates trade-offs associated with the revelation or protection of personal information (Acquisti *et al.*, 2016). A sizable strand of this literature has focused on the impact of privacy regulation. Empirical works from information systems, marketing, and economics scholars have shown that the impact of privacy regulation on economic outcomes is nuanced and context specific (Goldfarb and Tucker, 2012). For instance, in the health care domain, privacy legislation may affect innovation and reduce demand for electronic medical records via a suppression of network effects (Miller and Tucker, 2009); however, if privacy regulation is coupled with appropriate incentives for patients, it may have a *positive* impact on the development and adoption of health information exchanges (Adjerid *et al.*, 2015).

The online advertising market is a natural candidate for the study of how regulatory limits imposed on consumer data collection may affect stakeholders reliant on these data. Online ads are often targeted to individuals based on information tracked and collected about them. Targeted ads are likely to be more effective than non-targeted ones (Evans, 2009). Hence, regulation that restricts advertisers' ability to collect data on users can negatively affect advertising effectiveness (Goldfarb and Tucker, 2010). Accordingly, industry advocates have warned that restrictions on the ability to collect and use consumer data for targeted advertising may be harmful to content providers and Internet users, as they would impair websites' ability to provide quality content to their visitors (Castro, 2010; IHS Technology, 2015). To our knowledge, however, the link between privacy regulation, websites' responses, and websites' ability to provide content has not yet been thoroughly vetted in empirical research.

The impact of the GDPR. The GDPR has attracted a significant amount of attention across various research fields. One of the earliest studies in this stream found that the GDPR led to a decrease in investments in EU emerging technologies compared to US organizations (Jia *et al.*, 2021). Similar negative economic outcomes of the regulation have been reported in the literature, including drops in European web traffic and e-commerce revenues (Goldberg *et al.*, 2021), an increase in concentration among web technology vendors (Johnson and Shriver, 2019), an increase in search costs among users covered by the GDPR (Zhao *et al.*, 2022), and reduction of consumer surplus on the apps market due to more apps leaving the market and fewer ones entering (Janssen *et al.*, 2022).

Despite these early empirical findings, theoretical work has raised the prospect that the economic impact of the GDPR may be more nuanced than industry expectations. Lefouili and Toh (2018) argue that the effect of the GDPR on firm investments may be mixed: regulating information might reduce investments and yet be socially desirable when information and quality are not strong complements. Choi *et al.* (2019) argue that excessive collection of personal information in the market (and the resulting excessive loss of privacy compared to the social optimum) may be mitigated by regulatory interventions. The possibility of highly nuanced and contextual effects of the GDPR that emerges from the theoretical literature is consistent with some empirical studies. For instance, Zhuo *et al.* (2021) measure the impact of the GDPR on interconnection agreements between EU network providers with those outside the EU. Although the authors note a decrease in demand for data within EU networks, they estimate zero effects of the regulation on the number of types of interconnection agreements.

Recent work has focused on the impact of the GDPR on online tracking and the online advertising market. Peukert *et al.* (2020) observed increased concentration among web technology providers following the introduction of the GDPR (the market share for small firms decreased while that of large firms such as Google increased). Shortly after the enforcement of the GDPR, Libert *et al.* (2018) reported a 22% drop in third-party cookies on news websites. Later, Dabrowski *et al.* (2019) found that EU-based visitors were less likely to receive persistent cookies compared to US visitors, even as the number of US-based visitors decreased. Urban *et al.* (2020) found that syncing cookies (which allow the exchange of users' information between online advertising actors) decreased significantly around the time the GDPR came into effect. However, the authors found that the number of syncing cookies slightly increased again over the long-term. Sørensen and Kosta (2019) found that the number of third parties on EU websites declined slightly after the GDPR (although the authors ultimately concluded that the GDPR may not necessarily be responsible for that effect). Lukic *et al.* (2022) found that the overall number of trackers and tracking providers on news, entertainment, and business websites increased after the GDRP, but the relative increase was less for GDPR compliant websites. Aridor *et al.* (2020) found that the total number of consumers observed by a data intermediary in the online travel industry decreased by 12.5% after the GDPR, suggesting that a significant number of consumers decided to opt-out. Congiu *et al.* (2022) investigate the impact of the GDPR on traffic measures for $5,000$ websites in the US and EU. The authors find a 15% reduction in overall traffic for US and EU websites along with a reduction in traffic metrics for EU websites relative to US websites. They attribute this effect to a reduction in the effectiveness of display advertising and e-mail marketing for acquiring traffic. Our study differs from that work in several dimensions. Rather than investigating how GDPR has affected the different channels that websites use to acquire visitors, our aim is to determine whether the GDPR has affected the ability of news and media websites to continue to provide content and engage audiences. Our sample focuses on news and media sites that may use advertising for revenue, but do not rely on it for sourcing traffic (on average, less than 0.5% of the traffic of the websites in our samples is sourced through display advertising).[4] While we analyze some traffic related measures, our focus is on examining how different websites have responded to comply with the GDPR, and whether their responses have influenced their downstream outcomes. We attempt to identify both ecosystem-level effects, which affect all websites in the EU, and website-level effects, which depend on the specific responses that websites adopt.

The impact of the GDPR on websites' interface features (including consent mechanisms and visitors' reactions to them) has also been investigated in recent work. As we discuss in the analysis of our results (Section 7), interface features may affect consumer reactions to the regulation, and ultimately its impact. In an extensive analysis of a large telecom provider's data, Godinho de Matos and Adjerid (2022) found that user opt-in for the disclosure of different data types increased if GDPR-compliant consent was used. However, Degeling *et al.* (2019) found that while most websites adjusted their privacy policies and implemented consent mechanisms in the months immediately following GDPR enforcement, some did not comply and did not provide users with means to meaningfully consent to tracking. Dorfleitner *et al.* (2021) found that the readability of privacy policies for German financial technology firms decreased after the GDPR, while their length and the quantity of data processed increased.

[4] Based on data from SimilarWeb a month before GDPR became effective.

Sanchez-Rola *et al.* (2019) found that, despite the presence of the opt-out mechanism, it was still difficult for users to avoid being tracked. Additionally, about 90% of the websites involved in the study placed tracking cookies on users' browsers before they were given the chance to opt-out. Utz *et al.* (2019) examined common features of consent dialogs on websites post-GDPR and found that many elements can be leveraged to nudge users to accept tracking. As our longitudinal data-set includes both websites' responses to the GDPR and downstream outcomes, we document not just the evolution over time of content providers' responses to the regulation, but also the downstream impact of regulation and websites' responses, allowing us to capture the ecosystem-level effects and website-level effects of the regulation.

Online advertising and content providers. Research in the online advertising and media literature has investigated the relationship between ad-sponsored business models, content providers' incentives, and the provision of content. Several theoretical studies have argued that when content providers are supported by advertising revenue, they have an incentive to adjust their content to maximize traffic and attract advertisers (Anderson and Gabszewicz, 2006). Empirically, Monic and Feng (2013) found that the quality of blog posts tends to increase because of ad revenue. Shiller *et al.* (2018) investigated whether the increasing adoption of ad blockers by online users might decrease the quality of online content. The authors used traffic at the website level as a proxy for quality, and found that websites with a high proportion of ad blocking visitors experienced a deterioration in traffic ranking relative to websites with fewer ad blocking visitors. Athey *et al.* (2018) showed how consumer switching—that is, consumers consuming content from multiple websites—affects advertising strategies and increases competition among publishers, leading to an increase in a publisher's incentives to invest in quality content that attracts a greater share of consumers. To our knowledge, no study has investigated the link between privacy regulation (which may affect the availability of consumer data within the online advertising ecosystem and thus the ability to behaviorally target advertising) and a diverse set of downstream outcomes of relevance to content providers, such as their ability to create new content and their success in terms of traffic and social media engagement.

3 Privacy Regulation and the Ad-Supported Publishing Ecosystem

Because they often establish rules for how consumer data can be collected and used, privacy regulations can affect the circulation of personal data in the market and the ways consumer information can be used by online content providers. Under the GDPR, two justifications are generally accepted to apply to advertising practices: "user consent" and "legitimate interest" (IAB Europe, 2021).

Under the first justification, data collection can proceed if a user (the visitor) consents to the purpose for which the data is being collected. Under the second justification, data controllers (such as websites) can collect and process data if it is necessary "for the purposes of the legitimate interests pursued by the controller." When websites use legitimate interest to justify data collection, these interests must be communicated to the data subject. Typically, this is achieved by including verbiage referring to "legitimate interest" on the website's privacy policy.

While some GDPR requirements (such as obtaining user consent or invoking legitimate interest for collecting data) already existed in European privacy law, the GDPR brought about a drastic increase in sanctions for violations and a potentially more effective system of enforcement by national data protection authorities (Peukert *et al.*, 2022).[5] In turn, a higher likelihood and magnitude of fines creates the economic incentives for firms to comply with the regulation, and therefore the conditions for downstream economic outcomes. In the following subsections, we first establish why the requirements of the GDPR may result in specific downstream effects for content providers. We then break these effects into ecosystem effects and website-level effects, and note how they may operate both separately and in combination. Throughout, we propose a series economic expectations that describe how the GDPR may impact the market of online content providers based on where they are located, where their

[5] Under the GDPR, for instance, the Luxembourg National Commission for Data Protection (CNPD) imposed a record fine of €746 million on Amazon for mis-processing personal data (`https://www.sec.gov/ix?doc=/Archives/edgar/data/0001018724/000101872421000020/amzn-20210630.htm`, pp.13), and the French privacy regulatory authority (Commission Nationale de l'Informatique et des Libertés, or CNIL) fined Google €50 million for "lack of transparency, unsatisfactory information, and lack of valid consent for the personalization of advertising" (`https://www.cnil.fr/fr/la-formation-restreinte-de-la-cnil-prononce-une-sanction-de-50-millions-deuros-lencontre-de-la`).

traffic originates from, and their response to the regulation.

3.1 Downstream Impact

Article 6 of the GDPR includes requirements that may affect downstream outcomes for entities (including websites) generating some or most of their revenues through online advertising. First, the enactment of the GDPR may reduce the extent to which customers and visitors are tracked and their personal data collected. The reduction in tracking may be due to a number of factors and manifest itself in different ways: through a website's decision to block traffic from EU visitors altogether (in order to avoid potential fines associated with GDPR violations); through the decision by data controllers (such as websites) to reduce or altogether stop the tracking of EU customers or visitors (again, to avoid potential fines associated with GDPR violations); or through the adoption of consent mechanisms, which present users with options concerning the usage of their data (rather than assuming visitor consent and tracking their behavior by default) and which *may* therefore increase the portion of users who opt-out from tracking and targeting relative to the pre-GDPR *status quo*.[6]

A reduction in the ability to track data subjects will, in turn, adversely affect websites' ability to target visitors with personalized ads, as well as to target personalized ads on other channels in order to attract traffic. A reduction in the ability to personalize ads can decrease their effectiveness and their value. Personal information increases targeting efficiency, and ads that are tailored to visitors' preferences are more valuable (Tucker, 2012). Non-targeted impressions may therefore receive lower bids in ad auctions (Beales, 2010). Furthermore, a reduction in the ability to collect and use visitors' data would decrease the number of targeted impressions within ad auctions. Thus, online advertising may become less profitable as whole (Goldfarb and Tucker, 2011). As a result, websites that provide content may receive lower payments from selling advertising space for non-behaviorally targeted impressions (Sharma *et al.*, 2019), and may be less successful in converting traffic from the ads they purchase on other channels. Hence, overall revenues of websites may decrease (Lambrecht *et al.*, 2014).

[6] In Section 5, we present evidence of all those responses arising within our sample of websites. A reduction in tracking may be associated with decisions made by both individual websites and the online advertising/publishing ecosystem as a whole (for instance, data intermediaries such as advertising networks); these decisions may subsequently affect the websites and the ecosystem they are part of.

News and media websites, which typically rely on advertising for traffic and revenue generation, may be especially affected (Cook and Sirkkunen, 2013). Finally, revenue reduction may impact content provision (Angelucci and Cagé, 2019). Existing work has documented the prevalence of ad-sponsored business models among these websites (Casadesus-Masanell and Zhu, 2013; Goldfarb, 2004; Lambrecht *et al.*, 2014). Both theoretical and empirical works (pre-GDPR) have tied providers' content quality to advertising revenues (Anderson and Gabszewicz, 2006; Monic and Feng, 2013). In response to reduced revenue, websites may not be able to sustain the quantity and quality of output (content) they generated before the regulatory shock (Downes, 2018). The magnitude of these effects may vary both based on websites' location and their specific responses. Revenue-side effects may be compounded by cost-side effects. Compliance costs may have increased following the GDPR, further reducing profitability and the ability to generate new content.

Our empirical strategy focuses on capturing metrics correlated with websites' ability to provide new, quality content. The metrics include variables used in the previous literature, such as the amount of new content URLs generated by online publishers over time, the volume of traffic they receive, and the degree of social media engagement with new content (Shiller *et al.*, 2018; Gallea and Rohner, 2021; Ferreira *et al.*, 2021). In addition, we capture information such as content providers' exit from the market, and changes in revenue- and data-generating strategies, which may also signal websites' distress with the impact of the GDPR.

The chain of reactions espoused in the previous paragraphs may, however, *not* materialize for a number of reasons. Theoretical economic work considered in Section 2 suggests that the impact of regulations such as the GDPR on the ad-supported publishing ecosystem (and for online publishers in particular) may be more nuanced than the negative chain of events envisioned above. For instance, the GDPR may not, in fact, significantly reduce the ability to personalize ads if data holders found ways to keep capturing data without violating the regulation (for instance, by invoking legitimate interest; or because most visitors consent to tracking; or because websites adopt interface patterns in consent mechanisms that nudge visitors towards consent). Or, data available in the ecosystem may be reduced—and yet, under the new post-GDPR equilibria in the advertising market, online ad auction bids (and therefore the price of ads websites sell) may not significantly drop in response if global online ad spending does not change.[7] Under such alternative scenarios, we may expect to find that the GDPR did not, in fact, significantly affect websites' ability to provide content. We consider some of these possibilities in Section 7 as potential explanations for our findings.

3.1.1 Ecosystem Effects

By *ecosystem effects* we refer to the differential impact the GDPR may have on US vs. EU websites, as a function of website location and, therefore, the share of traffic a website receives from EU vs non-EU data subjects. Ecosystem effects materialize through two mechanisms. First, while the GDPR applies to both EU and non-EU websites,[8] EU-based data controllers (such as websites and advertising networks) are in principle required to comply with GDPR rules for *all* users and visitors, whereas data controllers based outside the EU can choose to apply GDPR-compliant practices only to the share of visitors originating from the EU. As non-EU websites are likely to receive a smaller portion of their traffic from the EU, relative to EU websites, they may opt to comply with the regulation for smaller shares of visitors. Hence, at parity of GDPR response relative to an EU counterpart, a US website may experience the chain of decline in tracking, targeting, and revenues described in Section 3.1 for smaller portions of its traffic than that counterpart, and may end up being relatively less affected by it.[9] Second, the responses to the GDPR by all other stakeholders within the EU (vs. US) online advertising/publishing ecosystems (such as individual websites, ad networks, and so forth) *collectively* affect the aggregate availability of consumer data within that ecosystem. Such changes in data availability can in turn affect the ability of individual websites within that ecosystem to target their respective visitors with behavioral ads.[10] Hence we expect EU

[7] Aridor *et al.* (2020) find that "the ability to predict consumer behavior by the intermediary's proprietary machine learning algorithm does not significantly worsen as a result of the changes induced by GDPR."

[8] The GDPR's regulatory scope encompasses any entity that operates in the EU or collects the personal data of EU data subjects (GDPR Article 1). Since the GDPR is extraterritorial in its scope (GDPR Article 3(2)), non-EU websites that utilize behavioral advertising and accept traffic from EU data subjects are subject to the requirements of the GDPR when interacting with EU visitors.

[9] This mechanism relies on two assumptions: US content providers receive different (in fact, significantly lower) shares of traffic from the EU than EU content providers; and US content providers respond differently (in terms of tracking, consent mechanisms, and so forth) to their US visitors relative to EU visitors. We test—and find support—for both assumptions in Section 5.2.2.

[10] If more stakeholders in the EU ecosystem respond to the GDPR by limiting data collection, relative to the US ecosystem, less personal information may become available in the EU ecosystem, and fewer individuals may be precisely profiled for behavioral advertising when they visit any given website in that ecosystem—affecting

websites, on average, to be more negatively affected by this mechanism than US websites, as the GDPR should more significantly reduce tracking and data availability across the EU-based data ecosystem, thus affecting all websites in it, regardless of their response.

In sum, while the GDPR is extraterritorial in scope, it is legitimate to expect that US-based and EU-based content providers would be differentially affected by it in terms of personal data acquisition and revenue generation. We expect the magnitude of ecosystem effects on a website's downstream outcomes to be moderated both by where the website is based and by the share of each website's traffic that originates from the EU, independently of a given website's GDPR response. Ultimately, we expect websites located in the EU, and websites with a higher percentage of EU-based traffic, to be more affected than websites with a lower percentage of EU traffic.

3.1.2 Website-Level Effects

By *website-level effects* we refer to the process through which responses to the GDPR by a given website may individually affect that website's ability to collect visitors' data and/or use it for behavioral advertising. At the website level, the personal information collected during each visit enables both the tracking of visitors and the targeting of ads to visitors on that website (the targeting, itself, may rely on a combination of user data coming from both the website and its partners in the ecosystem). Thus, websites' responses that limit data collection or usage (for instance, the adoption of consent mechanisms that allow visitors to opt-out of tracking or targeting) may affect that website's revenues from ads (Sharma *et al.*, 2019) and, ultimately, its ability to provide quality content.

Different websites' responses to the GDPR may have heterogeneous repercussions on the chain of effects discussed at the start of Section 3.1, and thus may disparately affect downstream outcomes. Below, we define five categories of website-level responses, starting with the responses more likely to curtail a websites' access to visitors' data, and ending with (arguably) less aggressive responses (see Appendix A, Figures 8-11 for examples).

First, faced with the compliance burden imposed by the GDPR, websites—especially those based outside the EU—have the option of exiting the EU data and advertising market

that website's revenues independently of its own distinct GDPR response.

altogether by blocking all traffic originating from the EU. In our analysis we refer to this response as *"Blocks EU."* This response could arguably be considered the most aggressive, as cutting off EU visitors would directly curtail potential future advertising revenue. This option may be attractive for websites based outside the EU which received only a small share of traffic from the EU prior to May 2018. The exit of these websites from the EU market may lead their former visitors to visit other websites, but would likely not result in large negative ecosystem effects on the tracking ability or targeting accuracy of other websites.

Second, websites may respond to the GDPR by curtailing the tracking and targeting of EU visitors while still allowing them to browse their content. This response may also negatively affect a website's advertising revenues, although arguably not as intensely as blocking EU visitors, as non-targeted ads may still be shown to EU users. We refer to this response as *"Stops EU Tracking."*

Third, websites may display consent mechanisms to visitors for the purpose of obtaining user consent to engage in data collection and data usage. We collectively refer to this type of response as *"Consent Mechanism."* Some implementations of consent mechanisms can diminish websites' ability to collect personal information—albeit arguably to a lesser extent than the unilateral curtailing of tracking and targeting by a website. Visitors to websites that implement consent dialogs may, for instance, not consent to tracking for the purposes of targeted advertising. From the perspective of websites, these visitors would no longer be linkable with interest profiles used for targeting ads. These effects can vary in magnitude depending on the specific manner in which websites choose to implement consent dialogs, including interface features and the possible deployment of dark patterns (Acquisti *et al.*, 2017). Websites that implement consent dialogs that make it easier for users to deny consent for tracking (such as dialogs that require only a single step to reject tracking) may incur stronger negative effects on tracking and targeting ability compared to websites that make denying consent for tracking more difficult (such as websites that implement consent dialogs that require multiple steps to reject tracking).

Fourth, websites may attempt to minimize the impact of having to implement consent mechanisms by instituting cookie walls. Cookie walls force users to consent to tracking before allowing them to view content. By forcing consent, these websites may not see a decrease in their ability to track visitors. However, visitors who do not wish to be tracked may react to the appearance of a cookie wall by turning away from the website altogether. Although the legality of this response under the GDPR is unclear (UK Information Commissioner's Office, 2019; Autoriteit Persoonsgegevens, 2019b), we observe multiple websites using cookie walls in our data (Section 5). We refer to this response as *"Cookie Wall."*

Fifth, websites may not take direct actions in response to the GDPR. This category is broad. Some websites may elect to not curtail tracking nor implement consent mechanisms, but rather invoke legitimate interest (see Section 3) to justify continuing their present data collection and usage practices.[11] Other websites may simply continue to comply with older EU privacy directives, merely displaying "cookie notices" (also known as cookie banners) which often appear as banners at the bottom of websites.[12] Functionally, the effects on tracking and targeting for the websites invoking legitimate interest and websites not bothering to do so are similar: either way, these websites do not engage in changes that are likely to affect their ability to track their visitors (in fact, they may end up benefiting from the reduced tracking ability of other websites, as a decrease in tracked advertising inventory may drive up advertisers' willingness to pay). Theoretically, they may experience a reduction in traffic from privacy-conscious and aware visitors who dislike the imposition of tracking, without consent, based on the legitimate interest rationale. In practice, one may expect this category of websites to experience the mildest effect on tracking, targeting, revenues, and thus on downstream outcomes. For our analysis, as these various responses are less likely than others to have a significant impact, we group them together and refer to them as *"No Response or Legitimate Interest."*

It is possible for a website to adopt more than one response at the same time (for example, we document in our analysis instances of websites implementing a consent mechanism while also invoking legitimate interest), as well as different responses at different moments in

[11] The legality of this justification for tracking is contested, and the compliance risk is potentially high. As early as 2019, the UK Information Commissioner's Office published an opinion stating that legitimate interest cannot be used as a legal basis for data collection in the context of behavioral advertising (UK Information Commissioner's Office, 2019). This has grown into a consensus among regulators and industry over time. In early 2021, IAB Europe published guidance stating that legitimate interest cannot be used as a basis for setting tracking cookies IAB Europe (2021).

