The Impact of the GDPR on Content Providers: A Longitudinal Analysis

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

Work in Progress. March 2022

Abstract

The European General Data Protection Regulation (GDPR) has received significant attention in economic research, but concerns that it would adversely affect websites' ability to provide content to visitors have not been adequately investigated. We use a longitudinal data-set to study how online content-providing websites adapted their response 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 provide 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, those initial reductions are followed, several months after the enactment of the regulation, by a reversal of the trend and an uptick in tracking among EU websites. We use difference-in-differences, LATE, and look ahead matching models to assess downstream effects, distinguishing between ecosystem effects (which affect all EU-based websites relative to US-based websites) and website-level effects (which depend on individual websites' specific responses to the GDPR). We document a small reduction in average page views per visitors in 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, ranking, and social media engagement with new content.

[*] 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.

The authors thank Madhukar Mohta and Paritosh Gupta for excellent research assistantship. The authors also gratefully acknowledge support from the Alfred P. Sloan Foundation. For a complete list of Acquisti's additional grants, please visit `https://www.heinz.cmu.edu/~acquisti/cv.html`. Work for this paper was supported by the CARNOT TSN.

1 Introduction

In May 2018, the European Union (EU) implemented the General Data Protection Regulation (GDPR) to enhance individuals' control over personal data. The enactment sparked interest among academics, policy-makers, and industry actors worldwide. Much empirical attention has been devoted to measuring websites' compliance with the GDPR, documenting changes in online consent mechanisms, and estimating compliance costs. Less attention has been devoted to understanding downstream consequences that the regulation might have on economically important metrics, such as the ability of websites to produce content, and the ability of Internet users to enjoy it. In a longitudinal study spanning data collected before and after the enactment of the GDPR, we track how online content providers (News and Media websites) adapted their response to the GDPR over time, and whether restrictions on online tracking enforced by the regulation ultimately affected websites' downstream outcomes, such as the quantity of content they offer to their visitors and visitors' engagement with such content.

A defining characteristic of the GDPR consists in restrictions it places on the collection and processing of EU residents' data by organizations. The GDPR establishes steep financial penalties for organizations that do not comply. For example, under the GDPR, the Luxembourg National Commission for Data Protection (CNPD) has imposed, on July 16, 2021, a record fine of €746 million on Amazon for mis-processing personal data,[1] and the French privacy regulatory authority (Commission Nationale de l'Informatique et des Libertés, or CNIL) fined Google €50 million on January 21, 2019 for "lack of transparency, unsatisfactory information and lack of valid consent for the personalization of advertising."[2]

The requirements introduced by the GDPR affect markets that rely on personal

[1] `https://www.sec.gov/ix?doc=/Archives/edgar/data/0001018724/000101872421000020/amzn-20210630.htm`, pp.13

[2] `https://www.cnil.fr/fr/la-formation-restreinte-de-la-cnil-prononce-une-sanction-de-50-millions-deuros-lencontre-de-la`.

data and profiling. Given how widespread personal data collection and use have become across different sectors, and given GDPR's restrictions on such collection, the aggregate economic impact of the GDPR was predicted, in early industry reports, to be negative and substantial. A 2013 Deloitte impact assessment report suggested that the potential economic impact of the GDPR (Deloitte, 2013) could amount to a loss of around 2.8 million jobs and a reduction of European GDP by around 1.34% (corresponding to around €173 billion). 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.[3] The imposition of limitations on the ability to collect and use data could have a negative impact on the effectiveness of online advertising campaigns (Goldfarb and Tucker, 2011). Furthermore, the GDPR may also impact the composition of the online advertising industry: regulation could impose comparably greater constraints and risks on small and medium-sized online advertising firms, leading to a further concentration of an industry where dominant players already have substantive power to define how the market operates and how benefits are allocated (Johnson and Shriver, 2019). This could in turn expose users, as well as firms upstream and downstream from advertising platforms, to other types of harms, such as monopolistic behavior.

Reductions in advertising effectiveness, spending, and competition were ultimately expected, in turn, to negatively affect publishers. Since advertising is a major revenue component for digital goods producers (Lambrecht *et al.*, 2014), constraints on online tracking and consumer data gathering may threaten the subsistence of free online content and services. Prior to the enactment of the GDPR, claims by both industry groups and think tanks asserted such hypothesis. In a 2015 report by IHS Technology, the

[3] For instance, in a recent study (which did not focus on GDPR, but on the impact of reductions in consumer tracking), Johnson *et al.* (2020) found that reductions in the ability to target advertising (through industry self-regulatory initiatives allows American consumers to opt-out from tracking) 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.

CEO of the Interactive Advertising Bureau of Europe, Townsend Feehan, suggested that overly burdensome privacy regulation 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 made stronger claims, stating, "[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) which is not in the interests of consumers" (Castro, 2010). Both sources capture the sentiment that overbearing privacy regulation could negatively impact publishers, resulting in a reduction in the availability of content or a degradation of its quality (Goldberg *et al.*, 2019).