[12] The banners inform users of the presence of cookies on a website. They are distinct from other privacy notices in that they do not ask for consent prior to tracking or notify users of legitimate interest claims.

time.

4 Data

We constructed a longitudinal panel with a sample of news and media websites located in the US and in several EU countries (Germany, France, UK, Italy, Spain, and the Netherlands). To select the websites to include in the panel, we used Amazon's Alexa Internet web metrics (`https://www.alexa.com/`) to identify the top 500 websites in world, in the US, and in each of the EU countries listed above, and augmented this set with a random sample of websites from Alexa's global top 1 million sites globally, considering only those that were from the US or one of the EU countries we previously identified. Next, we used SimilarWeb (`https://www.similarweb.com/`) to identify which of these websites were classified as "News and Media" and were located in the US or the EU. To determine the location of a website, we used the location of its headquarters as reported by SimilarWeb. When this information was not available, we inferred the location of a website by the country of the website's top-level domain (such as .fr or .us). If the website used a top level domain that was not country-specific (e.g., .com), we assigned a country based on where the most visitors originated from.[13] The resulting sample contains 909 websites news and media websites containing both top-ranked and long-tail (low-ranked) content providers. We provide a detailed description of sampling strategy in the Appendix B.

4.1 Technical Variables

For each website in the sample we captured both technical variables (discussed here) and downstream outcomes (discussed in Section 4.2) at regular time intervals. To construct technical variables, we mined several classes of website data by visiting each website using OpenWPM (a web privacy measurement framework: Englehardt and Narayanan (2016)), simulating a user browsing from a desktop. We refer to each round of visits as a "wave" of data collection. Each wave required on average between 4 and 5 days to complete collection of the various classes of data from all websites in the sample, with an average period of 45 days between

[13] The results we present in this manuscript are robust to classifying websites solely based on the origin of the majority of the traffic they receive.

the waves. The data collected spans a period of time of over 19 months (from April 2018 to November 2019). During each wave, we visited every website twice, simultaneously, from two different visitor IP addresses, one located in Europe (France) and one in the US. This design allows us to compare, before and after the enactment of the GDPR, whether and how websites adapted their data collection behavior according to the geographical location of a visitor.

Technical variables are constructed based on the different types of raw website data we mined over time. They measure websites' interactions with their visitors, including tracking activities, advertising choices, the provision of consent mechanisms, and privacy policies. We use technical variables to construct categories of website responses to the GDPR. The raw website data include: over 5.5M cookies (including first- and third-party cookies) set by the websites on visitors' browsers; more than 40M HTTP responses (including all the information exchanged between the browser and the websites visited), which we use to measure websites' advertising patterns; over 20,000 screenshots (including visual interface elements such as consent mechanisms, buttons to accept cookies, user-facing messaging, or subscription options), which we use to classify visual elements of websites that may indicate a website's response to the GDPR; and HTML data (including over 18,000 privacy notices), which we use to detect a number of variables, including references to legitimate interest to justify data collection, and websites' usage of paywalls.[14] The technical variables are discussed below, and are used in Section 5.2 to construct categories of websites' responses to the GDPR. Details on their construction can be found in Appendix B.

Cookies: The variable *1st Party Cookies* counts the number of cookies set by the website being browsed. The variable *3rd Party Cookies* counts cookies that are set by entities other than the original website and that could be used to track users' behavior across websites, construct users' profiles, and improve the targeting of behavioral ads. We identify which of these cookies are known to be used for tracking or advertising using scripts included in popular ad-blockers that flag advertising content (see Appendix B). The variable *Advertising Cookies* counts the number of cookies set by advertising companies, and the variable *Tracking Cookies*

[14] Paywalls can be used by websites to restrict access to content to paying users.

counts the number of identified cookies used for user tracking. We rely on a drop to zero in either advertising or tracking cookies to identify when a website responds to the GDPR by halting the tracking of (EU) visitors and to create a dummy website response variable (*Stops EU Tracking*).

Advertising Intensity: To analyze the volume of advertising displayed to a website's visitors, we measure the amount of HTML content related to advertising. The variable *Advertising Intensity* captures the size, in kilobytes, of the quantity of advertising content on a website's homepage.[15] We identify advertising content using the same method explained above for cookies.

Interactions with Visitors: We manually inspected and labelled over 20k captured websites' screenshots to determine how websites' interaction with their visitors evolved in response to the GDPR. We use screenshots to distinguish between websites that block EU visitors, implement consent mechanisms, or use cookie walls and cookie banners.[16] We are able to identify US websites that decided to block EU visitors by spotting static pages shown to EU visitors informing them that the website is unavailable (see Figure 8 in Appendix A). We consider a consent mechanism to be a banner or pop-up that offers users the ability to reject tracking. This can be either through a "reject" button or through sub-menus such as a "settings" menu (for example Figure 9a and Figure 9b in Appendix A). By contrast, cookie banners inform users about cookies, but do not provide them with a way to reject tracking (see Figure 11 in Appendix A). We identify cookie walls by virtue of the fact that they prevent visitors from viewing content and do not provide a means (through buttons or links) to reject tracking (see Figure 10 in Appendix A). For each of the responses so identified, we create a dummy website response variable (*Blocks EU Visitors, Consent Mechanism, Cookie Wall, Cookie Banner*) that takes on the value 1 if the corresponding response is implemented by a given website, and 0 otherwise at a particular point in time. We then use those variables to define categories of GDPR responses in Section 5.2.

[15] We constructed other advertising metrics, such as the number of ads on the page or the type of ads (video, image). The results presented in the rest of the manuscripts are consistent across different specifications of the *Advertising Intensity* variable.

[16] As noted above, we use cookie data to determine which websites halt the tracking of visitors.

Legitimate Interest: We analyze websites' HTML to extract their privacy policies and track their changes over time. We use text analysis on over 18,000 privacy notices to infer which websites invoke legitimate business interest as a justification for data collection. We use references to legitimate business interest in the construction of an hybrid dummy website response variable *No Response or Legitimate Interest* (see Section 5.2).

4.2 Downstream Outcomes

We use third-party repositories to measure the quantity of content generated by websites over time and user engagement with that content (a proxy for its quality). These metrics do not change as function of the country of the visitor and they are aggregated across different sources. We collect these metrics from April 2017 to November 2019.

Content Quantity: To measure content quantity, we use the Global Database of Events, Language, and Tone, or GDELT (`https://www.gdeltproject.org`). GDELT gathers and provides metadata for articles from news and media websites going back to 2015 from both domestic (US) and international sources. The database provides metadata including the URL, publication date, and publisher website for each article, and has been used in studies that examined global events (Gallea and Rohner, 2021; Ferreira *et al.*, 2021). We use GDELT data to count the number of new URLs of content (*GDELT URLs*) published by each website in the sample in the week surrounding each wave of data collection (three days before and after each OpenWPM observation).

User Engagement: As proxies for content quality, we use two sets of metrics that capture user engagement with websites' content: web traffic metrics and social media reactions. Following prior work (Luo and Zhang, 2013; Shiller *et al.*, 2018; Utz *et al.*, 2019; Sørensen and Kosta, 2019), we use Alexa web metrics to measure user traffic. The underlying premise is that, were the quality of the content provided by a website to decrease, users might substitute for other content and, therefore, we should observe a decrease in the number of visits to a given website. We use *Reach Per Million*, a measure of the number of unique users visiting a website;[17] *Page Views Per Million*, a measure of the number of pages viewed by visitors; *Page Views Per User*, which represents the average number of unique pages viewed per user, per day, by the users visiting a website; and *Rank*, a measure of a website's popularity that combines measures of page views and unique visitors.

Following Cagé *et al.* (2020), we complement Alexa's data by mining the Facebook Graph API to capture social media reactions related to the content published by websites in the sample. For each URL posted by each website during the week surrounding the data collection in each wave (as retrieved via GDELT), we collect the number of reactions on Facebook and calculate their average number across all new URLs by website/wave. We call this the *FB Average Reaction*.

Table 1: Descriptive Statistics — Before the GDPR

	Mean	Std. Dev.	Min	Max	N
	Technical variables				
Tracking:					
1st Party Cookies EU Visitor	12.502	8.133	0.0	45.0	3,367
3rd Party Cookies EU Visitor	47.715	44.264	0.0	272.0	3,367
1st Party Cookies US Visitor	12.981	8.476	0.0	45.0	3,396
3rd Party Cookies US Visitor	51.958	47.143	0.0	281.0	3,396
Advertising Cookies EU visitor	23.618	25.991	0.0	180.0	3,367
Tracking Cookies EU visitor	15.116	13.589	0.0	102.0	3,367
Advertising:					
Advertising Intensity (KB) EU Visitor	680.697	675.452	0.0	5,941.6	3,367
Advertising Intensity (KB) US Visitor	725.949	785.895	0.0	9,321.6	3,396
Website Visitors:					
Share of EU Visitors	0.430	0.420	0.0	1.0	13,624
Share of US Visitors	0.395	0.403	0.0	1.0	13,624
	Downstream Outcomes				
Log GDELT URLs	5.072	1.699	0.0	9.6	11,218
Reach Per Million	266.301	924.108	0.9	18,714.3	13,624
Page Views Per Million	15.380	64.167	0.0	1,451.4	13,624
Page Views Per User	2.059	0.931	0.6	14.7	13,624
Rank	58,734.104	93,125.165	0.0	1,729,171.1	13,624
FB Average Reaction	105.665	459.430	0.0	10,690.5	11,218

Notes: This table presents descriptive statistics before the enactment of the GDPR. The technical variables were collected for the first time in April 2018, while the downstream variables were collected from April 2017, which explains the difference in the number of observations.

[17] Unique visitors are determined by the number of unique Alexa users who visit a website on a given day.

5 Empirical Patterns

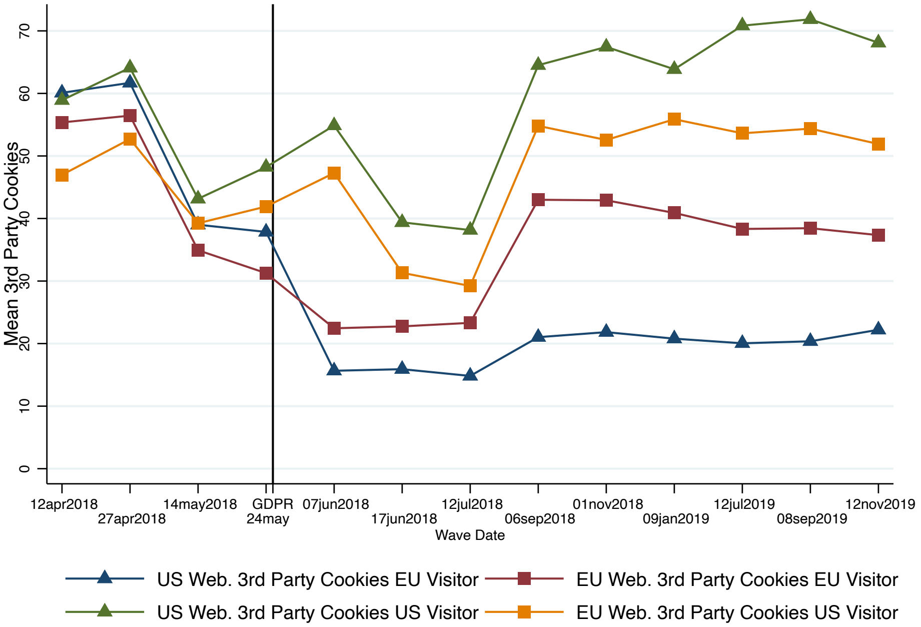
Table 1 presents descriptive statistics for technical variables and downstream outcomes across the whole sample, before the implementation of the GDPR. In the rest of this section, we first investigate whether the GDPR had any effect on the extent of tracking and advertising in EU versus US websites (Section 5.1). We then consider the five different website-level responses previously described under our theoretical framework, and describe the characteristics of the websites adopting them, as well as how those responses evolved over time (Section 5.2). Finally, we discuss changes in downstream outcomes following the enactment of the GDPR 5.3.

5.1 Changes in Cookies and Advertising Patterns

We start by analyzing changes in cookies and advertising patterns for the websites in our sample after the GDPR became effective. We contrast EU- versus US-based websites and how the results change if the websites are browsed by EU- or US-based visitors.

We first consider third-party cookies, which are typically used to track users across websites. Figure 1 shows how, before the GDPR, the number of third-party cookies used by EU and US websites were similar for both EU and US visitors. Shortly before the GDPR came into effect, we observe a drop in the number of third-party cookies being used in EU/US websites for both EU/US visitors. Right after the GDPR became effective, the sharpest drop happens in US websites for EU visitors, followed by EU websites for EU visitors. However, these drops are short lived: we observe a rebound in the number of third-party cookies set by websites roughly three months after the GDPR became effective. The rebound is not the same for all websites and visitors. US websites continue to set, for EU visitors, a much lower number of third-party cookies than they did before the GDPR. In the case of EU websites visited from the EU, however, the number of third-party cookies rebounds to pre-GDPR levels.

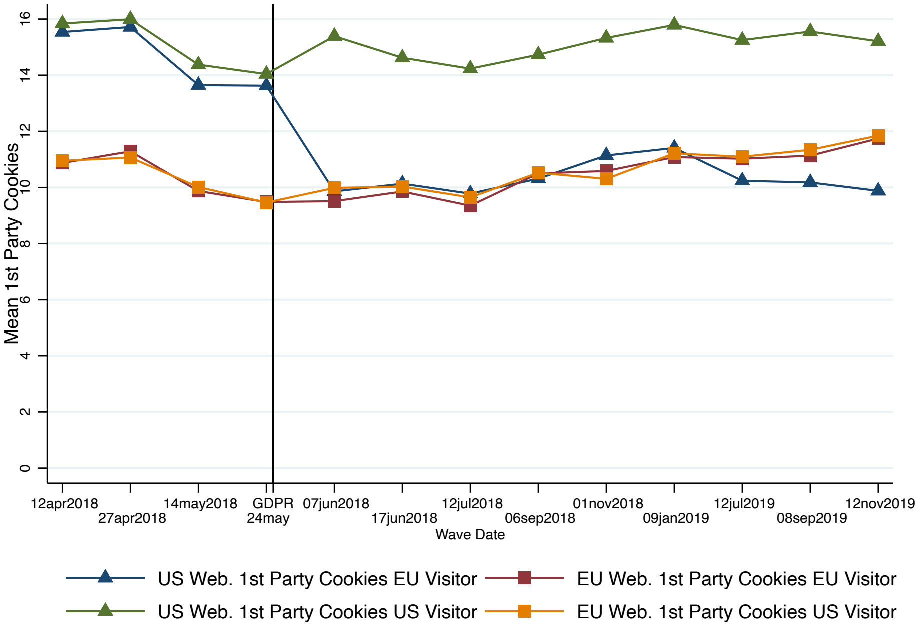
Fig. 1 *3rd Party Cookies Set by EU/US Websites for EU/US Visitors*



Next, we examine whether the number of first-party cookies used by websites changed over time. While third-party cookies are typically used to track users across websites, first-party cookies are typically related to particular websites' functionalities. For example, a website may use first-party cookies to remember visitors' login information, products they have browsed, or news articles they have read. However, since first-party cookies can also be used for advertising purposes, we are interested in examining whether third-party cookies are being replaced by first-party cookies for that purpose (such an option was introduced by Facebook in 2018: Flynn (2018)). Figure 2 suggests that the number of first-party cookies set by websites remains unchanged over time, except for the case of US websites when visited from the EU, for which we observe a persistent drop after the GDPR.[18] It is also clear that EU websites seem to set, on average, fewer first-party cookies than US websites.

[18] The drop persists also when we exclude websites that block EU traffic.

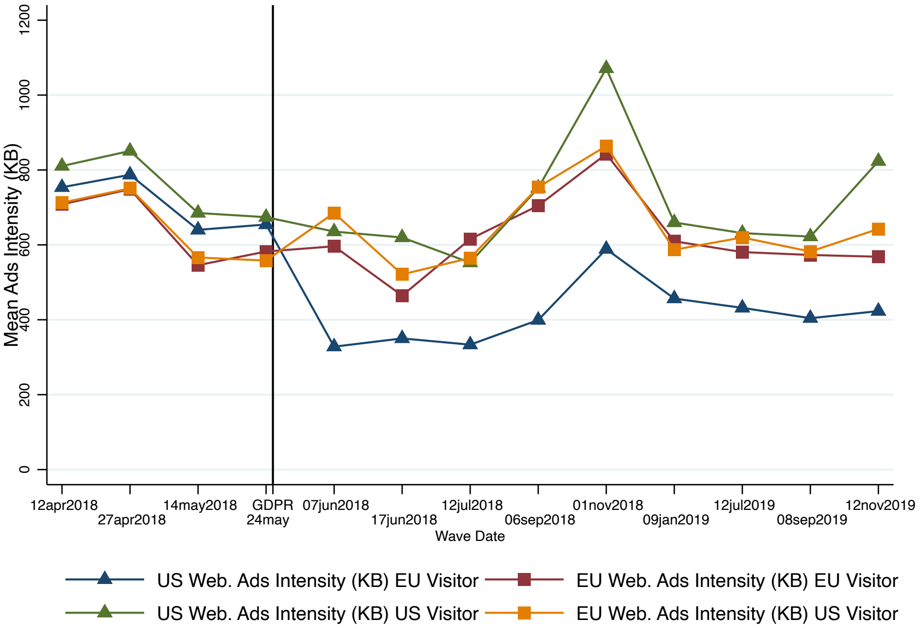
Fig. 2 *1st Party Cookies Set by EU/US Websites for EU/US Visitors*



While third-party cookies are typically used by advertising technology firms to track users across websites, they can also be used for other purposes. To get a more precise measure of the amount of data being sent to websites' visitors for advertising purposes, and the reliance of websites in our sample on advertising before and after the GDPR, we explore how advertising intensity (as defined in Section 4.1) evolved over time (Figure 3). EU websites experience a drop right before the GDPR, followed by seasonal fluctuations (such as the peak around the 2018 Christmas shopping season) and ending on levels not dissimilar from pre-GDPR advertising intensity. Within US websites, the response is more nuanced and dependent on the country of origin of the visitor. Although over the long-term advertising intensity for US visitors seems to return to pre-GDPR levels, it remains at a much lower level for EU visitors. This downward trend is robust to the exclusion of the fraction of US websites that blocked EU traffic.

In short, both tracking and advertising patterns reveal subtle but meaningful differences in EU vs. US websites' behaviors before and after the GDPR. While the shorter-term findings we present are consistent with the post-GDPR concentration dynamics documented in earlier literature on the GDPR (Johnson and Shriver, 2019; Peukert *et al.*, 2020), extending the analysis over time reveals an increase in tracking and a return of companies to the online tracking market several months after the enactment of the GDPR.

Fig. 3 *Advertising Intensity on EU/US Websites for EU/US Visitors*



5.2 Website-Level Responses

Table 2 shows descriptive statistics of pre-GDPR website-level characteristics for the five website-level responses we identified in Section 3.1.2. Websites are clustered based on their most prevalent response when browsed by a EU visitor.

Table 2: US and EU Websites Characteristics Before the GDPR (Based on Their Most Prevalent Response to GDPR)

	EU Websites				US Websites				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Stops EU Tracking	Consent Mechanism	Cookie Wall	No Response or Legitimate Interest	Blocks EU	Stops EU Tracking	Consent Mechanism	Cookie Wall	No Response or Legitimate Interest
	mean/sd	mean/sd	mean/sd	mean/sd	mean/sd	mean/sd	mean/sd	mean/sd	mean/sd
Websites characteristics									
Rank	99,453.26	36,796.21	85,299.29	31,347.35	102,979.21	102,081.91	23,609.57	2,024.84	72,057.52
	(136,676.03)	(77,245.86)	(70,021.51)	(51,685.60)	(75,067.00)	(109,215.09)	(59,542.12)	(2,315.01)	(107,602.52)
Share of EU Visitors	0.82	0.78	0.95	0.81	0.01	0.04	0.07	0.05	0.03
	(0.26)	(0.23)	(0.05)	(0.23)	(0.01)	(0.05)	(0.12)	(0.02)	(0.04)
Share of US Visitors	0.02	0.04	0.00	0.03	0.91	0.77	0.64	0.67	0.79
	(0.05)	(0.07)	(0.00)	(0.11)	(0.04)	(0.23)	(0.22)	(0.13)	(0.18)
Ads Intensity (KB) EU Visitor	140.23	784.85	231.88	716.88	1127.23	513.40	681.32	916.53	729.10
	(308.61)	(779.44)	(359.03)	(671.94)	(707.60)	(532.07)	(546.22)	(644.71)	(649.53)
Cookie Banner EU Visitor	0.11	0.10	0.05	0.11	0.00	0.02	0.02	0.09	0.01
	(0.31)	(0.31)	(0.22)	(0.31)	(0.00)	(0.12)	(0.15)	(0.28)	(0.09)
Privacy									
3rd Party Cookies EU Visitor	7.29	56.20	14.60	49.13	59.60	42.72	60.28	82.67	48.01
	(17.86)	(46.92)	(26.06)	(45.01)	(30.02)	(43.13)	(43.85)	(56.69)	(41.99)
1st Party Cookies EU Visitor	4.84	12.23	5.45	10.89	10.51	13.57	16.04	14.48	15.74
	(3.58)	(6.39)	(4.37)	(6.75)	(5.71)	(8.93)	(9.45)	(4.08)	(9.07)
Advertising Cookies EU Visitor	3.16	29.38	6.92	23.97	29.30	21.69	29.26	35.59	23.01
	(9.15)	(29.52)	(14.45)	(25.79)	(18.10)	(25.14)	(25.38)	(24.79)	(25.29)
Tracking Cookies EU Visitor	1.88	16.76	4.00	15.21	21.36	12.73	19.97	31.85	16.05
	(5.64)	(13.36)	(7.84)	(13.07)	(9.47)	(13.33)	(13.80)	(20.37)	(13.46)
Obs.	840	2,850	285	2,983	673	1,650	1,019	105	3,219
Unique websites	56	190	19	199	45	110	68	7	215

Blocks EU Visitors. A number of websites (45) exit the EU market altogether by blocking EU visitors' access. The websites in our sample that implement such a response are all US-based and the overwhelming majority of their visitors are US visitors. Before the GDPR, US websites blocking EU visitors received, on average, 91% of their visits from the US, while US websites not blocking EU visitors received 76.4% of their traffic from US visitors. This type of response was quickly implemented after the GDPR, and the share of US websites using this strategy remains fairly constant over time. Websites that block EU visitors rank lower than other websites (and therefore receive less traffic) and seem to rely on advertising to a greater extent than other websites.

Stops EU Tracking. Instead of blocking EU visitors, websites may choose to stop tracking EU visitors after the enactment of the GDPR. We identify all websites that either decrease their number of third-party cookies to zero or decrease both advertising and tracking cookies to zero (we also include in this group websites that, before the GDPR, were not using third-party cookies and continue not doing so after the GDPR). In our sample, 110 out of 445 US websites stop tracking EU visitors, and only 56 out of 464 EU websites do so. US websites that decide to stop EU tracking have a larger proportion of EU visitors than US websites that decide to block EU visitors, but seem to rely less on advertising. EU websites that decide not to track have a large share of EU visitors, and their average advertising intensity is much lower than EU websites that respond in other ways.

Consent Mechanism. Before the GDPR, we find that almost no US websites implemented consent mechanisms, while about 16.8% of EU websites did. Over time, we observe that the presence of consent mechanisms sharply increases for EU visitors on EU and US websites right before the GDPR became effective, and continues to rise until reaching a stable level with nearly 60% of EU websites in our sample using them (see Figure 14 in Appendix C). Websites that choose to implement a consent mechanism tend to be highly ranked (thus, have more traffic), compared to the other groups; they also have a sizeable share of EU visitors (both in the case of US and EU websites) and have a greater reliance on advertising, as suggested by the average advertisement intensities on their websites.

Cookie Wall. About 4.4% of EU websites and 1% of US websites use cookie walls that force users to consent to tracking in order to access the website's content. EU websites that fall in this category have a large proportion of EU visitors, but tend to be smaller websites (by ranking) and do not rely as much on advertising. The US websites that fall in this category tend, instead, to perform better in terms of ranking (they have more overall traffic) and rely heavily on advertising.

Legitimate Interest or No Response. The last response category includes websites that claim legitimate interest (and therefore continue to collect or use data as before the GDPR) or simply decide to not actively respond to the GDPR in a manner detectable by our metrics. About 35% of EU websites and 37% of US websites fall into this group. Among those, we are able to identify the portion of websites which, specifically, invoke legitimate interest by collecting and analyzing websites' privacy policies. We are able to collect privacy policies for about 45% of the observations in this category; among those, about 18.7% included language suggesting the websites' reliance on legitimate interest. For the purpose of our analysis, we combine these two types of responses (legitimate interest and no response), as there are reasons to expect that websites in these categories will experience similar effects following the enactment of the GDPR.[19] We do not expect major website-level effects for this group of content providers, because these websites do not actively change the way they interact with visitors following the GDPR. If they do experience any impact from GDPR, however, this may be attributed to ecosystem-level effects: although these websites do not adopt any action to curtail tracking, they may still be impacted if there is an overall reduction of visitors' data in the ecosystem these websites are part of, due to the actions of other stakeholders. EU websites that fall in this category are higher ranking (i.e., more traffic) compared to the other websites and rely heavily on advertising (Table 2). US websites that fall into this category also rely considerably on advertising but rank lower compared to US websites that decide to implement a consent mechanism. They rank higher than US websites that decide to not track or completely block EU visitors.

5.2.1 Evolution of Responses over Time

EU and US websites' responses to the GDPR evolved over time in different ways. Figure 4 and Figure 5 summarize the dynamics of websites' responses using Sankey diagrams. To make the diagrams readable, we divide the post-GDPR period into three time windows. The first is from May 24, 2018 to July 2018; the second from September 2019 to January 2019; and the last from July 2019 to November 2019. The grey area captures the magnitude of movements from one response to another.

EU websites clearly exhibit significant variation in response strategies over time (Figure 4). The number of EU websites identified in the group *No Response or Legitimate Interest* steadily decreases over the three periods. Most of this drop is explained by a steadily increasing number of websites using consent mechanisms in response to the GDPR. In contrast, Figure 5 illustrates that US websites' responses tend to be more stable over time.

[19] The results presented in Section 6 are robust to separating websites that invoke legitimate interest from websites for which we did not detect any type of response to the GDPR.

Fig. 4 *Sankey Diagram: Evolution of EU Website-Level Responses over Time*

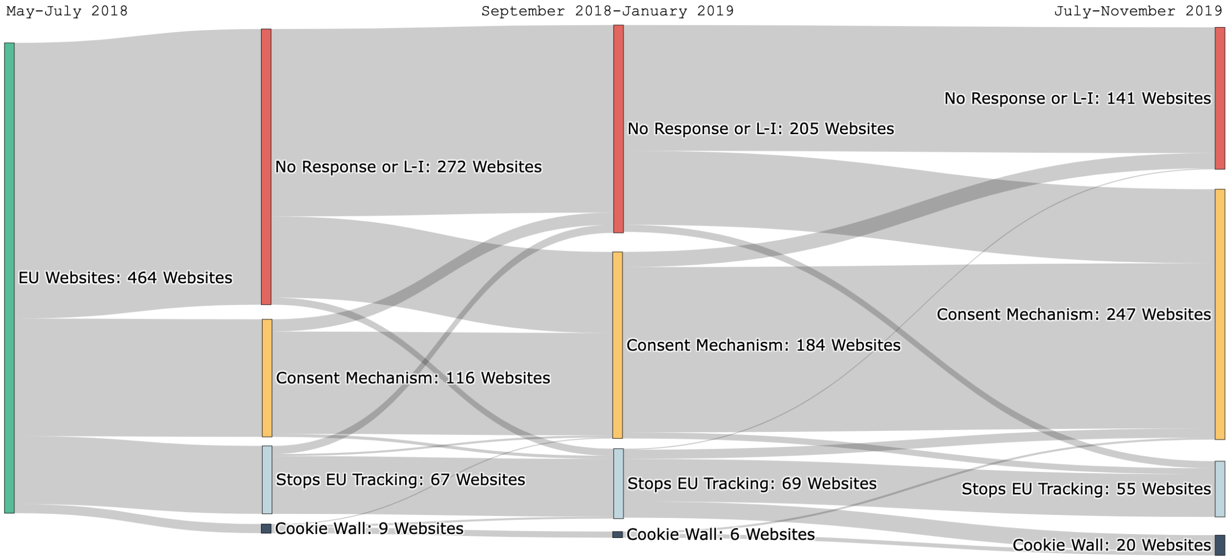
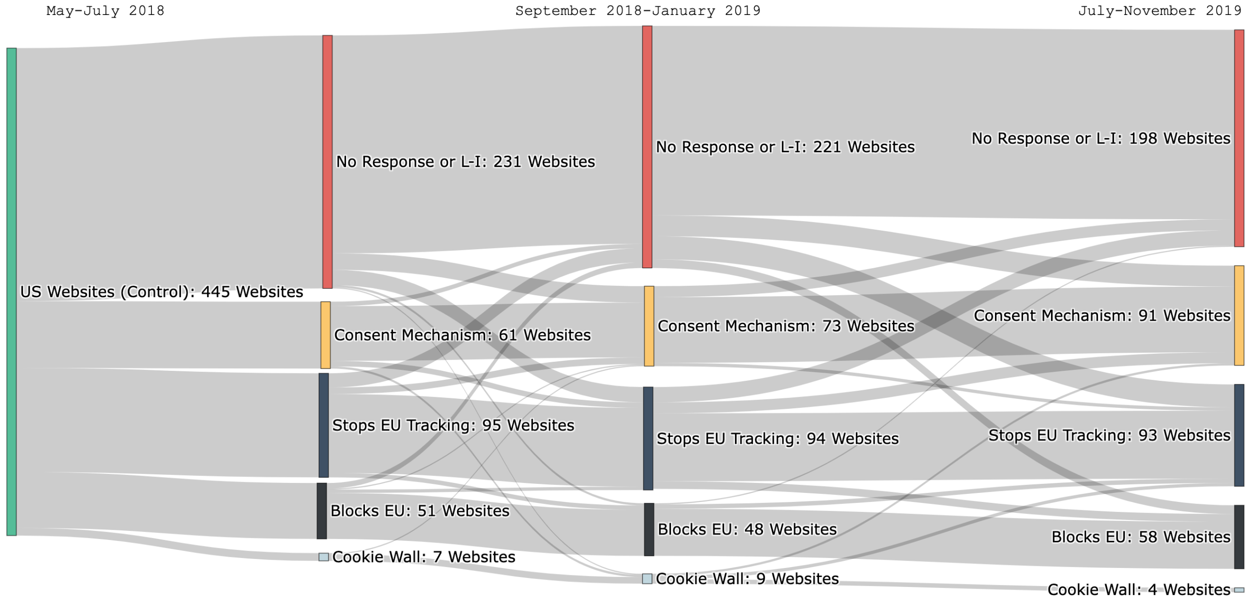


Fig. 5 *Sankey Diagram: Evolution of US Website-Level Responses over Time*



5.2.2 Shares of Traffic and Differences in Response by Country of Origin

Our empirical strategy for detecting differences in ecosystem effects on downstream outcomes for EU-based vs. US-based websites is contingent on two assumptions described in Section 3.1.1: that US content providers in our sample receive significantly lower shares of their traffic from EU visitors than EU content providers; and that US content providers interact differently with their US visitors compared to how they interact with their EU visitors—namely, when visited from US locations, they are less likely to engage in responses that may reduce data collection or revenue generation.

A comparison of the right side of Table 2 to its left side confirms that, unlike EU websites, US websites in our sample, regardless of their response category, receive tiny portions of traffic from the EU: across all US websites, the mean percentage of traffic from the EU is 0.04% (median: 0.02%).

Heterogeneity in website response behavior by country of origin of the visitor for US websites, but not for EU websites, is confirmed by several Figures presented in Section 5.1. No US websites stopped tracking US visitors (but 110 stopped tracking EU visitors; all of them received tiny fractions of their traffic from the EU); only 18.4% percent of the US websites that reduced tracking of EU visitors also reduced tracking for US visitors; and only 12% percent of US websites that implemented consent mechanisms for EU visitors also implemented consent mechanisms for their US visitors.

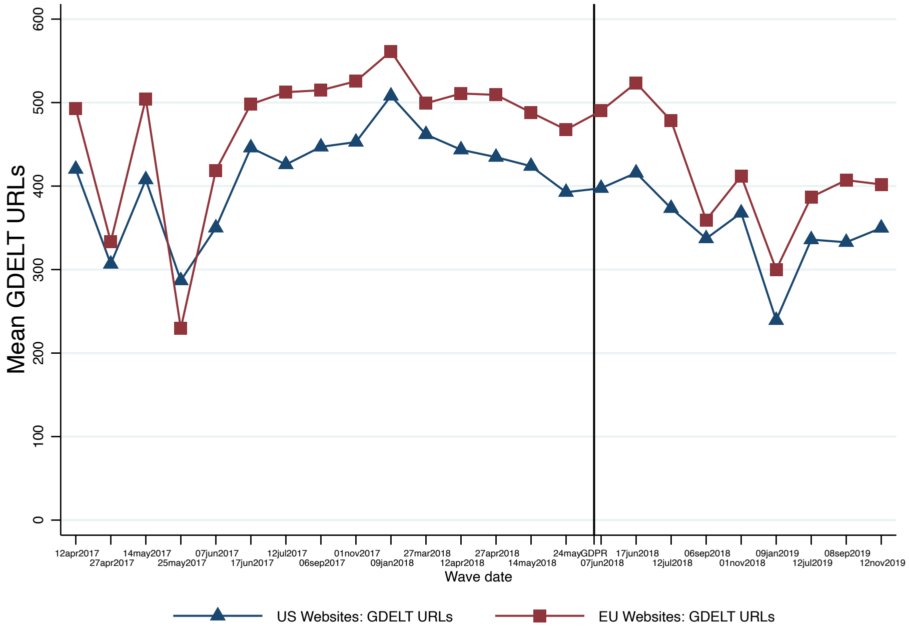
The data confirm the theoretical prediction that US websites respond to GDPR selectively, enacting strategies that may curtail their access to user data only for a minority of their traffic.

5.3 Changes in Downstream Outcomes

5.3.1 Content Quantity

Figure 6 shows similar initial declines in the absolute number of new URLs of content published by both EU and US websites, immediately after the enactment of the GDPR. Considering that the median proportion of EU visitors for US websites is not greater than 2%, and that EU and US websites show analogous trends, we deem the generalized decline to be likely seasonal or due to factors other than the GDPR (such as competition from streaming services). The number of new URLs increases for both EU and US websites a few months following the enactment of the GDPR. The provision of new content seems to follow largely similar trends over time in the two groups of websites.

Fig. 6 *GDELT URLs*



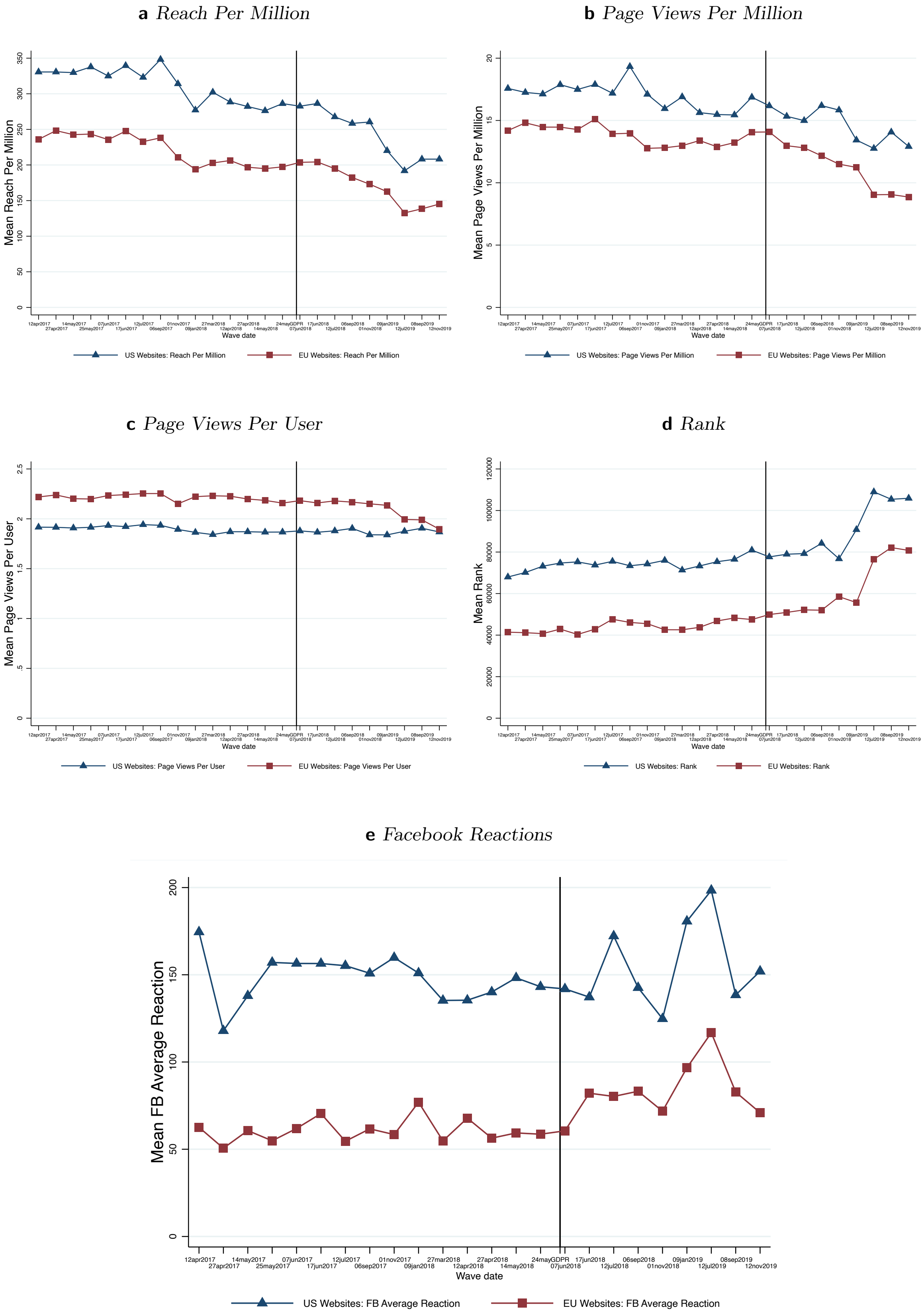
5.3.2 User Engagement

Figure 7 shows that both *Reach* and *Page Views Per Million* exhibit a decline after the GDPR in both EU and US websites. The downward trend is likely unrelated to the GDPR, as it predates its implementation date; a potential cause may be shifts in patterns of news consumption.[20] The reach for EU websites starts increasing towards the end of the period of observation, while page views per million seem to stabilize. These combined patterns could lead to a decrease in page views per user for EU websites, when compared to (the stable pattern for) US websites. Figure 7c is consistent with this prediction. *Page Views Per User* follows similar, stable trends for US and EU websites, but there appears to be a

[20] News is increasingly being consumed through social media and other sources rather than through traditional media channels. According to the 2021 Reuters Institute Digital News Report, only 25% of users go directly to a news and media website when consuming news, with the rest accessing articles though social media, search results, mobile alerts, news aggregators, or e-mail (Newman *et al.*, 2021).

small downward trend for EU websites, relative to US websites, near the end of the period of observation. The decline in reach and page views per million explains the increasing trend in rank (Figure 7d) that we observe for both EU and US websites (an increase in rank number implies the website is getting fewer visits relative to other sites). Figure 7e shows, initially, a stable trend for reactions on Facebook, followed by broadly similar fluctuations in later periods for EU and US websites.

Fig. 7 *User Engagement*



a *Reach Per Million*

b *Page Views Per Million*

c *Page Views Per User*

d *Rank*

e *Facebook Reactions*

5.3.3 Website Survival and Changes in Monetization Strategies

We discuss in this subsection website dynamics that are not covered under our dependent variables and do not represent compliance responses to the GDPR, but which nevertheless capture indirect impacts that the regulation may have had on content providers: website survival rates and changes in monetization strategies.

We use both GDELT and screenshot data to investigate whether the GDPR may have caused interruptions in content production or interruptions of service by content providers. Only a small fraction of websites (around 1%) shut down during the period of observation (their main page URL was no longer accessible, or they had stopped producing content) as of November 2019. In total, 4 websites in the EU and 6 websites in the US shut down during the period of observation or stopped producing content. The difference is not significant.[21]

We use both screenshot and HTML data to investigate changes in monetization strategies, such as subscription options and paywalls. Using screenshot data, we do not find statistically significant differences in the percentage of US and EU websites that, following the enactment of the GDPR, start highlighting payment/subscription options on their homepages. Comparing waves before the GDPR to the waves following it, the fraction of EU and US websites that engage in those activities remains similar (on average, 12% of EU websites and 36% of US websites had subscriptions options before the GDPR was enacted; 12% of EU websites and 41% of US websites had them in the period following the enactment; the changes, within each area, are not statistically significant). Next, we use HTML data and a methodology inspired by Papadopoulos *et al.* (2020) to identify paywalls on websites' front pages during the period of observation. We find a significant increase in paywalls both for US sites (on average, 25% of US websites had paywalls before the GDPR; this percentage went up to 32% in the waves following its enactment; $p < 0.05$) as well as for EU sites (from 14% before the GDPR to 16% after; $p < 0.05$). However, a difference-in-difference analysis shows the increase in paywalls to be *larger* for US sites visited from US IP addresses than for EU sites visited from EU addresses—suggesting that factors other than the GDPR may be at play (see analysis and discussion in Section 7).