Despite the numerous claims and predictions about the potential effects of the GDPR on the profitability of ad-supported content providers, the evidence is limited and contradictory. For instance, anecdotal evidence suggests that at least some online publishers that reduced their use of behaviorally targeted advertising in the EU, post-GDPR, have continued to enjoy stable advertising revenue (Davies, 2019). Additionally, it is unclear how consistently different companies (including, but not limited to, online publishers) interpreted and applied the regulation. Small firms could find implementing the GDPR costly, and their size may not justify the investment necessary to use personal data in a compliant way. Larger technology firms may exploit data protection to achieve competitive advantage (for instance, Microsoft and Apple declared before the enforcement of the GDPR that they would voluntarily implement GDPR protections worldwide (Brill, 2018; Phelan, 2018)). Thus, the ultimate implications of the GDPR on the ad-supported publishing ecosystem may be more nuanced than the negative scenarios being proposed by the online advertising industry. To date, no empirical study has tested the relationship between the regulation and websites' ability to provide high quality content to visitors have. Understanding the impact of the GDPR on online content providers (including, but not limited to, news websites) and their visitors, and specifically websites' ability to provide content (including free content), is the focus of this manuscript.

We analyze data for 909 content providers—news and media websites in both the European Union (EU) and the United States (US). We regularly mine website-level data, before and after the enactment of the GDPR, browsing each website from both EU and US IP addresses. The data collected spans 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). This longitudinal panel captures how EU and US websites responded to the GDPR over time, including how they interacted over time with visitors and how they managed the collection of visitors' information. We refer to these variables as *technical variables* in our analysis (Section 4.2). In addition, we use multiple available sources to capture how websites changed their content offerings over time, and how visitors reacted to websites' content over time. We refer to these variables as downstream outcomes in our analysis (Section 4.2).

We focus on content providers as they consist, predominantly, of websites that rely heavily on online advertising. We consider and compare both EU and US websites as the GDPR applies to every European organization (regardless of country of origin of the organization's customers/users), but provides protection to every European resident regardless of whether the organization providing the service is based in Europe or not. That means that websites located in the US should, in principle, comply with the GDPR when interacting with EU visitors. However, the impact of the regulation might be expected to be more evident for European websites, whereas non-European websites (including US websites), whose user bases are largely non-European, may be only marginally affected (Section 3).

In our empirical analysis, we distinguish between ecosystem effects (affecting all EU-based websites relative to US-based websites) and website-level effects (which depend on individual websites' specific responses to the GDPR). We first use a difference-in-differences approach to estimate how the regulation impacts EU websites, compared to US websites, regardless of whether the individual websites do respond and the type of response adopted. To take into consideration that not every website explicitly responds to the regulation, we complement the initial analysis with an Instrumental Variable approach (local average treatment effect or LATE) which allows us to estimate the effect of the regulation for websites that do decide to respond. Finally, in order to investigate the impact of specific responses adopted by websites (website-level effects), we use a look ahead matching analysis. This analysis 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 on outcomes.

The longitudinal nature of our data, and the fact that our data include both websites' responses and downstream outcomes, allow us to paint a rich picture of the evolution of the online publishing ecosystem post-GDPR, and to consider the economic impact of the GDPR in light of websites' different choices of response to the regulation over time. We find 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, and also 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 cookies and visitors tracking among both EU and US websites following the enactment of the GDPR. However, those initial reductions are followed, several months after the enactment of the regulation, by a reversal of the trend and an uptick in tracking among EU websites. Websites that receive a significant proportion of traffic from EU visitors either do not implement measurable changes (for instance, they invoke "legitimate business interest" in the collection of visitors' data), or adjust their responses over time. Websites with the stronger and longer-lasting responses to the GDPR (such as curtailing tracking over an extend period of time), instead, are those that receive only a small fraction of traffic from EU visitors, and thus do not rely on the latter for economic success. Our various econometric specifications are consistent in failing to reject the null hypothesis of no significant differences in downstream economic outcomes for EU and US websites. While we find a small reduction (-0.09) in average number of page views per user in EU websites relative to US websites, we find no statistically significant impact of the regulation on EU websites' ability to provide content, on the amount of visitors' traffic they receive, and on visitors' social media engagement with new content. In short, the results suggest that websites that did respond more strongly to the GDPR were those not likely, in fact, to be affected by such response; whereas websites that did rely in great part on EU visitors found, over time, ways to avoid being negatively affected by the regulation.

2 Literature Review

This paper builds upon and contributes to three strands of literature: the literature on the economics of privacy (and, in particular, the economic impact of privacy regulations); the growing body of economic and non-economic work on the impact of the GDPR; and the economic literature on the online advertising industry and the media industry, including content providers and online publishers.