[21] The results presented in Section 6 are robust to the exclusion of these websites from the analysis.

6 Empirical Analysis

The descriptive patterns presented so far suggest significant differences in EU and US websites' handling of visitor data following the GDPR, nuanced and complex variations in websites' responses over time, but little evidence of differences in long-term downstream outcomes for EU versus US websites. In this section we estimate in a more exacting way the impact of the GDPR on content providers' downstream outcomes by accounting for both ecosystem and website-level response-driven effects.

6.1 Identification Strategy

All empirical analyses of the GDPR face some common hurdles. First, the GDPR applies to all data subjects in the EU regardless of the location of the data controllers—and therefore affects both EU and US websites, as outlined in previous sections. Second, the mode and intensity of websites' responses should be expected to affect the impact of GDPR's enactment on downstream outcomes. Third, both the decision to respond to the regulation and the decision about how to respond are endogenous decisions of individual websites and are, therefore, correlated to websites' observable and unobservable characteristics.

We use the technical variables mined from individual websites visited from different IP addresses over time to address these identification challenges. We start with a difference-in-differences (DID) approach aimed at estimating the overall impact of the GDPR on downstream outcomes. Based on the theoretical arguments presented in Section 3.1.1, and the empirical validation of the assumptions they rely upon in Section 5.2.2, it is legitimate to expect the GDPR to affect, foremost, EU websites, and only to a lesser extent US websites. Thus we begin by using a definition of treatment and control groups that only considers geographical location, where all the EU websites in our sample are considered as treated and all US websites in our sample are considered as controls. As it is not obvious which, and to what extent, US websites will be affected by GDPR, in our next estimations we use different definitions of treatment and control based on how "exposed" websites are to the regulation. Our definition of exposure is based on the location of the website as well as the location of its visitors. Therefore we repeat the DID analysis by using a definition of treatment and control that considers both the geographical location of the websites and the geographical location of the visitors: we include, among the treated websites, all EU-based websites as well as US websites with a considerable proportion of EU visitors (we use two fairly conservatives thresholds to consider a US website as treated: having at least 10% or at least 5% of visitors from the EU). The effects estimated by these DID models are intention to treat (ITT) effects, since the estimation includes all the websites subject to the treatment assignment. For the great part, they capture what we referred to as the ecosystem effects of the GDPR (Section 3.1.1), notwithstanding the fact that the responses chosen by individual websites still play a role: through the DID estimates, we measure the average impact of the GDPR for websites exposed to the regulation, relative to non-exposed sites, regardless of whether they responded to the GDPR or not, and regardless of the type of response potentially implemented.[22]

Next, we use website response data to estimate website-level effects and take into consideration the fact that not all the websites respond, or respond in the same manner, to the enactment of the GDPR (Section 3.1.2). To account for endogeneity in websites' responses, we use two strategies: instrumental variable and look-ahead matching. First, we use an instrumental variable (IV) approach aimed at estimating a local average treatment effect (LATE)—that is, the effect of the GDPR for those websites that do respond to the regulation in any way (Section 6.3). Second, we take into account that the predominant response over time for EU websites is the adoption of consent mechanisms and attempt to estimate the effect of that specific response, instead of any response as we do in the LATE analysis. We focus on EU websites that decide to adopt a consent mechanism over the period of observation and exploit variation in timing of adoption to utilize a look-ahead matching methodology (Bapna *et al.*, 2018) (Section 6.4).

Finally, we leverage the richness of the data we collected to explore the existence of heterogeneity in the estimated effects over time as well as based on websites' features (Section 6.5).

[22] Recent contributions have highlighted how the DID methodology can produce misleading results when the treatment is staggered or if the treatment effect changes over time (Goodman-Bacon, 2021; De Chaisemartin and D'Haultfoeuille, 2022). Our treatment is not staggered, as GDPR became effective for all affected entities on the same date. We examine heterogeneous treatment effects over time in section 6.5

6.2 Difference-in-Differences or Intention to Treat

We start with a traditional difference-in-differences model to tease out potential changes in content quantity and user engagement after the GDPR, for websites more likely to be exposed to the regulation relative to websites less likely to be exposed to it. Our framework controls for websites' fixed effects and time-specific fixed effects. The specification of our regressions is as follows:

$$Y_{i,t} = \beta_0 + \beta_1 Post\ GDPR \times Exposed\ to\ GDPR_{i,t} + \omega_t + \mu_i + \epsilon_i \tag{1}$$

where $Y_{i,t}$ represents our variable of interest for a website i at wave t; ω_t is a vector of time fixed effects, and μ_i is a vector of website fixed effects. $Post\ GDPR \times Exposed\ to\ GDPR_{i,t}$ is equal to 1 if the website i is exposed to the GDPR and wave t was collected after the GDPR became effective, and 0 otherwise. Standard errors ϵ_i are clustered at the website level. The coefficient β_1 corresponds to the DID estimator of the effect of the implementation of the GDPR for websites exposed to the regulation compared to websites not exposed to it.

For a difference-in-difference estimator to produce unbiased estimations it is necessary that, without the treatment, the treatment and control groups would have followed a similar trend in outcome(s) (the parallel trend assumptions). This assumption cannot be tested, but it is customary to inspect how the treatment and control groups evolved over time before the date of the intervention and assume that if their trends where similar, without the treatment they would have continued to evolve in a similar way. When there is only one treatment and control group, visual inspection of outcome trends is commonly used (Angrist and Pischke, 2009). Figures 6 and 7, in Section 5.3.1, are useful for this purpose. We can see that the pre-GDPR trends for our outcome variables, for EU and US websites, follow very similar patterns.

The results of the DID analysis are presented in Table 3. Our regressions include data from from April 2017 to November 2019. Column (1) presents the results using the log of *GDELT URLs* as the dependent variable. We use a logarithmic transformation to take into account that our dependent variable is a count of new URLs.[23] Columns (2) to (6) present

[23] The results are robust to using a Poisson specification.

the results using *Reach Per Million*, *Page Views Per Million*, *Page Views Per User*, *Rank* and *FB Average Reactions* as dependent variables.

Table 3 is separated into three panels, each presenting the results for the analysis implemented using different definitions of which websites are exposed to the regulation (i.e. treated). The first panel shows the results for our basic specification, where the group exposed to GDPR includes all EU-based websites and the control group includes all US-based websites. In the second panel, the group exposed to GDPR includes EU-based websites and US websites with a share of EU visitors greater than 10%. In the last panel, the group exposed to GDPR includes EU-based websites and US websites with a share of EU visitors greater than 5%.

The results are consistent across the panels. We do not find any significant effect for GDELT URLs (1)—that is, we do not find evidence that the GDPR negatively impacted EU websites' ability to provide new content, relative to their US counterparts. We also do not find evidence of significant changes for *Reach*, *Page Views Per Million*, and *Rank of EU Websites* (Columns 2,3,5). Finally, we do not find a negative effect in terms of social media engagement (Facebook reactions, Column 6). We do find a negative, small, but statistically significant effect for *Page Views Per User* (Column 4): after the enactment of the GDPR, EU websites experience an average decrease in the number of pages browsed in a day by their visitors by about 0.09 pages per user. The result for page views per user seems to be driven by changes that happen at the very end of the period of observation. Looking back at Figure 7c, we observe that the trends for page views per user are stable for both EU and US websites until the very last waves included in our analysis. Towards the end of the period of observation, we see a decline in page views per user for EU websites. One possible interpretation is that the reduction in the number of pages visited may be a signal of reduction in the quality of the content offering: if the quality of the content is reduced, users may decide to spend less time on the website and divert their attention to other websites. This interpretation, however, is not supported by the results of the other engagement variables in the data-set. Another possible interpretation is that the observed trend could be related to the evolution of websites' responses to the GDPR and, in particular, the spreading adoption of consent mechanisms (or even more restrictive responses) by EU websites over time, and the resulting user fatigue from having to interact with consent dialogs across multiple websites. We discuss further possible interpretations of the results in Section 7.

In summary, the results of the DID analysis suggest that the enactment of the GDPR has not greatly affected the outcomes experienced by websites more likely to be exposed to it (EU websites, and US websites with a noticeable fraction of EU visitors). It has not affected the amount of content they are able to publish, or the degree of average social media engagement with such content, but may have negatively affected, to a small degree, their average number of page views per user.

Table 3: DID Estimations

	Content Quantity	User Engagement				
	(1)	(2)	(3)	(4)	(5)	(6)
	Log GDELT URLs	Reach Per Million	Page Views Per Million	Page Views Per User	Rank	FB Average Reaction
Intention to Treat 1 - Exposed to GDPR: EU Websites						
$Post\ GDPR \times Exposed\ to\ GDPR_{i,t}$	0.005	19.002	-0.160	-0.093***	2,746.876	11.994
	(0.041)	(14.080)	(0.881)	(0.033)	(3,322.209)	(15.023)
Constant	5.015***	240.055***	14.494***	2.050***	64,537.281***	108.379***
	(0.007)	(2.695)	(0.169)	(0.006)	(635.879)	(2.668)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	17,577	21,797	21,797	21,797	21,797	17,577
Intention to Treat 2 - Exposed to GDPR: EU Websites + US Websites with more than 10% of EU Visitors						
$Post\ GDPR \times Exposed\ to\ GDPR_{i,t}$	0.004	24.430*	-0.048	-0.119***	2,529.713	13.988
	(0.041)	(13.906)	(0.869)	(0.032)	(3,326.240)	(14.867)
Constant	5.015***	239.107***	14.472***	2.054***	64,588.247***	108.081***
	(0.007)	(2.610)	(0.163)	(0.006)	(624.290)	(2.581)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	17,577	21,797	21,797	21,797	21,797	17,577
Intention to Treat 3 - Exposed to GDPR: EU Websites + US Websites with more than 5% of EU Visitors						
$Post\ GDPR \times Exposed\ to\ GDPR_{i,t}$	0.004	19.512	-0.134	-0.114***	3,172.219	15.511
	(0.041)	(13.969)	(0.873)	(0.032)	(3,323.893)	(14.912)
Constant	5.015***	240.006***	14.488***	2.053***	64,463.728***	107.797***
	(0.007)	(2.639)	(0.165)	(0.006)	(627.967)	(2.607)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	17,577	21,797	21,797	21,797	21,797	17,577

Notes: Standard errors in parentheses and clustered at the website level. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

6.3 LATE

The analysis presented above provides us with an estimate of the effect of the GDPR that puts more emphasis on the overall ecosystem effects, without distinguishing between respondent and non-respondent websites. Additionally, the DID analysis does not consider that some of the US websites may voluntarily extend GDPR stipulations to their US visitors. We use an instrumental variable (IV) approach to estimate the effect of the GDPR for the websites that do respond to the regulation. The effect estimated represents a local average treatment effect (LATE) or the average effect for websites that choose to respond to the GDPR. With this analysis we attempt to better capture the effect of website-level responses (notwithstanding that ecosystem effects will still play a role). Although the decision to respond to the GDPR is endogenous, we follow Angrist and Imbens (1994) and exploit the fact that the (exogenous) enactment of the GDPR can be used as an instrument for the decision of a website to respond to the regulation. To implement this approach we first need to identify whether a website is responding to the GDPR or not. We use a conservative approach and assume that a website is responding to the GDPR if it implements a response that is clearly detectable and able to induce changes in a website's tracking capability. This includes: Stopping the tracking of its visitors, implementing a consent mechanism, or implementing a cookie wall or a cookie banner. Unlike prior work, by using websites' response data to visitors with different IP addresses we can determine whether a website decides to implement the GDPR's requirements for the majority of its visitors, rather than focusing on just EU visitors, as the objective of the LATE analysis is not only to correct the ITT estimates by the fraction of "compliers" in the treated group, but also by the number of "defiers" in the control group, which in our case would correspond to US websites that decide to extend GDPR protections to their US visitors even when they are not required to do so. Note that for the purpose of the LATE analysis our definition would not consider that a US website responds to GDPR if it implements a response only for its EU visitors (and not its US visitors) because, as discussed in section 3.1.1, this response would have negligible downstream effects as the average share of EU visitors at US websites is very small.

In the first stage specification, we estimate the probability of a website responding to the GDPR as function of the enactment of the GDPR and how exposed websites are to the regulation:

$$GDPR\ Response_{i,t} = \alpha_0 + \alpha_1 Post\ GDPR \times Exposed\ to\ GDPR_{i,t} + \omega_t + \mu_i + \zeta_{i,t} \qquad (2)$$

where $GDPR\ Response_{i,t}$ is equal to 1 if the website responded to the GDPR (based on the definition of response provided above), and 0 otherwise.[24] $Post\ GDPR \times Exposed\ to\ GDPR_{i,t}$ is our instrument. $Post\ GDPR$ is equal to 1 for waves collected after the GDPR became effective, and $Exposed\ to\ GDPR_{i,t}$ is equal to 1 if website i at time t is exposed to the regulation and thus should comply with it; ω_t is a vector of time fixed effects, μ_i is a vector of website fixed effects, and $\zeta_{i,t}$ is the error. In the second stage, we regress the outcomes of interest (the log of *GDELT URLs*, *Page Views Per User*, *Reach Per Million*, *Rank*, and *FB Average Reactions*) onto the predicted GDPR response and the time and website level fixed effects. The specification of the second stage is:

$$Y_{i,t} = \beta_0 + \beta_1 \widehat{GDPR\ Response}_{i,t} + \omega_t + \mu_i + \epsilon_{i,t} \qquad (3)$$

where $Y_{i,t}$ represents the outcome variable of interest for a website i at wave t; $\widehat{GDPR\ Response}_{i,t}$ is the predicted response from the first stage, ω_t is a vector of time fixed effects, μ_i is a vector of website fixed effects, and $\epsilon_{i,t}$ corresponds to the error term.

In our LATE estimations, we use data from from April 2017 to November 2019. Note that the specifications outlined above consider that sites may respond to GDPR even before the regulation becomes effective to account for anticipation effects (although previous studies on the GDPR, such as Peukert *et al.* (2020), did not find any). Table 4 presents the results of the IV approach. As in the DID analysis, we use three possible definitions of websites exposed to the GDPR (see Section 6.2). Results are consistent across the different panels: for all our outcomes of interest, we do not find a statistically significant effect of the GDPR for websites that do choose to respond to the regulation, for the overwhelmingly majority of our outcome variables. The only exception is, again, page views per users, which shows a negative and statistically significant coefficient.

[24] We only have accurate GDPR response data starting in April 2018, so we set the $GDPR\ Response_{i,t}$ to 0 before April 2018.

Table 4: LATE Estimations

	Log GDELT URLs		Reach Per Million		Page Views Per Million		Page views per user		Rank		FB Average Reaction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	GDPR Response	Log GDELT URLs	GDPR Response	Reach Per Million	GDPR Response	Page Views Per Million	GDPR Response	Page Views Per User	GDPR Response	Rank	GDPR Response	FB Average Reaction
LATE 1 - Exposed to GDPR: EU Websites												
Post GDPR × *Exposed to* $GDPR_{i,t}$	0.487***		0.474***		0.474***		0.474***		0.474***		0.487***	
	(0.020)		(0.017)		(0.017)		(0.017)		(0.017)		(0.020)	
GDPR Response		0.010		40.095		-0.338		-0.197***		5,796.157		24.652
		(0.084)		(29.681)		(1.858)		(0.069)		(6,982.403)		(30.828)
Constant	0.001		0.000		0.000		0.000		0.000		0.001	
	(0.006)		(0.006)		(0.006)		(0.006)		(0.006)		(0.006)	
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Share of response inside Control		0.184		0.184		0.184		0.184		0.184		0.184
Share of response inside Treatment		0.799		0.799		0.799		0.799		0.799		0.799
Underidentification (LM)		333.277		413.138		413.138		413.138		413.138		333.277
P-value (LM-Stat)		0.000		0.000		0.000		0.000		0.000		0.000
Weak identification		614.166		757.476		757.476		757.476		757.476		614.166
P-value (J-Stat)		0.000		0.000		0.000		0.000		0.000		0.000
Obs.	17,588	17,577	21,797	21,797	21,797	21,797	21,797	21,797	21,797	21,797	17,588	17,577
LATE 2 - Exposed to GDPR: EU Websites + Websites with more than 10% of EU Visitors												
Post GDPR × *Exposed to* $GDPR_{i,t}$	0.488***		0.475***		0.475***		0.475***		0.475***		0.488***	
	(0.020)		(0.017)		(0.017)		(0.017)		(0.017)		(0.020)	
GDPR Response		0.008		51.449*		-0.102		-0.251***		5,327.610		28.693
		(0.084)		(29.282)		(1.829)		(0.069)		(6,978.311)		(30.439)
Constant	0.001		0.000		0.000		0.000		0.000		0.001	
	(0.006)		(0.006)		(0.006)		(0.006)		(0.006)		(0.006)	
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Share of response inside Control		0.187		0.187		0.187		0.187		0.187		0.187
Share of response inside Treatment		0.804		0.804		0.804		0.804		0.804		0.804
Underidentification (LM)		334.423		416.150		416.150		416.150		416.150		334.423
P-value (LM-Stat)		0.000		0.000		0.000		0.000		0.000		0.000
Weak identification		620.947		766.516		766.516		766.516		766.516		620.947
P-value (J-Stat)		0.000		0.000		0.000		0.000		0.000		0.000
Obs.	17,588	17,577	21,797	21,797	21,797	21,797	21,797	21,797	21,797	21,797	17,588	17,577
LATE 3 - Exposed to GDPR: EU Websites + Websites with more than 5% of EU Visitors												
Post GDPR × *Exposed to* $GDPR_{i,t}$	0.489***		0.477***		0.477***		0.477***		0.477***		0.489***	
	(0.019)		(0.017)		(0.017)		(0.017)		(0.017)		(0.019)	
GDPR Response		0.008		40.907		-0.280		-0.239***		6,650.510		31.722
		(0.084)		(29.251)		(1.829)		(0.069)		(6,939.220)		(30.443)
Constant	0.001		0.000		0.000		0.000		0.000		0.001	
	(0.006)		(0.006)		(0.006)		(0.006)		(0.006)		(0.006)	
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Share of response inside Control		0.184		0.184		0.184		0.184		0.184		0.184
Share of response inside Treatment		0.804		0.804		0.804		0.804		0.804		0.804
Underidentification (LM)		337.446		420.428		420.428		420.428		420.428		337.446
P-value (LM-Stat)		0.000		0.000		0.000		0.000		0.000		0.000
Weak identification		629.437		781.364		781.364		781.364		781.364		629.437
P-value (J-Stat)		0.000		0.000		0.000		0.000		0.000		0.000
Obs.	17,588	17,577	21,797	21,797	21,797	21,797	21,797	21,797	21,797	21,797	17,588	17,577

Notes: Standard errors in parentheses and clustered at the website level. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

6.4 Look-Ahead Matching

In the analyses presented in the previous subsections, the estimates capture ecosystem- and website-level response effects to different extents, with the DID analysis giving more emphasis to ecosystem effects, and the LATE analysis focusing more on website-level effects. In this section we attempt to isolate the effect of website-level responses. We focus on the implementation of consent mechanisms, as this was the predominant response to GDPR and, in terms of sample size, should be easiest way to observe a statistically significant effect.

The response websites implement is likely correlated with their characteristics, many of which may not be observable. This endogeneity problem prevents us from directly comparing websites that adopt a particular response versus websites that do not. We address this challenge using look-ahead matching (Bapna *et al.*, 2018) strategy. We compare websites that have adopted a response with websites that have not adopted such response but will adopt it some time in the future (or that adopted the response and later abandoned it). This approach isolates the analysis from the endogeneity problem, as we only consider websites that will end up adopting a response and exploit the temporal variation in adoption to identify the impact of the response on our variables of interest.

Of the 465 EU websites in our sample, 316 used a consent mechanism in at least one of the waves. Considering only the subsample of EU websites that use a consent mechanism for at least one wave, and using only observations after the GDPR was implemented, we estimate the following linear regressions:

$$Y_{i,t} = \beta_0 + \beta_1 \times Response_{i,t} + \omega_t + \mu_i + \epsilon_i \tag{4}$$

In this equation, $Y_{i,t}$ corresponds to the outcomes we study for website i at time t; Response is equal to 1 if website i has adopted the response of interest for EU visitors at time t; ω_t is a vector of time fixed effects, and μ_i is a vector of website fixed effects. Standard errors are clustered at the website level. In this estimation, β_1 corresponds to the effect of the website-level response on our outcome variables of interest.

Table 5: Look-Ahead Matching - Consent Mechanism

	Log GDELT URLs	Reach Per Million	Page Views Per Million	Page Views Per User	Rank	FB Average Reaction
Consent Mechanism	0.0347	-6.275	-0.639	0.00734	-5,741	-3.467
	(0.0485)	(6.410)	(0.537)	(0.0281)	-3,981	(8.513)
Constant	5.253***	244.3***	16.43***	2.086***	47,395***	64.53***
	(0.0441)	(7.149)	(0.575)	(0.0224)	-3,206	(12.25)
Website Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,216	2,838	2,838	2,838	2,838	2,247
R-squared	0.887	0.975	0.978	0.710	0.708	0.633

Notes: Standard errors in parentheses and clustered at the website level. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

Table 5 shows the Look-Ahead Matching estimation of the website-level effect of consent mechanisms. We do not find any statistically significant effect on any of the outcomes we study. This contrasts with the results of the DID and LATE analysis, where we observed a small but statistically significant effect on page views per user. The difference is probably due to the look-ahead matching attempting to identify an effect, which was already small, by relying solely on differences that happen within EU websites that adopt the same response in different points in time after GDPR. This result suggests that the use of consent mechanisms is not the only reason behind the negative effects of GDPR on page views per user, and instead the effect is due to a combination of ecosystem and website level effects.

Table 6: Look-Ahead Matching - Cookie Wall

	Log GDELT URLs	Reach Per Million	Page Views Per Million	Page Views Per User	Rank	FB Average Reaction
Cookie Wall	0.148	-0.149*	-0.790	6,671	-0.569	-10.15
	(0.203)	(0.0801)	(6.366)	(6,492)	(0.734)	(10.87)
Constant	4.324***	2.503***	55.83***	81,438***	4.501***	37.21**
	(0.116)	(0.0677)	(8.389)	(9,111)	(1.101)	(15.75)
Website Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	243	333	333	333	333	243
R-squared	0.870	0.770	0.934	0.778	0.852	0.643

Notes: Standard errors in parentheses and clustered at the website level. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

Given our inability to find a website-level effect associated with the implementation of consent mechanisms, it is interesting to explore whether there is some other website-level response that would by itself (that is, without considering ecosystem effects) negatively impact downstream outcomes. Considering the case of page views per user (the only outcome for which we found a statistically significant impact in our previous estimations), the responses that could arguably impact it the most are blocking EU visitors or implementing a cookie wall. Capturing the effect of blocking EU users is uninteresting, as we already reported that this response was only adopted by US websites with minimal EU audiences. Exploring the website-level impact of using cookie walls is more insightful, as it is a response that was used by some EU websites. This estimation is shown in Table 6. We do find a negative impact of the use use cookie walls on page views. Conceivably, this explains why cookie walls were only used by a small number of websites, and frequently abandoned by the websites that used them. In total, 37 EU websites in our sample use a cookie wall in at least one wave. The maximum number of EU websites using cookie walls at the same time happens a few weeks after GDPR (with 30 websites using them). In our last wave of data collection, only 22 websites were using cookie walls. The limited and declining popularity of cookie walls contrasts with the increasing use of consent mechanism. Our estimation of website level effects provides a possible reason behind this trend, as it suggests that adopting cookie walls directly impacts the outcomes experienced by websites.

6.5 Heterogeneous effects

Our findings so far suggest the GDPR had no impact on EU websites' ability to provide content, relative to their US counterparts, or on traffic and engagement measures, with the exception of a negative, albeit small, effect on page views per user. In this section we report on additional analyses we conducted to account for heterogeneity in websites' reliance on advertising and heterogeneity in ranking, and to look at differences in short- versus long-run changes in downstream outcomes.

We repeat our difference-in-difference analysis for the sub-sample of websites that rely more heavily on advertising to monetize their content before the GDPR. We separate our sample in two groups—"low" and "high" advertising—respectively representing websites below and above the median advertising intensity before the GDPR. Table 8 in the Appendix shows that the negative effect we found on page views per user is similar for the low and the high advertising group. As in previous results, we find no statistically significant changes in the quantity of new content published or other measures of user engagement, including average Facebook reactions.

Next, we investigate the effect of the GDPR on downstream outcomes for the top ranked and bottom ranked websites. We split the sample into two groups: the first group consists of websites ranking in the top 10% of websites in their respective region (EU/US) with respect to websites in our sample (92 websites); the second group includes the bottom 10% of websites in their respective region (EU/US) (91 websites). Table 9 presents the DID estimation of the effect the GDPR on content for the top ranking and bottom ranking EU websites. Columns (1), (3), (5), and (7) report the estimations for the sub-sample of the top ranking websites. The results confirm no effect of the GDPR on GDELT URLs, ranking, and FB reactions in both the top and bottom ranked websites samples, and a negative effect on page views per user.

As observed in Section 5, both EU and US websites reacted rapidly to the enactment of the GDPR by reducing the magnitude of visitor tracking, but such reduction only lasted a few months. In fact, responses to the GDPR (including intensity of visitor tracking) evolved over time for a majority of EU websites. It stands to reason that downstream outcomes of the regulation may change across our period of observation. To compare short- and long-run effects of the GDPR, we split our sample into two groups. The short-run subsample includes all pre-GDPR waves (from April 2017 to May 25, 2018) as well as early post-GDPR waves up to January 2019. In Table 10 in the Appendix, columns (1), (3), (5), (7), (9), and (11) present the results for the estimation on the short-run subsample. The long-run subsample, instead, *excludes* from the analysis the time period just after the GDPR (the period from June 2018 to January 2019), and instead includes, in addition to all the pre-GDPR waves, only the latest post-GDPR waves (from July 2019 to November 2019). Columns (2), (4), (6), (8), (10), and (12) present the estimation for the long-run analysis subsample. For the most part, results do not change from our previous analysis. We find little evidence of an impact of the GDPR on websites' ability to provide content or on visitors' engagement. Columns (7) and (8) confirm that the decrease in page views per user for EU websites compared to US websites arises only in the long-run.

7 Discussion

The scant evidence for ecosystem or website-level effects of the GDPR on various measures of content quantity and user engagement is a surprising result of our analysis. The ability of EU-based outlets to produce content and engage audiences does not seem to have been substantially affected by the regulation; their ranking, relative to US websites, does not seem to have changed. Furthermore, EU content providers do not appear to exit the market at higher rates than US counterparts, or start highlighting subscription options on their homepages, or switch to paywalls at higher frequencies. In short, our results suggest that, by and large, the GDPR, one and a half year after its enactment, had not impaired content providers' downstream outcomes. Considering the pre-GDPR expectations on these matters (Section 1), as well as theoretical arguments supporting the hypothesis of a negative effect (Section 3), these findings call for explanation. In this section we consider a number of possible alternative mechanisms that may have produced them.

One possible explanation for the lack of starker downstream effects on content provision is that revenues from ads did in fact decrease following the GDPR, but websites' aggregate revenues did not (and hence quantity and quality of content did not vary), because affected EU websites switched to other sources of revenue/business models. As noted in Section 5, we do not find evidence supporting this explanation in our data-set. First, only a small proportion of EU websites decided to implement cookie walls; this number increased from 2.3% to 4.1% during the period of observation, but by the end of that period ended up reverting to levels nearly identical to pre-GPDR levels (roughly 2.5%). Thus, while a few more websites implemented cookie walls following the enactment of the GDPR, we do not find an increase in the number of EU websites permanently switching to cookie walls. (Our look ahead matching estimation provides a rationale for this pattern, as we found that websites that implement cookie walls performed worst in terms of page views per user during the periods they were using them compared to periods they were not using them.) Second, as noted in Section 5.3.3, we do not observe a significant increase in the number of EU or US websites showing subscription options on their front page during the period of observation. Third, we do not find evidence that the usage of paywalls increased among EU websites more than among US websites. Figure 20 in the Appendix is insightful, as it leverages data on differential changes in website behavior by visitor's location. The figure shows that while EU sites added paywalls for both EU and US visitors, US sites did so mainly for US visitors—that is, visitors *not* covered by the GDPR. In fact, while the number of paywalls identified on websites grew during the period of observation within both EU sites and US sites (see Section 5.3.3), a difference-in-differences regression shows the change to be actually *larger* for US sites.[25] Both pieces of evidence suggest that the generalized increase in paywalls was not GDPR-related, and that factors other than the GDPR—such as the progressive decline of the news industry's financial fortunes and subscriber base (Pew Research Center, 2021b), and the competition for audience online newspapers face from alternative online channels, including social media (Pew Research Center, 2021a)—may be potentially at play.

Another possible explanation for the lack of starker downstream effects on content

[25] Results available from the authors upon request.

provision is that EU websites (particularly those that responded to the GDPR in more forceful manners, such as curtailing tracking) attempted to compensate for revenue losses due to reduced tracking by increasing ad intensity (the volume of ads displayed on their pages).We do not find evidence supporting this mechanism, either. As Figure 15 in the Appendix shows, the few EU websites (about 40) that decreased tracking (orange line) did not experience a systematic change in advertising intensity. Ad intensity decreased somewhat after the enactment of the GDPR, and picked up again soon after. Even EU websites and US websites that kept tracking constant following the GDPR (red line and blue line, respectively) exhibit relatively stable patterns of ad intensity (the peak observed for both EU and US websites is associated with the Christmas shopping period). If anything, ad intensity for US websites that chose to decrease tracking (green line) seems to decrease over time, suggesting that a reduction in tracking is correlated with less advertisement, not more. We further confirm the results presented above by examining ad intensity on EU websites by type of response to the GDPR. While ad intensity fluctuates over time (with, again, some decrease after GDPR enactment, followed by an uptick around the Christmas season), by the end of our period of observation ad intensity among websites with the more common reactions (adopting a consent mechanism or not responding to the GDPR) is close to where it was before the GDPR (see Figure 12 in the Appendix).

Another possible explanation for the lack of starker downstream effects on content provision is that advertising revenues did not substantially change for EU websites because the amount of data available for targeting in the EU ecosystem, or advertisers' ability to target adverting content, ultimately did not vary to a significant enough degree. This may have occurred through dynamic adaptations to the regulation at both the website and ecosystem levels.

First, over time, publishers may have evolved and adapted responses and compliance postures in manners strategically designed not to hurt them (the Sankey diagram we presented in Section 5.2.1 provides evidence of EU websites varying their response strategies to the GDPR over time: Figure 4). In particular, a number of third-party players entered the market several months following the enactment of the GDPR and helped publishers manage compliance requirements (rather than altogether curtail tracking). In turn, publishers felt pressure to adopt these platforms, as ad buyers placed higher value on inventory with information on user consent (Davies, 2018). Industry reports suggest that the growing popularity of consent mechanisms among EU websites was driven by the rise of intermediary Consent Management Platforms (CMPs) such as OneTrust, Quantcast, or Trustarc that help publishers collect and communicate consent (Davies, 2018). These reports are supported by empirical measurements that track a rise in these platforms following GDPR enforcement (Hils *et al.*, 2020). These reports are consistent with our data. Sankey diagrams of EU websites (Figure 4 in Section 5) indicate a more dynamic reaction by EU websites over time, compared to their US counterparts, and Figure 16 in the Appendix confirms that the number of EU websites in our sample that introduce a consent mechanism kept increasing overtime.

Although consent mechanisms may, in theory, reduce the amount of data available to the publishers (by allowing consumers to opt-out), multiple studies have reported the emergence of dark patterns in GDPR consent dialogs to nudge visitors towards acquiescence to tracking (see (Nouwens *et al.*, 2020) and Section 2). We know from other contexts that, when tracking choices are made easily accessible to users, or no tracking is the default, few users autonomously choose to be tracked (Godinho de Matos and Adjerid, 2022). In the case of the GDPR, few websites made opt-out choices easily accessible to visitors. Using our data, we can differentiate between consent mechanisms (CMs) that require a single action for users to reject tracking ("Single-Step CM") from those that require more than one step ("Multi-Step CM"). While we observe that both increase over time in our sample following the enforcement of the GDPR, multi-step CMs are much more prevalent and are adopted at a faster rate [26]. These CMs are arguably more likely to dissuade visitors from actually completing the process of opting out of tracking. Additionally, in Section 6.4, when we attempt to isolate the website-level effect of the presence of a consent mechanism on outcomes, we do not find any significant effect.

The increasing presence of multi-step CMs among EU websites may also be one of the explanations for the negative effects of the GDPR on page views per user that we have documented in prior sections. Rather than originating from the GDPR directly, the slight reduction in page views per users may be an unexpected effect of the "fatigue" that CMs

[26] See Figure 18 and Figure 19 in the Appendix.

impose on visitors (Ursu *et al.*, 2022). Almost 60% of websites in our sample use consent mechanisms at the end of our data collection. CMs capture the attention of visitors on each websites, which can cumulatively reduce their ability to view pages per websites. For example, Yan *et al.* (2022) find that the use of Ad-Blockers increase the amount of news consumption by keeping user attention. We offer that CMs can induce the opposite effect. This means that CMs can have a negative effect on all websites, not just those that implement them. Consistent with the fact that the number of EU websites using CMs kept increasing over time, the empirical analysis in Section 6.5) reports a reduction in page views per user materializing only in the long-run. Additionally, our look-ahead analysis suggests that this reduction cannot be attributed only to the presence or absence of a CM: when comparing within sites that use consent mechanisms at different points in time, we do not find any statistically significant difference in terms of page views per user (or any other outcome variable) between the periods when a CM is in use and those when it is not.

Second, the market of third-party players (especially tracking firms) in the EU data ecosystem may too have adapted and evolved over time. Prior work has documented an early concentration effect of GDPR on data markets (see Johnson and Shriver (2019); Peukert *et al.* (2020) and Section 2). In Section 5, we presented descriptive results *initially* compatible with prior work: EU-website-level tracking (as measured by the number of third party-trackers) decreased, initially, following the enactment of the GDPR (Figure 1). However, we also presented additional evidence suggesting a more nuanced evolution of this market in the longer-run: Figure 1 shows how the number of third-party cookies on EU websites actually picked up again several months after the enactment of the GDPR.[27] Most importantly, the number of *unique* third-party tracking companies in the market also evolved over time. Figure 17 in the Appendix shows a surprising result: the distribution of the number of different third-party companies over time tracks precisely the graph of the number of third-party cookies we presented in previous sections. In other words, some third-party companies left the market shortly after GDPR, making it more concentrated (consistent with early prior work), yet

[27] Figure 13 in the Appendix confirms that the websites that adopted the most common response (either adopting a consent mechanism or not responding/invoking legitimate interest) showed exactly the same patterns we have presented in prior sections, with the number of third-party trackers first decreasing and then increasing back to pre-GDPR levels.

they did so temporarily, as in the long-term they re-enter the market. A possible effect of these dynamics (EU websites increasing the number of third-party cookies over time, and the number of third-party trackers expanding in the long-run after a post-GDPR shrinkage) is that the amount of visitor data available to EU websites and in the EU ecosystem may have stabilized in the long-run following the enactment of the GDPR.

Third, even when regulations such as the GDPR may affect the availability of cross-session and cross-device tracking data, online publishers may still have other ways to target individuals with valuable advertising. For example, a website may infer a visitors' preference, interests, and income with information such as the location (IP) of the user, their operating system and browser, and, most notable, the type of content they are browsing. By contextually targeting ads (Zhang and Katona, 2012) from these instantaneous data, publishers may partially offset the loss of targeting precision (and revenue) associated with a decrease in tracking across the ecosystem.

In summary, we considered possible explanations for the lack of more pronounced negative downstream effects on websites' content. We lean towards ruling out as legitimate dynamics an increase (in EU websites, relative to US ones) in subscription-based revenue models that do not rely on tracking or an increase in advertising intensity. On the other hand, we were not able to rule out the possibility that EU websites (and the EU data ecosystem as a whole), after an initial decrease in tracking, over time reached levels of tracking comparable to pre-GDPR levels, or adopted data gathering responses and compliance postures in manners strategically designed not to hurt them.

8 Limitations

Despite our data-sets covering a period that extends for nearly two years into the GDPR, we acknowledge that it may still be too early to detect changes in the content produced by publishers. Firms, weighing the cost of compliance against potential fines that may result from enforcement actions, may be inclined to wait until EU authorities provide further clarification on the requirements for compliance. Others still may be justifying data collection and processing under the legitimate interest clause of Article 6. Indeed, a December 2019 report by the Dutch Data Protection Authority found that many popular websites were still placing tracking cookies on the browsers of EU visitors (Autoriteit Persoonsgegevens, 2019a). If a significant number of websites are currently not fully compliant with the GDPR requirements, this would make the impact of the regulation on publishers' content weaker and thus more difficult to detect. It is possible that future clarifications or enforcement actions by the EU will trigger smaller scale market shocks as publishers are steered towards compliance in areas such as consent.

While we used multiple measures to capture content quantity and quality, they are only proxies that may not fully capture the potential effect of the GDPR. Additionally, although we classified cookies and HTTP requests to identify tracking and advertising related activity, and devised a way to detect the presence of consent mechanisms, our technical variables are capturing only a part of the technical changes that are possible.

Additionally, the GDPR may have impacted the ability of Alexa to collect traffic data, which in turn may spuriously affect some of our downstream metrics. However, the findings we obtain from Alexa data are consistent with those we obtain for other sources, such as GDELT and FB API, as well as the screenshots and HTML data we mined directly from the websites in our sample. This may be because the impact of the GDPR on Alexa's data collection practices was, in fact, likely more limited compared to website traffic analytics services that rely primarily on third-party cookies and tracking pixels placed on websites, whereas Alexa's data collection relies on multiple sources.[28]

9 Conclusions

The enactment of the GDPR was accompanied by concerns over possible unintended economic consequences—in particular, potential detrimental effects on websites' ability to produce quality, free content. We assess the impact of the GDPR on ad-supported content providers by tracking downstream effects of the regulation. Whereas previous work focused on measur-

[28] In their product literature, Alexa claims to gather traffic data from multiple sources, many of which do not depend on cookies, such as users of its Alexa Toolbar and over 25,000 other browser extensions Yesbeck (2016). If we assume a similar diversity in responses to the GDPR among these browser extensions as there was among websites (including browser extensions that choose not to respond to the GDPR at all), we may expect any effect of the GDPR on Alexa traffic data to be blunted.

ing the effects of the GDPR on advertising technologies (such as cookies) and short-term effects, we captured the evolution over time of a number of metrics related to tracking, traffic, and content variables over several months, both leading up to and immediately following the enforcement of the GDPR.

The GDPR initially reduced the number of third-party cookies and tracking responses, suggesting decreased tracking of users by websites. This decrease is more evident for EU visitors to US websites, indicating that US websites are taking a conservative approach when dealing with the requirements of the GDPR. The short-term reduction in tracking among EU and US sites, was followed, for EU websites, by an uptick in tracking several months after the enactment of the regulation. We do not find evidence of EU content providers exiting the market at higher rates than US counterparts or switching to alternative revenue models (e.g., cookie-walls or paywalls) with higher frequency.

We use multiple identification strategies including DID estimations, LATE models, and a look-ahead matching to estimate ecosystem and website-level effects of the regulation. We do not find any statistically significant impact of the regulation on EU websites' ability to provide content. While we find a small reduction in the average number of page views per user in EU websites relative to US websites, we find no statistically significant impact on other measures of visitor engagement, including the amount of visitors' traffic EU websites receive, their rank, or on visitors' social media reactions to new content. The robustness of this result was confirmed by using different methodologies to account for endogeneity concerns, and by the absence of significant differences in content providers' survival in the EU vs. the US.

In short, while industry predictions forebode dire consequences of the GDPR for content providers, our data collected for a period of almost two years suggest this did not materialize. Our analyses indicate this is most likely due to the fact that websites that did respond more strongly to the GDPR were those not likely, in fact, to be affected by such a response. In contrast, websites that did rely to a larger extent on EU visitors found, over time, ways to avoid being negatively affected by the regulation.

References

Acquisti, A., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L. F., Komanduri, S., Leon, P. G., Sadeh, N., Schaub, F., Sleeper, M. *et al.* (2017). Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Computing Surveys (CSUR).* 50(3), 1–41.

Acquisti, A., Taylor, C. and Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature.* 54(2), 442–92.

Adjerid, I., Acquisti, A., Telang, R., Padman, R. and Adler-Milstein, J. (2015). The Impact of Privacy Regulation and Technology Incentives: The Case of Health Information Exchanges. *Management Science.* 62(4), 1042–1063.

Anderson, S. and Gabszewicz, J. (2006). The Media and Advertising: A Tale of Two-Sided Markets. In *Handbook of the Economics of Art and Culture.* (p. 568–614). vol. 1. Elsevier B.V., Amsterdam, chap. 18. (1st ed.).

Angelucci, C. and Cagé, J. (2019). Newspapers in Times of Low Advertising Revenues. *American Economic Journal: Microeconomics.* 11(3), 319–64.

Angrist, J. and Imbens, G. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica.* 62(2), 467–475.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Aridor, G., Che, Y.-K., Nelson, W. and Salz, T. (2020). The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR. *NBER Working Paper.* (w26900).

Athey, S., Calvano, E. and Gans, J. S. (2018). The Impact of Consumer Multi-Homing on Advertising Markets and Media Competition. *Management Science.* 64(4), 1574–1590.

Autoriteit Persoonsgegevens (2019a). *AP: Veel Websites Vragen Pp Onjuiste Wijze Toestemming Voor Plaatsen Tracking Cookies.* Technical report. Autoriteit Persoonsgegevens. Retrievable at `https://autoriteitpersoonsgegevens.nl/nl/nieuws/ap-veel-websites-vragen-op-onjuiste-wijze-toestemming-voor-plaatsen-tracking-cookies`.

Autoriteit Persoonsgegevens (2019b). *Websites Moeten Toegankelijk Blijven Bij Weigeren Tracking Cookies.* Technical report. Autoriteit Persoonsgegevens. Retrievable at `https://autoriteitpersoonsgegevens.nl/nl/nieuws/websites-moeten-toegankelijk-blijven-bij-weigeren-tracking-cookies`.

Bapna, R., Ramaprasad, J. and Umyarov, A. (2018). Monetizing Freemium Communities: Does Paying for Premium Increase Social Engagement? *MIS Quarterly.* 42(3), 719–736.

Beales, H. (2010). The Value of Behavioral Targeting. *Network Advertising Initiative.* 1, 2010.

Cagé, J., Hervé, N. and Viaud, M.-L. (2020). The Production of Information in an Online World. *The Review of Economic Studies.* 87(5), 2126–2164.

Casadesus-Masanell, R. and Zhu, F. (2013). Business Model Innovation and Competitive Imitation: The Case of Sponsor-Based Business Models. *Strategic Management Journal.* 34(4), 464–482.

Castro, D. (2010). *Stricter Privacy Regulations for Online Advertising Will Harm the Free Internet.* Technical report. Information Technology and Innovation Foundation. Retrievable at https://itif.org/publications/2010/09/08/stricter-privacy-regulations-online-advertising-will-harm-free-internet#:~:text=Stricter%20Privacy%20Regulations%20for%20Online%20Advertising%20Will%20Harm%20the%20Free%20Internet,-Daniel%20Castro%20September&text=A%20study%20shows%20that%20overly,effectiveness%20of%20the%20Internet%20ecosystem.

Choi, J. P., Jeon, D.-S. and Kim, B.-C. (2019). Privacy and Personal Data Collection with Information Externalities. *Journal of Public Economics.* 173, 113–124.

Competition and Markets Authority (CMA) (2020). *Online Platforms and Digital Advertising.* Technical report. Market study final report. Retrievable at https://assets.publishing.service.gov.uk/media/5fa557668fa8f5788db46efc/Final_report_Digital_ALT_TEXT.pdf.

Congiu, R., Sabatino, L. and Sapi, G. (2022). *The Impact of Privacy Regulation on Web Traffic: Evidence From the GDPR. Working Paper.*

Cook, C. and Sirkkunen, E. (2013). What's in a Niche? Exploring the Business Model of Online Journalism. *Journal of Media Business Studies.* 10, 63–82. ISSN 1652-2354. doi: 10.1080/16522354.2013.11073576.

Dabrowski, A., Merzdovnik, G., Ullrich, J., Sendera, G. and Weippl, E. (2019). Measuring Cookies and Web Privacy in a Post-GDPR World. In *International Conference on Passive and Active Network Measurement.* Springer, 258–270.

Davies, J. (2018). *Under GDPR, Publishers Are Adopting CMPs for Fear of Losing Out on Ad Revenue.* Retrievable at https://digiday.com/media/gdpr-publishers-adopting-cmps-fear-losing-ad-revenue/, accessed: 2021-13-08.

Davies, J. (2019). *After GDPR, The New York Times Cut Off Ad Exchanges in Europe — and Kept Growing Ad Revenue.* Retrievable at https://digiday.com/media/gumgumtest-new-york-times-gdpr-cut-off-ad-exchanges-europe-ad-revenue/, accessed: 2021-05-08.

De Chaisemartin, C. and D'Haultfoeuille, X. (2022). Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey. *NBER Working Paper.* (w29691).

Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F. and Holz, T. (2019). We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. In *Network and Distributed Systems Security (NDSS) Symposium 2019.* 345–346.

Deloitte (2013). *Economic Impact Assessment of the Proposed General Data Protection Regulation.* Technical report. Deloitte. Retrievable at https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/about-deloitte/deloitte-uk-european-data-protection-tmt.pdf.

Dorfleitner, G., Hornuf, L. and Kreppmeier, J. (2021). *Promise Not Fulfilled: Fintech, Data Privacy, and the GDPR. Working Paper.*

Downes, L. (2018). GDPR and the End of the Internet's Grand Bargain. *Harvard Business Review.* Retrievable at `https://hbr.org/2018/04/gdpr-and-the-end-of-the-internets-grand-bargain`.

Englehardt, S. and Narayanan, A. (2016). Online tracking: A 1-Million-Site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security.* ACM, 1388–1401.

Evans, D. S. (2009). The Online Advertising Industry: Economics, Evolution, and Privacy. *Journal of Economic Perspectives.* 23(3), 37–60.

Ferreira, L. N., Hong, I., Rutherford, A. and Cebrian, M. (2021). The Small-World Network of Global Protests. *Scientific Reports.* 11.

Flynn, K. (2018). *What Are Facebook's First-Party Cookies for Pixel?* Retrievable at `https://digiday.com/marketing/wtf-what-are-facebooks-first-party-cookies-pixel/`.

Gallea, Q. and Rohner, D. (2021). Globalization Mitigates the Risk of Conflict Caused by Strategic Territory. *Proceedings of the National Academy of Sciences.* 118(39), e2105624118.

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

Goldberg, S., Johnson, G. and Shriver, S. (2021). *Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic & E-Commerce Outcomes. Working Paper.*

Goldfarb, A. (2004). Concentration in Advertising-Supported Online Markets: An Empirical Approach. *Economics of Innovation and New Technology.* 13(6), 581–594.

Goldfarb, A. and Tucker, C. (2011). Online Display Advertising: Targeting and Obtrusiveness. *Marketing Science.* 30(3), 389–404.

Goldfarb, A. and Tucker, C. (2012). Privacy and innovation. *Innovation policy and the economy.* 12(1), 65–90.

Goldfarb, A. and Tucker, C. E. (2010). Privacy Regulation and Online Advertising. *Management Science.* 57(1), 57–71.

Goodman-Bacon, A. (2021). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics.* 225(2), 254–277.

Hils, M., Woods, D. W. and Böhme, R. (2020). Measuring the Emergence of Consent Management on the Web. In *Proceedings of the ACM Internet Measurement Conference.* 10. ACM, 317–332.

IAB Europe (2021). *GDPR Guidance: Legitimate Interests Assessments (LIA) for Digital Advertising.* Technical report. Retrievable at `https://iabeurope.eu/wp-content/uploads/2021/03/IAB-Europe-GDPR-Guidance-Legitimate-Interests-Assessments-LIA-for-Digital-Advertising-March-2021.pdf`.

IHS Technology (2015). *Paving the Way: How Online Advertising Enables the Digital Economy of the Future.* Technical report. Retrievable at `https://www.iabfrance.com/sites/www.iabfrance.com/files/atoms/files/iab_ihs_euro_ad_macro_finalpdf.pdf`.

Janssen, R., Kesler, R., Kummer, M. E. and Waldfogel, J. (2022). *GDPR and the Lost Generation of Innovative Apps. Working Paper.*

Jia, J., Jin, G. Z. and Wagman, L. (2021). The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment. *Marketing Science.* 40(4), 661–684.

Johnson, G. and Shriver, S. (2019). *Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR. Working Paper.*

Johnson, G. A., Shriver, S. K. and Du, S. (2020). Consumer Privacy Choice in Online Advertising: Who Opts Out and at What Cost to Industry? *Marketing Science.* 39(1), 33–51.

Lambrecht, A., Goldfarb, A., Bonatti, A., Ghose, A., Goldstein, D. G., Lewis, R., Rao, A., Sahni, N. and Yao, S. (2014). How Do Firms Make Money Selling Digital Goods Online? *Marketing Letters.* 25(3), 331–341.

Lefouili, Y. and Toh, Y. L. (2018). *Privacy Regulation and Quality Investment. Working Paper.*

Libert, T., Graves, L. and Nielsen, R. K. (2018). *Changes in Third-Party Content on European News Websites after GDPR.* Factsheet. Reuters Institute for the Study of Journalism Reports.

Lukic, K., Miller, K. and Skiera, B. (2022). The Impact of the General Data Protection Regulation (GDPR) on the Amount of Online Tracking. In *Workshop on the Economics of Privacy.* NBER.

Luo, X. and Zhang, J. (2013). How Do Consumer Buzz and Traffic in Social Media Marketing Predict the Value of the Firm? *Journal of Management Information Systems.* 30(2), 213–238.

Miller, A. R. and Tucker, C. (2009). Privacy Protection and Technology Diffusion: The Case of Electronic Medical Records. *Management Science.* 55(7), 1077–9.

Monic, S. and Feng, Z. (2013). Ad Revenue and Content Commercialization: Evidence from Blogs. *Management Science.* 59(10), 2314–2331.

Newman, N., Fletcher, R., Schulz, A., Andi, S., Robertson, C. T. and Nielsen, R. K. (2021). *Reuters Institute Digital News Report 2021.* Technical report. Reuters Institute for the Study of Journalism.

Nouwens, M., Liccardi, I., Veale, M., Karger, D. and Kagal, L. (2020). Dark Patterns After the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 4. ACM, 1–13.

Papadopoulos, P., Snyder, P., Athanasakis, D. and Livshits, B. (2020). Keeping out the masses: Understanding the popularity and implications of internet paywalls. In *Proceedings of The Web Conference 2020.* 1433–1444.

Peukert, C., Bechtold, S., Batikas, M. and Kretschmer, T. (2020). *European Privacy Law and Global Markets for Data. Working Paper.*

Peukert, C., Bechtold, S., Batikas, M. and Kretschmer, T. (2022). Regulatory spillovers and data governance: Evidence from the GDPR. *Marketing Science.*

Pew Research Center (2021a). News consumption across social media in 2021.

Pew Research Center (2021b). *State of the News Media.* Technical report. SSRN 3477686. Retrievable at `https://www.pewresearch.org/journalism/fact-sheet/newspapers/`.

Sanchez-Rola, I., Dell'Amico, M., Kotzias, P., Balzarotti, D., Bilge, L., Vervier, P.-A. and Santos, I. (2019). Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security.* 340–351.

Sharma, P., Sun, Y. and Wagman, L. (2019). *The Differential Effects of New Privacy Protections on Publisher and Advertiser Profitability. Working Paper.*

Shiller, B., Waldfogel, J. and Ryan, J. (2018). The Effect of Ad Blocking on Website Traffic and Quality. *The RAND Journal of Economics.* 49(1), 43–63.

Sørensen, J. and Kosta, S. (2019). Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites. In *The World Wide Web Conference.* 1590–1600.

Tucker, C. (2012). The Economics of Advertising and Privacy. *International Journal of Industrial Organization.* 30(7), 326–329.

UK Information Commissioner's Office (2019). *Update Report into Adtech and Real Time Bidding.* Technical report. Retrievable at `https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf`.

Urban, T., Tatang, D., Degeling, M., Holz, T. and Pohlmann, N. (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. New York, NY, USA: Association for Computing Machinery, 222–235.

Ursu, R. M., Zhang, Q. and Honka, E. (2022). Search Gaps and Consumer Fatigue. *Marketing Science.* Forthcoming.

Utz, C., Degeling, M., Fahl, S., Schaub, F. and Holz, T. (2019). (Un)Informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security.* CCS '19. New York, NY, USA: Association for Computing Machinery, 973–990.

Yan, S., Miller, K. M. and Skiera, B. (2022). How Does the Adoption of Ad Blockers Affect News Consumption? *Journal of Marketing Research.* Forthcoming.

Yesbeck, J. (2016). *Your Top Questions About Alexa Data and Ranks, Answered.* Retrievable at `https://web.archive.org/web/20180302091519/https://blog.alexa.com/top-questions-about-alexa-answered/`.

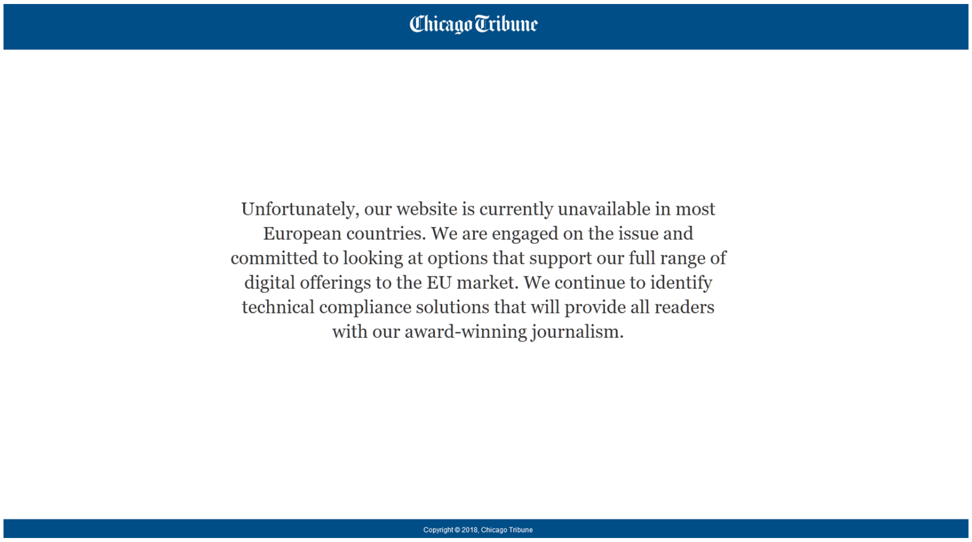
Zhang, K. and Katona, Z. (2012). Contextual Advertising. *Marketing Science.* 31(6), 980–994.

Zhao, Y., Yildirim, P. and Chintagunta, P. (2022). Privacy Regulations and Online Search Friction: Evidence from GDPR. In *Workshop on the Economics of Privacy.* NBER.

Zhuo, R., Huffaker, B., Claffy, K. and Greenstein, S. (2021). The Impact of the General Data Protection Regulation on Internet Interconnection. *Telecommunications Policy.* 45(2), 102083.

Appendix A:
Types of response

Fig. 8 *Example of Blocks EU Visitors*



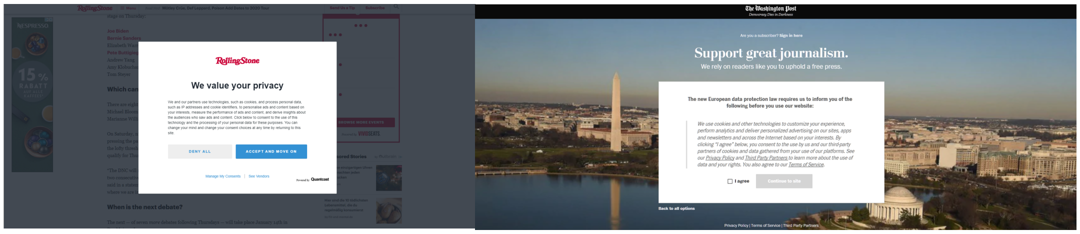
ChicagoTribune

Unfortunately, our website is currently unavailable in most European countries. We are engaged on the issue and committed to looking at options that support our full range of digital offerings to the EU market. We continue to identify technical compliance solutions that will provide all readers with our award-winning journalism.

Copyright © 2018, Chicago Tribune

Notes: This figure presents an example of Blocks EU websites

Fig. 9 *Examples of Consent Mechanisms:*



a *With a direct Opt-out button*

RollingStone

We value your privacy

We and our partners use technologies, such as cookies, and process personal data, such as IP addresses and cookie identifiers, to personalise ads and content based on your interests, measure the performance of ads and content, and derive insights about the audiences who saw ads and content. Click below to consent to the use of this technology and the processing of your personal data for these purposes. You can change your mind and change your consent choices at any time by returning to this site.

b *Without a direct Opt-out button*

The Washington Post
Democracy Dies in Darkness

Are you a subscriber? Sign in here

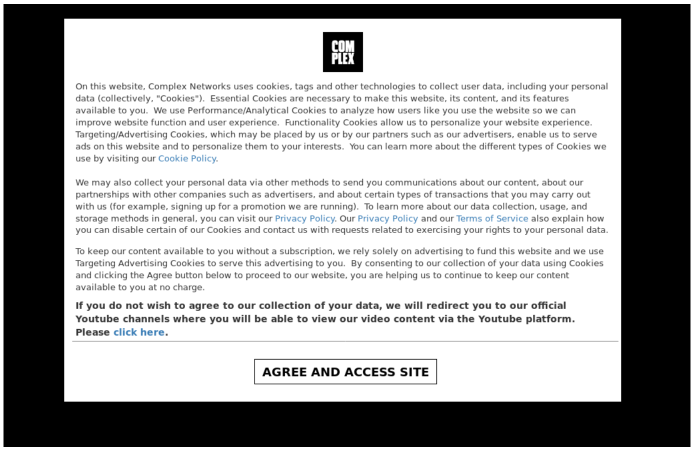
Support great journalism.
We rely on readers like you to uphold a free press.

The new European data protection law requires us to inform you of the following before you use our website:

We use cookies and other technologies to customize your experience, perform analytics and deliver personalized advertising on our sites, apps and newsletters and across the Internet based on your interests. By clicking "I agree" below, you consent to the use by us and our third-party partners of cookies and data gathered from your use of our platforms. See our Privacy Policy and Third Party Partners to learn more about the use of data and your rights. You also agree to our Terms of Service.

Notes: This figure presents different kinds of consent mechanisms

Fig. 10 *Example of Cookie Wall*



On this website, Complex Networks uses cookies, tags and other technologies to collect user data, including your personal data (collectively, "Cookies"). Essential Cookies are necessary to make this website, its content, and its features available to you. We use Performance/Analytical Cookies to analyze how users like you use the website so we can improve website function and user experience. Functionality Cookies allow us to personalize your website experience. Targeting/Advertising Cookies, which may be placed by us or by our partners such as our advertisers, enable us to serve ads on this website and to personalize them to your interests. You can learn more about the different types of Cookies we use by visiting our Cookie Policy.

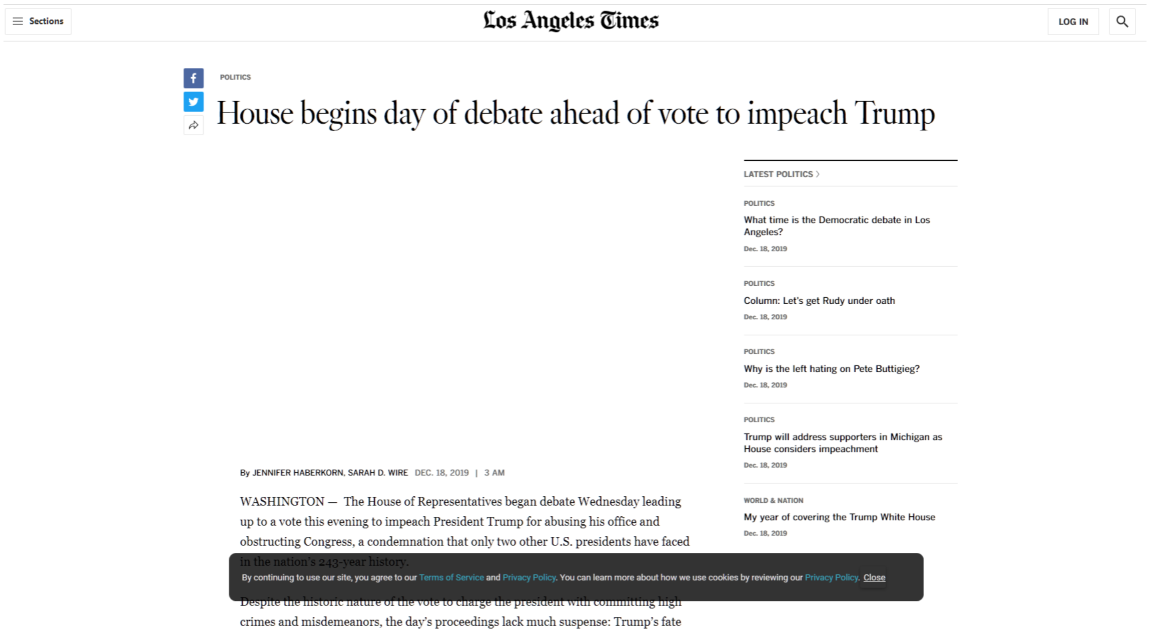
We may also collect your personal data via other methods to send you communications about our content, about our partnerships with other companies such as advertisers, and about certain types of transactions that you may carry out with us (for example, signing up for a promotion we are running). To learn more about our data collection, usage, and storage methods in general, you can visit our Privacy Policy. Our Privacy Policy and our Terms of Service also explain how you can disable certain of our Cookies and contact us with requests related to exercising your rights to your personal data.

To keep our content available to you without a subscription, we rely solely on advertising to fund this website and we use Targeting Advertising Cookies to serve this advertising to you. By consenting to our collection of your data using Cookies and clicking the Agree button below to proceed to our website, you are helping us to continue to keep our content available to you at no charge.

If you do not wish to agree to our collection of your data, we will redirect you to our official Youtube channels where you will be able to view our video content via the Youtube platform. Please click here.

Notes: This figure presents a Cookie wall example

Fig. 11 *Example of Cookie Notice*



POLITICS

House begins day of debate ahead of vote to impeach Trump

LATEST POLITICS >

POLITICS

What time is the Democratic debate in Los Angeles?

Dec. 18, 2019

POLITICS

Column: Let's get Rudy under oath

Dec. 18, 2019

POLITICS

Why is the left hating on Pete Buttigieg?

Dec. 18, 2019

POLITICS

Trump will address supporters in Michigan as House considers impeachment

Dec. 18, 2019

WORLD & NATION

My year of covering the Trump White House

Dec. 18, 2019

By JENNIFER HABERKORN, SARAH D. WIRE DEC. 18, 2019 | 3 AM

WASHINGTON — The House of Representatives began debate Wednesday leading up to a vote this evening to impeach President Trump for abusing his office and obstructing Congress, a condemnation that only two other U.S. presidents have faced in the nation's 243-year history.

By continuing to use our site, you agree to our Terms of Service and Privacy Policy. You can learn more about how we use cookies by reviewing our Privacy Policy. Close

Despite the historic nature of the vote to charge the president with committing high crimes and misdemeanors, the day's proceedings lack much suspense: Trump's fate

Notes: This figure presents different kinds of consent mechanisms

Appendix B:
Data Collection Strategy

In this section we provide additional details on our website sampling strategies and data collection strategies.

B.1 Websites Sample Selection

The data panel used in this paper is based on a broader data collection effort by the authors that scrapes top-ranked and long-tail (low-ranked) websites in the US and several EU countries (Germany, France, UK, Italy, Spain, and the Netherlands). While that panel includes a variety of websites, in this manuscript we focus exclusively on content providers (publishers), such as news websites and online magazines, because our aim is to determine how GDPR may affect content providers that greatly rely on online tracking and behavioral targeted advertising for revenues. The websites included in this study are all sites from our broader panel that are classified as news and media sites by SimilarWeb[29]. Below we provide a short description of how our broader panel was constructed.

Using 2018 Alexa data we identify the top 500 websites from various geographical areas (Global, Germany, France, UK, Italy, Spain, and the Netherlands) and 5 popular content categories (News, Sports, Society, Health, and Games). Alexa's top 500 websites by country correspond to the websites most *visited* by users in that country (rather than the most popular websites that are *based* in that country). To include enough top websites *based* in each of our areas of interest (EU and US), we used Alexa's global top 1 million websites to complement the data set with the top 500 websites for the top-level domains associated with our countries of interest (*.de, .fr, .uk, .it, .es, .nl, .com, .net* and *.us*). Finally, to also include long-tail websites we add a random sample of websites (considering only sites from our countries of interest) ranked between 200,000 and 1 million. Specifically, we included in the panel 500 random websites for each 100k websites ranking interval, i.e., 500 websites ranked between 200k and 300k, 500 websites ranked between 300k and 400k, and so on until reaching 1 million.

To obtain a set of website level characteristics we use data from SimilarWeb. For each

[29] See `https://www.similarweb.com/`.

website we capture: Its content category, the share of its users originating from each country, the location of its headquarters, among others. We use this data to classify websites as EU or US based by looking at the location of its headquarters. If this is not available, we infer the location by the country of the its top-level domain (such as .fr or .us). If the site is not assigned to a country top level domain (e.g., .com or .org), we assign it to the country where most visitors originated from. We also use this data to exclude websites that receive less than 10% of its visitors from the EU or the US. Finally, for this paper we only focus on websites in the "News and Media" category. The resulting dataset used in this paper consist of 909 news and media sites located in the US or the EU. No news and media website that was classified as based in the EU or the US received less than 10% of its traffic from either the US or the EU.

B.2 Data Collection

For each News and Media website in our sample we collect two categories of data. The first category includes data we mine directly from each website at regular intervals, such as HTML data, cookies, screenshots, and HTTP responses. We use these raw data to extract "technical variables"(see Section 4.1). The goal of the technical variables is to capture websites' behavior (including provision of consent mechanisms, tracking, privacy, and advertising choices) and changes in that behavior following the implementation of the GDPR.

The second category of data is obtained from third parties' repositories. We use these repositories to measure changes in the quantity of content offered by the websites in the sample as well as traffic to and user engagement with such content (a proxy for its quality). We refer to the metrics extracted from repositories data as "downstream outcomes" (see Section 4.2). These metrics do not change as function of the country of the visitor. However, we do expect to find differences depending on the location of registration of the website, as websites registered in different locations (EU vs. US) should be affected differently by the GDPR.

The data collected span a period of time of at least 19 months for technical variable metrics (from April 2018 to November 2019), and 31 months for downstream outcomes (from April 2017 to November 2019).

B.2.1 Technical Variables

We extract technical variables from raw website data collected directly from each website. We use OpenWPM—a web privacy measurement framework (Englehardt and Narayanan, 2016)—to simulate user browsing and capture the website's interaction with its visitors. The framework is implemented within an instrumented web browser that automates the process of visiting a set of websites and records a series of variables. We refer to each round of visits to all websites as a "wave" of data collection. During each wave, we visit each website twice at the same time from two different visitor IP addresses, one located in Europe (France) and one in the US.

This design allows us to compare, before and after the enactment of the GDPR, whether and how websites adapted their data collection behavior according to the geographical location of a visitor. The categories of data collected include screenshots (including visual interface elements such as buttons to accept cookies and user-facing messaging) to classify visual elements of websites that may indicate a website's response to the GDPR; cookies (including third-party cookies) set by the websites on visitors' browsers; HTML data (including privacy notices) to capture a website's references to relying on legitimate interest to justify data collection; and HTTP responses (including all the information exchanged between the browser and the websites visited) to capture a website's advertising patterns. From these data we construct a number of technical variables that capture websites' behaviors (including tracking, privacy notices, advertising choices, consent mechanisms) and changes in behaviors in response to the GDPR. Below, we discuss the variables that we extract from these different categories of data.

Cookies: Cookies are small files stored on visitors' browsers and often embedded on websites to provide additional functionality. Cookies are extensively used for advertising purposes—for example, to store information on the websites or products visited by a user. Our data collection focused on two types of cookies: 1st party and third party cookies. The variable *1st Party Cookies* measures the cookies that are set by the website being browsed. The variable *3rd Party Cookies* represents cookies that are set by entities other than the original website, and that could be used to track users' behavior across different websites in order to construct users' profiles aimed, in part, at improving behaviorally targeted ads.

We also identify, among these cookies, which are known to be used for tracking or advertising. The variable *Advertising Cookies* counts the number of cookies set by advertising companies. We identify these by using scripts included in popular ad blockers that flag advertising content. [30] We use the same methodology to identify the number of tracking cookies, which are recorded in the variable *Tracking Cookies.* We rely on a drop to zero in either advertising or tracking cookies to identify when a website responds to the GDPR by halting the tracking of (EU) visitors.

Advertising Intensity: To analyze the volume of advertising displayed to visitors when browsing websites in our panel, we captured the length (in bytes) of certain types of websites' HTML content, using scripts included in popular ad blockers to flag advertising content within the HTTP response content we extracted from each website. The variable *Advertising Intensity* captures the size, in kilobytes, of the quantity of advertising content on a website's homepage. It is constructed by measuring the length of the content that is identified as advertising by *Adblock Easylist*[31].

Website Responses: We use the visual elements of websites' interfaces that appear within screenshots to distinguish between types of website responses. Specifically, we use screenshots to distinguish between websites that implement consent mechanisms, cookie walls, cookie banners, or block EU visitors. We consider a consent mechanism to be a banner or pop-up that offers users the ability to reject tracking. This can be either through a "reject" button or through sub-menus such as a "settings" menu (for example, Figure 9a and Figure 9b in the Appendix). By contrast, cookie banners inform users about cookies, but do not provide them

[30] An ad blocker is a small piece of software or module incorporated into a user's browser (add-on) that prevents the display of banners and other advertising formats. Ad-blockers filter advertisements using community maintained lists that contain the URLs and HTML tags used by the main ad servers and advertising networks (these lists are known as blocklists). We cross-reference the data we collected using OpenWPM with these blocklists to identify advertising related cookies and content. We rely on two blocklists (last retrieved in February 2020): Adblock Plus (`https://adblockplus.org/fr/subscriptions`), and Disconnect (`https://disconnect.me/`).

[31] AdBlock Easylist consists of a set of rules used by AdBlockers to detect and hide elements that correspond to advertising. We re-purpose these rules to identify and measure the length of advertising instead of hiding it. The list is available at `easylist.to`

with a way to reject tracking (see Figure 11, in the Appendix). We distinguish cookie walls by the fact that the cookie walls prevent visitors from viewing content and do not provide a means (through buttons or links) to reject tracking (see figure 10, in the Appendix). Finally, we are able to identify US websites that decided to block EU consumers (visitors) by identifying a static page shown to EU visitors informing them that the website is unavailable (see Figure 8, in the Appendix). For each of the responses so identified, we create a dummy variable that takes on the value 1 if the corresponding response is implemented by a given website, and 0 otherwise.

Privacy Policies: We analyze websites' HTML to extract their privacy policies over time. We then use text analysis to infer which websites invoke legitimate business interest as a justification for data collection under GDRP.

B.2.2 Downstream Outcomes

We collect content-related metrics from third parties' repositories to measure downstream changes in quantity of content generated by websites in the panel, and changes in traffic and user engagement with such content.

To measure content quantity, we use the *Global Database of Events, Language, and Tone* (GDELT).[32] GDELT gathers and provides metadata for articles from news and media websites going back to 2015 from both domestic (US) and international sources. The database provides metadata including the URL, publication date, and publisher website for each article, and has been used in studies that examine global events (Gallea and Rohner, 2021; Ferreira *et al.*, 2021). We use GDELT data to count the number of new URLs of content published by each website in our sample in the week surrounding each observation from OpenWPM (three days before and after each OpenWPM observation). Because we visit each website multiple times to construct our longitudinal data set, we collect multiple observations of the new URL counts for each website over time.

We use websites' traffic metrics (Page Views Per User, Page Views Per Million, Reach, and Rank) and visitors' engagement (as measured by social media reactions) as a proxy for

[32] `gdeltproject.org`

content quality. The underlying premise is that, were the quality of the content provided by the website to decrease, users might try to substitute for other content and, therefore, we should observe a decrease in the number of visits to a given website.

Websites traffic metrics are obtained from Amazon Alexa web metrics (Shiller *et al.*, 2018; Luo and Zhang, 2013; Utz *et al.*, 2019; Sørensen and Kosta, 2019).[33] We use Alexa's *Rank*, a measure of a website's popularity that is calculated (by Alexa) by combining measures of page views and unique visitors. We use Alexa's *Reach Per Million* as a measure of the number of (unique) users visiting a website. [34] We use Alexa's *Page Views Per Million* as a measure of the number of pages viewed by visitors. Finally, we use Alexa's *Page Views Per User*, which represents the average number of unique pages viewed per user, per day, by the users visiting a website.

We capture social media "reactions" related to the content published on the websites in our sample using the Facebook Graph API, in line with Cagé *et al.* (2020) methodology, who used the same metric as a proxy of quality for online news websites. For each new URL of content posted by each website in our sample during the week surrounding the data collection in each wave (as retrieved via GDELT), we collect the number of reactions on the Facebook platform and calculate the average number of *Facebook Reactions* across all new URLs by website/wave. We call this average the *FB Average Reaction.* Such reactions can be used to measure users' engagement with a piece of content, and can be interpreted as a proxy for content quality. Table 7 presents the descriptive statistics of the technical variables and the downstream outcomes, for the overall sample, across all waves.

[33] See `https://www.alexa.com/`.

[34] Unique visitors are determined by the number of unique Alexa users who visit a website on a given day.

Appendix C:
Additional figures

Fig. 12 *Mean Advertising Intensity for Sites by Type of Website-Level Response to the GDPR*

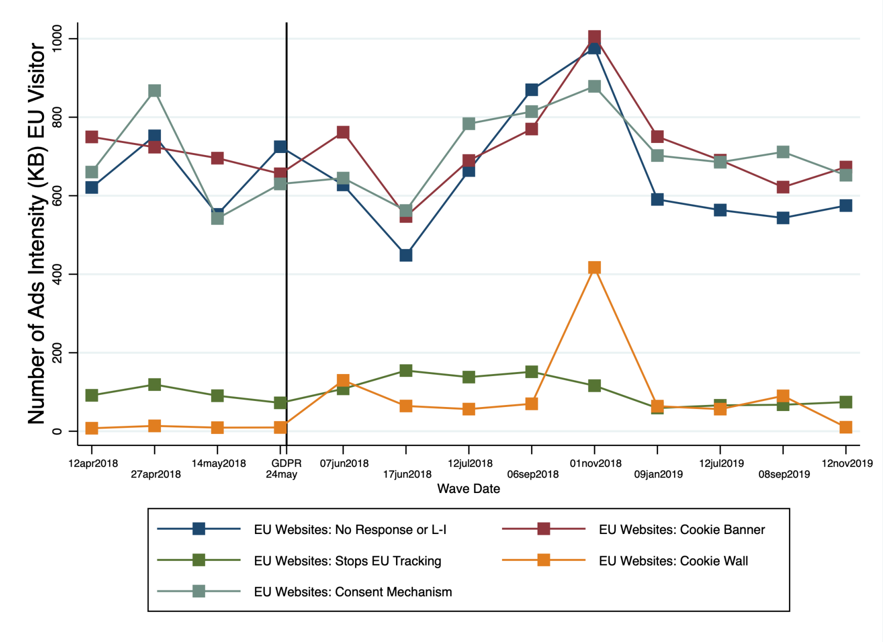


Fig. 13 *3rd Party Cookies By Type of Response*

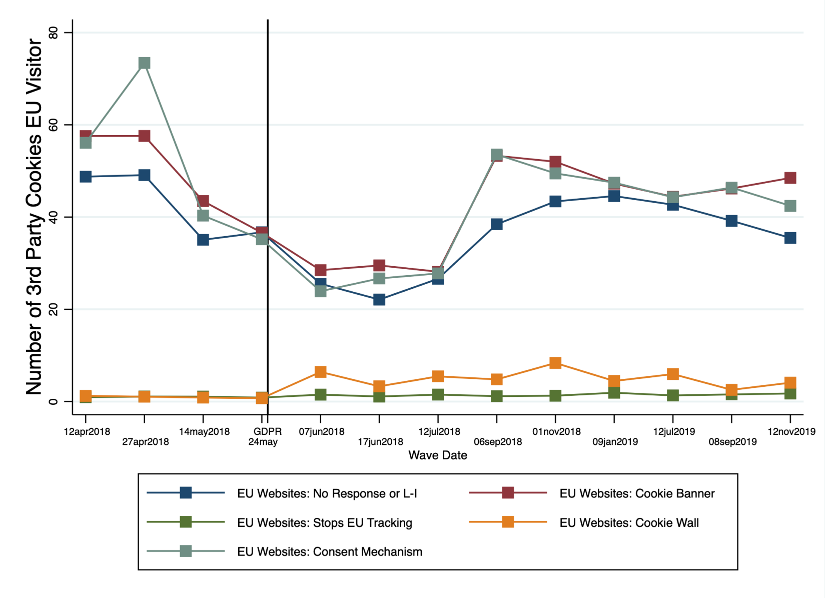


Fig. 14 *Consent Mechanism EU/US Websites for EU/US Visitors*

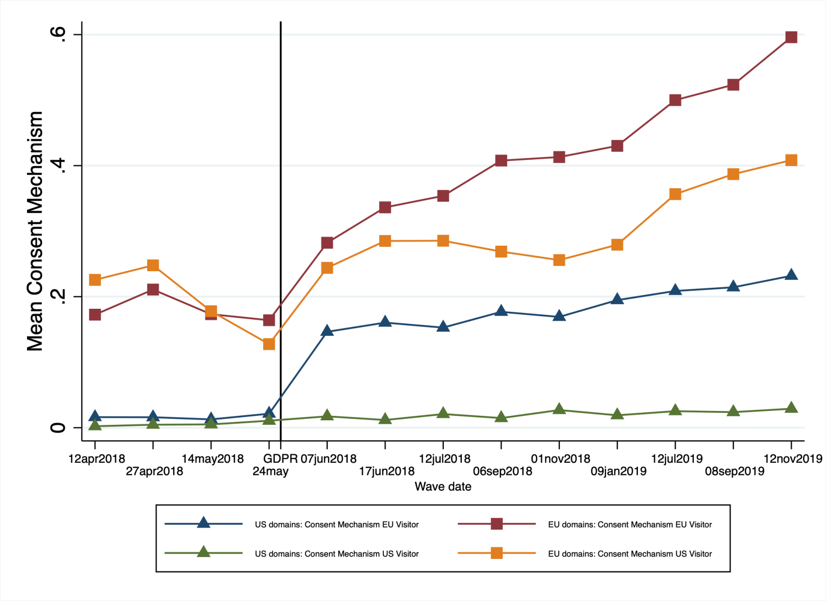


Fig. 15 *Mean Advertising Intensity for Sites that Decrease/Do not Decrease 3rd Party Cookies*

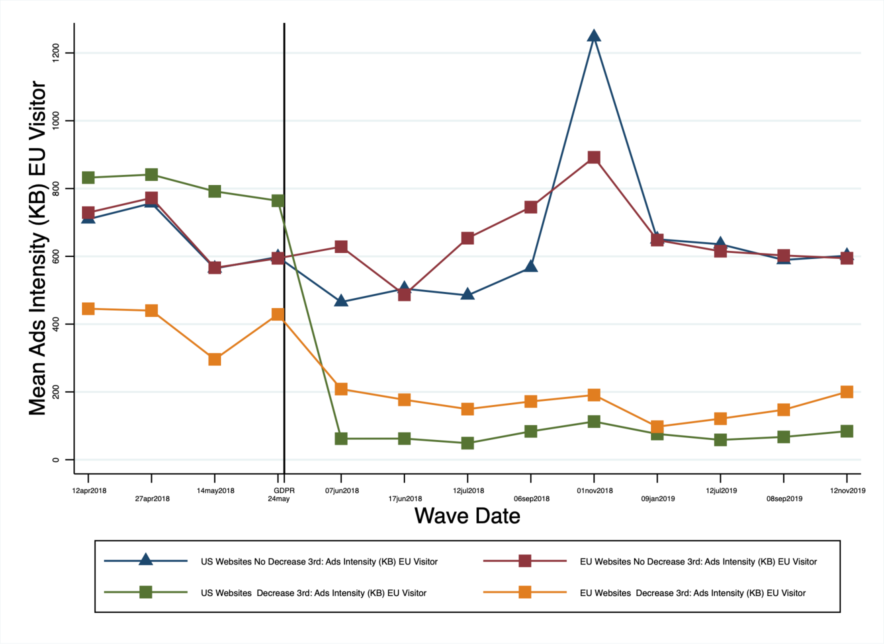


Fig. 16 *Website-Level Responses to GDPR by EU Websites*

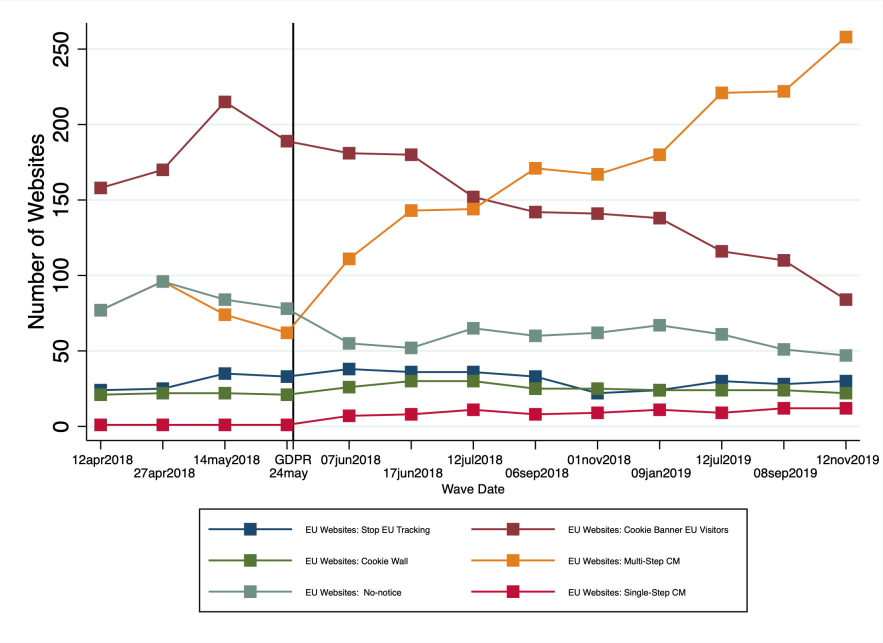


Fig. 17 *3rd Party Cookies Companies EU/US Websites for EU/US Visitors*

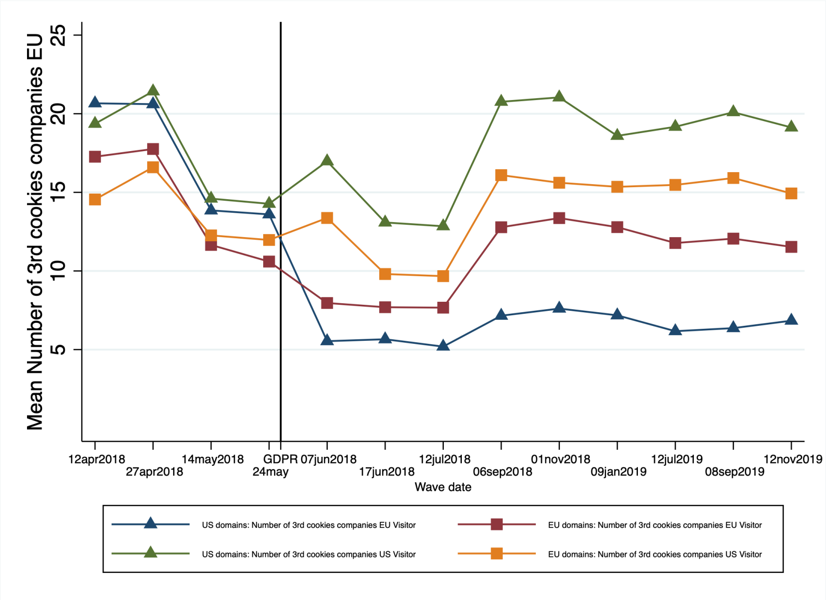
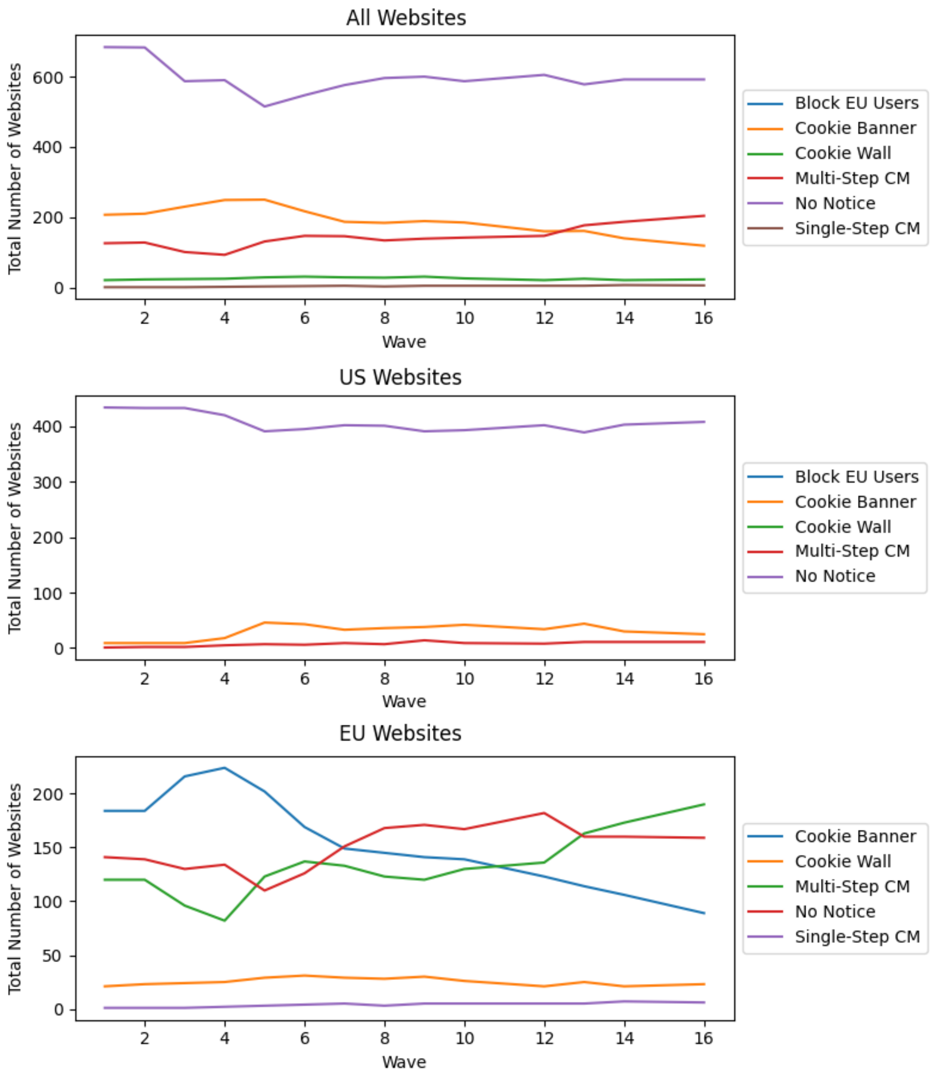


Fig. 18 *Counts of Websites Responses to the GPDR by Wave for US Visitors from Screenshot Data (GDPR Enforcement after Wave 4)*


All Websites

US Websites

EU Websites

Fig. 19 *Counts of Websites Responses to the GPDR by Wave for EU Visitors from Screenshot Data (GDPR Enforcement after Wave 4)*

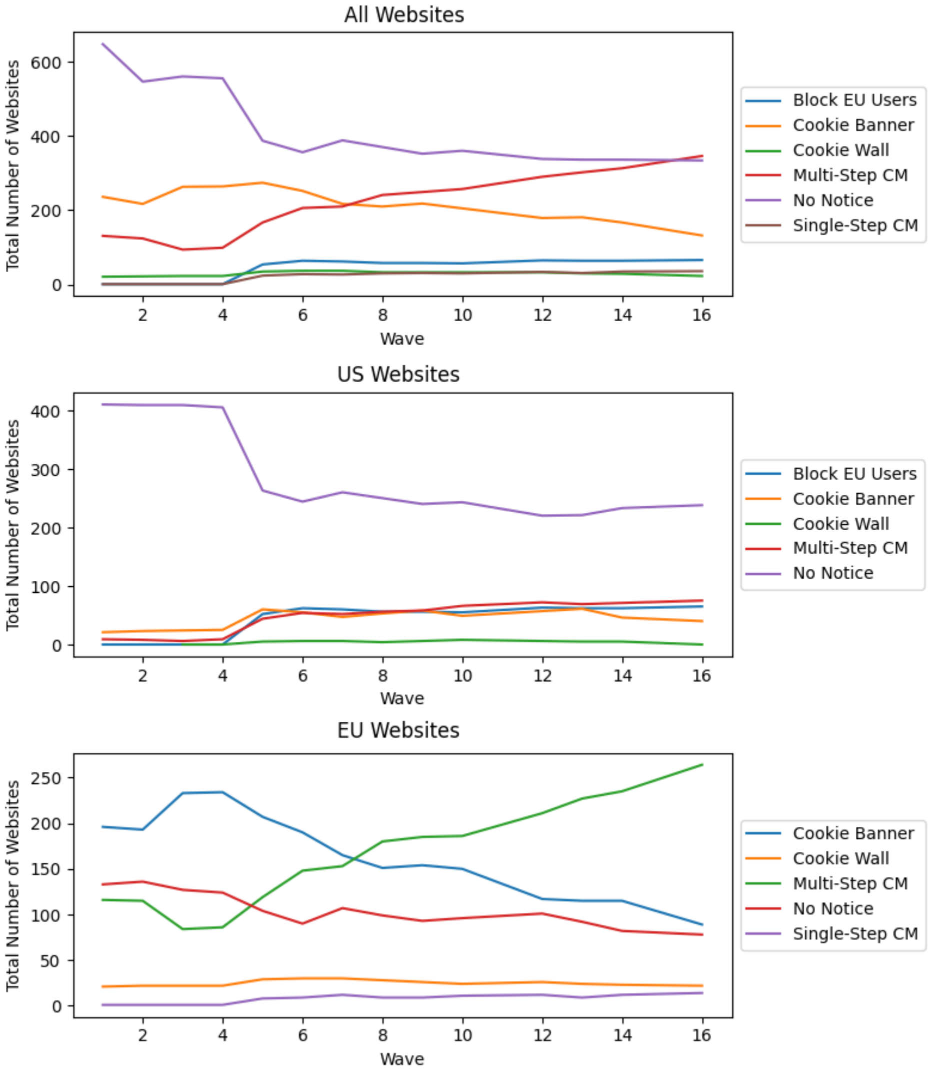
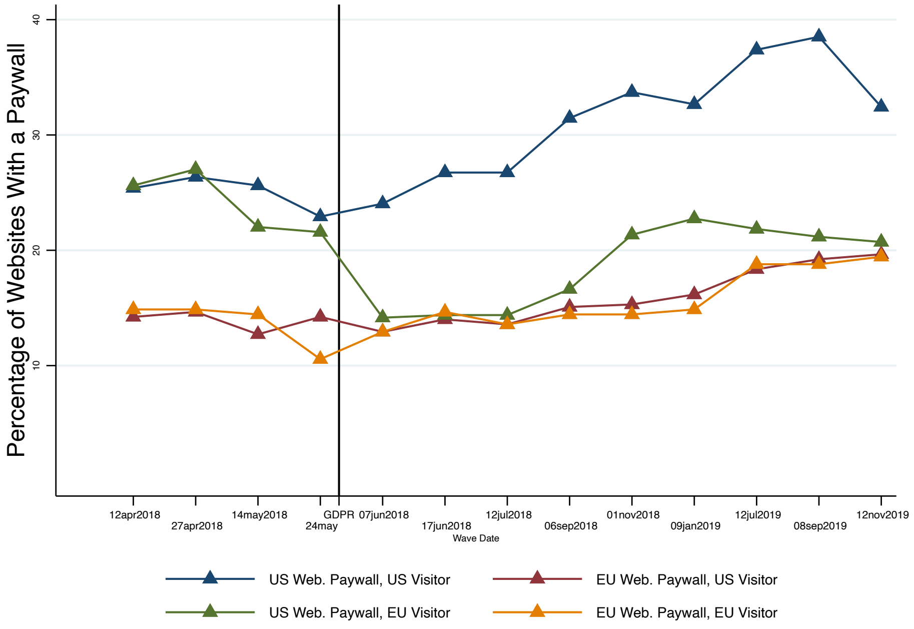


Fig. 20 *Paywalls on EU/US Websites for EU/US Visitors*



Appendix D:
Additional tables

Table 7: Descriptive Statistics — Entire Sample

	Mean	Std. Dev.	Min	Max	N
	Technical variables				
Tracking:					
1st Party Cookies EU Visitor	11.058	8.008	0.0	52.0	11,114
3rd Party Cookies US Visitor	33.269	37.620	0.0	272.0	11,114
1st Party Cookies US Visitor	12.942	8.288	0.0	50.0	11,175
3rd Party Cookies EU Visitor	53.892	50.060	0.0	351.0	11,175
Advertising:					
Advertising Intensity EU Visitor	576.564	1,279.351	0.0	111,046.5	11,107
Advertising Intensity US Visitor	695.740	932.966	0.0	28,571.3	11,172
Website-Level Responses:					
Blocking EU Visitor	0.023	0.151	0.0	1.0	21,798
Stop EU Tracking	0.118	0.322	0.0	1.0	11,114
Consent Mechanism EU Visitor	0.133	0.340	0.0	1.0	21,798
Cookie Wall EU Visitor	0.017	0.130	0.0	1.0	21,798
Cookie Banner EU Visitor	0.101	0.302	0.0	1.0	21,798
Website Visitors:					
Share of EU Visitors	0.430	0.420	0.0	1.0	21,798
Share of US Visitors	0.395	0.403	0.0	1.0	21,798
	Downstream Outcomes				
Log GDELT URLs	5.014	1.705	0.0	9.6	17,588
Page Views Per Million	14.463	61.499	0.0	1,451.4	21,797
Reach Per Million	243.692	862.081	0.0	18,714.3	21,797
Page Views Per User	2.032	0.953	0.6	19.1	21,797
Rank	65,063.040	100,954.740	0.0	1,832,762.9	21,797
FB Average Reaction	110.463	466.708	0.0	12,476.9	17,588

Table 8: DID Estimations by Low and High Advertising

	Log GDELT URLs		Reach Per Million		Page Views Per Million		Page views per user		Rank		FB Average Reaction	
	(1) Low Ads	(2) High Ads	(3) Low Ads	(4) High Ads	(5) Low Ads	(6) High Ads	(7) Low Ads	(8) High Ads	(9) Low Ads	(10) High Ads	(11) Low Ads	(12) High Ads
Intention to Treat 1: Treatment group based on EU Websites												
Treated (EU Web) × Post GDPR	0.058	-0.015	40.265	11.333	0.383	0.223	-0.063	-0.135***	346.684	2847.824	3.405	11.129
	(0.067)	(0.055)	(27.587)	(14.621)	(1.662)	(0.834)	(0.057)	(0.036)	(5653.017)	(3787.949)	(23.361)	(18.528)
Constant	4.663***	5.288***	284.931***	195.507***	17.914***	11.008***	2.064***	2.036***	67926.571***	60787.870***	128.237***	95.349***
	(0.013)	(0.008)	(5.840)	(2.409)	(0.352)	(0.138)	(0.012)	(0.006)	(1196.752)	(624.157)	(4.730)	(2.811)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	7819	9196	10614	10481	10614	10481	10614	10481	10614	10481	7819	9196
Intention to Treat 2: Treatment group based on EU Websites + Websites with more than 10% of EU Visitors												
EU Websites and > 10% EU visitors × Post GDPR	0.052	-0.015	47.111*	14.310	0.444	0.334	-0.118**	-0.136***	54.909	2972.581	10.198	10.096
	(0.066)	(0.055)	(26.872)	(14.564)	(1.611)	(0.832)	(0.058)	(0.036)	(5636.333)	(3791.712)	(23.198)	(18.451)
Constant	4.664***	5.288***	283.792***	195.027***	17.904***	10.990***	2.074***	2.036***	67988.702***	60769.582***	126.940***	95.514***
	(0.013)	(0.008)	(5.512)	(2.389)	(0.330)	(0.136)	(0.012)	(0.006)	(1156.048)	(621.883)	(4.519)	(2.783)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	7819	9196	10614	10481	10614	10481	10614	10481	10614	10481	7819	9196
Intention to Treat 3: Treatment group based on EU Websites + Websites with more than 5% of EU Visitors												
EU Websites and > 5% EU visitors × Post GDPR	0.053	-0.015	40.189	11.333	0.392	0.223	-0.107*	-0.135***	1484.354	2847.824	12.512	11.129
	(0.066)	(0.055)	(26.970)	(14.621)	(1.624)	(0.834)	(0.058)	(0.036)	(5618.205)	(3787.949)	(23.259)	(18.528)
Constant	4.664***	5.288***	285.144***	195.507***	17.914***	11.008***	2.072***	2.036***	67692.996***	60787.870***	126.468***	95.349***
	(0.013)	(0.008)	(5.577)	(2.409)	(0.336)	(0.138)	(0.012)	(0.006)	(1161.858)	(624.157)	(4.569)	(2.811)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	7819	9196	10614	10481	10614	10481	10614	10481	10614	10481	7819	9196

Notes: Standard errors in parentheses and clustered at the website level. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$

Table 9: DID Estimations for Top and Bottom Subsamples of Website in the EU and in the US

	Log GDELT URLs		Reach Per Million		Page Views Per Million		Page views per user		Rank		FB Average Reaction	
	(1) Top	(2) Bottom	(3) Top	(4) Bottom	(5) Top	(6) Bottom	(7) Top	(8) Bottom	(9) Top	(10) Bottom	(11) Top	(12) Small
Intention to Treat 1: Treatment group based on EU Websites												
Treated (EU Web) × Post GDPR	0.088	-0.091	225.053**	-0.342	0.704	-0.048	-0.255***	-0.179*	1816.087	37442.148*	110.594	19.397
	(0.137)	(0.121)	(109.755)	(1.127)	(7.969)	(0.044)	(0.072)	(0.096)	(2018.569)	(21801.794)	(100.603)	(17.404)
Constant	5.955***	3.938***	1671.769***	6.652***	106.206***	0.310***	2.307***	1.889***	912.707**	2.47e+05***	319.995***	24.087***
	(0.026)	(0.016)	(21.036)	(0.214)	(1.527)	(0.008)	(0.014)	(0.018)	(386.885)	(4145.981)	(19.073)	(2.370)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean rank of US websites	996.843	275140.590	996.843	275140.590	996.843	275140.590	996.843	275140.590	996.843	275140.590	996.843	275140.590
Mean rank of Treatment group	3096.626	250088.450	3096.626	250088.450	3096.626	250088.450	3096.626	250088.450	3096.626	250088.450	3096.626	250088.450
Number of Control group	45	44	45	44	45	44	45	44	45	44	45	44
Number of EU websites	47	47	47	47	47	47	47	47	47	47	47	47
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	1978	1395	2207	2156	2207	2156	2207	2156	2207	2156	1978	1395
Intention to Treat 2: Treatment group based on EU Websites + Websites with more than 10% of EU Visitors												
EU Websites and > 10% EU visitors × Post GDPR	0.082	-0.107	230.227**	-0.321	0.394	-0.049	-0.254***	-0.191**	2054.285	37813.341*	124.993	19.611
	(0.137)	(0.121)	(103.959)	(1.149)	(7.464)	(0.045)	(0.071)	(0.095)	(2201.759)	(21953.529)	(96.575)	(17.204)
Constant	5.957***	3.939***	1674.533***	6.647***	106.272***	0.310***	2.303***	1.891***	900.563**	2.47e+05***	319.540***	24.185***
	(0.024)	(0.016)	(18.229)	(0.214)	(1.309)	(0.008)	(0.012)	(0.018)	(386.081)	(4083.193)	(16.552)	(2.232)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean rank of US websites	962.583	272785.833	962.583	272785.833	962.583	272785.833	962.583	272785.833	962.583	272785.833	962.583	272785.833
Mean rank of Treatment group	3330.995	251904.428	3330.995	251904.428	3330.995	251904.428	3330.995	251904.428	3330.995	251904.428	3330.995	251904.428
Number of Control group	49	45	49	45	49	45	49	45	49	45	49	45
Number of EU websites	43	46	43	46	43	46	43	46	43	46	43	46
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	1978	1395	2207	2156	2207	2156	2207	2156	2207	2156	1978	1395
Intention to Treat 3: Treatment group based on EU Websites + Websites with more than 5% of EU Visitors												
EU Websites and > 5% EU visitors × Post GDPR	0.079	-0.107	209.030*	-0.321	0.527	-0.049	-0.251***	-0.191**	1935.518	37813.341*	137.821	19.611
	(0.137)	(0.121)	(107.124)	(1.149)	(7.700)	(0.045)	(0.071)	(0.095)	(2106.032)	(21953.529)	(97.977)	(17.204)
Constant	5.957***	3.939***	1676.545***	6.647***	106.244***	0.310***	2.304***	1.891***	905.603**	2.47e+05***	316.087***	24.185***
	(0.025)	(0.016)	(19.658)	(0.214)	(1.413)	(0.008)	(0.013)	(0.018)	(386.472)	(4083.193)	(17.683)	(2.232)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean rank of US websites	988.423	272785.833	988.423	272785.833	988.423	272785.833	988.423	272785.833	988.423	272785.833	988.423	272785.833
Mean rank of Treatment group	3198.744	251904.428	3198.744	251904.428	3198.744	251904.428	3198.744	251904.428	3198.744	251904.428	3198.744	251904.428
Number of Control group	47	45	47	45	47	45	47	45	47	45	47	45
Number of EU websites	45	46	45	46	45	46	45	46	45	46	45	46
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	1978	1395	2207	2156	2207	2156	2207	2156	2207	2156	1978	1395

Notes: Standard errors in parentheses and clustered at the website level. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$

Table 10: DID Estimations by Short and Long-Run Subsample

	Log GDELT URLs		Reach Per Million		Page Views Per Million		Page views per user		Rank		FB Average Reaction	
	(1) Short Run	(2) Long Run	(3) Short Run	(4) Long Run	(5) Short Run	(6) Long Run	(7) Short Run	(8) Long Run	(9) Short Run	(10) Long Run	(11) Short Run	(12) Long Run
Intention to Treat 1: Treatment group base on EU Websites												
Treated (EU Web) × Post GDPR	-0.007	0.019	15.193	26.686	0.310	-1.101	-0.023	-0.234***	1967.806	3912.734	13.815	10.722
	(0.038)	(0.067)	(10.923)	(21.211)	(0.743)	(1.299)	(0.031)	(0.057)	(3289.256)	(5155.478)	(15.083)	(24.070)
Constant	5.037***	5.041***	251.965***	248.028***	14.902***	14.757***	2.051***	2.056***	60792.165***	64104.940***	106.496***	108.150***
	(0.005)	(0.005)	(1.594)	(1.801)	(0.108)	(0.110)	(0.004)	(0.005)	(480.017)	(437.824)	(2.055)	(1.819)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	15566	13222	19077	16344	19077	16344	19077	16344	19077	16344	15566	13222
Intention to Treat 2: Treatment group base on EU Websites + Websites with more than 10% of EU Visitors												
EU Websites and > 10% EU visitors × Post GDPR	-0.010	0.024	20.648*	32.061	0.516	-1.180	-0.039	-0.279***	1043.402	5107.264	16.283	11.719
	(0.038)	(0.067)	(10.782)	(20.964)	(0.731)	(1.287)	(0.031)	(0.056)	(3311.108)	(5130.506)	(14.960)	(23.715)
Constant	5.037***	5.041***	251.227***	247.624***	14.873***	14.761***	2.053***	2.060***	60930.021***	64011.932***	106.210***	108.095***
	(0.005)	(0.005)	(1.543)	(1.746)	(0.105)	(0.107)	(0.004)	(0.005)	(473.834)	(427.228)	(1.992)	(1.751)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	15566	13222	19077	16344	19077	16344	19077	16344	19077	16344	15566	13222
Intention to Treat 3: Treatment group base on EU Websites + Websites with more than 5% of EU Visitors												
EU Websites and > 5% EU visitors × Post GDPR	-0.010	0.024	16.686	25.225	0.431	-1.266	-0.036	-0.270***	2234.922	4650.988	16.944	15.165
	(0.038)	(0.067)	(10.834)	(21.052)	(0.735)	(1.291)	(0.031)	(0.057)	(3302.734)	(5138.329)	(14.994)	(23.822)
Constant	5.037***	5.041***	251.779***	248.180***	14.885***	14.769***	2.053***	2.059***	60757.401***	64047.366***	106.107***	107.831***
	(0.005)	(0.005)	(1.561)	(1.765)	(0.106)	(0.108)	(0.004)	(0.005)	(475.752)	(430.709)	(2.010)	(1.773)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	15566	13222	19077	16344	19077	16344	19077	16344	19077	16344	15566	13222

Notes: Standard errors in parentheses and clustered at the website level. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.