Privacy regulation and economic outcomes. The economics of privacy literature investigates the trade-offs associated with the revelation or protection of personal information (Acquisti *et al.*, 2016). Within this literature, an important strand of work has focused on the impact of privacy regulations. Policy interventions that regulate the collection or usage of consumer data tend to be aimed at protecting individuals' privacy, but may have nuanced and unpredictable consequences for innovation, market structure, and the economic welfare of different stakeholders.

For instance, Goldfarb and Tucker (2012) argue that privacy regulation might affect the extent and direction of data-based innovation. Empirical works have showed that the impact of privacy regulation can be heterogeneous and context specific. For instance, in the health care domain, Miller and Tucker (2009) have found that privacy legislation primarily can reduce demand for electronic medical records (EMR) via a suppression of network effects; whereas Adjerid *et al.* (2015) find that, although privacy regulation can result in a reduction in health information exchanges (HIE)'s operational effectiveness, if the right privacy incentives are provided to patients, regulation can also have a *positive* impact on the development and adoption of HIEs.

The online advertising market is a conspicuous candidate for the study of how limits imposed on the type or amount of data that can be collected and used may affect an industry reliant on these data. In online advertising, ads are often targeted to individuals based on information tracked and collected online. Personalized (targeted) ads are likely to be more effective than non-targeted ones (Evans, 2009). Goldfarb and Tucker (2010) empirically investigated how the 2002 EU Privacy and Electronic Communications Directive, which restricted advertisers' ability to collect data on users, affected advertising effectiveness captured by hypothetical purchasing intentions. Their results show that after the regulation, certain types of display advertising were less effective relative to display advertising in other countries. Accordingly, the advertising industry has been quick to complain that restrictions on the ability to collect and use consumer data for targeted advertising may be harmful to both advertising companies 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 more broadly (and the GDPR specifically) and websites' ability to provide content has not been vetted in empirical research.

Economic studies of the GDPR. Within the large body of work on privacy regulation, our paper builds upon the significant wave of empirical studies on the impact of the GDPR. One of the very first studies in this stream by Jia *et al.* (2021) investigated the impact of the GDPR on investments in EU emerging technologies. The authors found that the regulation led to a decrease in such investments for EU companies, compared to US organizations. Similarly, Goldberg *et al.* (2021) examined the effect of the GDPR on European web traffic and e-commerce sales and found that *recorded* page-views and recorded revenues both fall by about 12%; nevertheless, the authors estimate that the *real* effect of the GDPR is lower and provide a conservative estimate of 0.4%, for the reduction in page-views, and of 0.6%, for the reduction in e-commerce revenues.

The possibility that the economic impact of the GDPR may be more nuanced than what industry estimates suggested arises from theoretical work. Lefouili and Toh (2018) argue that the effect of the GDPR on investments may be mixed: in a fully covered market, regulating information might reduce investments and yet may be socially desirable when information and quality are not strong complements. Choi *et al.* (2019) investigated consumers' privacy choices with a model in which consumers are required to consent to the collection of their data and consumers are fully aware of the consequences of giving such consent. They found that information externalities and coordination failures among users are drivers of excessive loss of privacy. The possibility of highly nuanced and contextual effects of the GDPR emerging from the theoretical literature is consistent with some empirical studies. Zhuo *et al.* (2021) measure the impact of the GDPR on interconnection agreements between EU network providers with those outside the EU. While 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 in the short-run.

Recent work has also examined how the GDPR has shaped the advertising market. In a theoretical paper, Sharma *et al.* (2019) argued that burdens imposed by regulations such as the GDPR would negatively impact the revenues derived from ad networks for smaller publishers more than larger publishers. In an empirical investigation, Peukert *et al.* (2020) found a decrease in connections to third-party websites among $110,000$ websites following the introduction of the GDPR. The authors observed increased concentration among web technology providers as the market share for small firms decreases, while large firms such as Google see significant increases. Johnson and Shriver (2019) found a similar increase in concentration among web technology vendors from a sample of $27,000$ websites.

The impact of the GDPR on websites' interface features (including consent mechanisms, and visitors' reactions to them) has also been investigated, in particular within economics and computer science. Shorty 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) browsed websites from EU and US visitor addresses and 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. In the same vein, Urban *et al.* (2020) found that a particular type of cookie—syncing cookies, which allow the exchange of users' information between online advertising actors such as Ad networks and Ad exchanges—decreased across more than 2.6 million websites by approximately 40% 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. In a related longitudinal study, Sørensen and Kosta (2019) assessed the effect of the GDPR on the presence of third parties on EU websites. While they found that the number of third parties did slightly decline after the GDPR, they ultimately concluded that the GDPR may not necessarily be responsible for that effect. Degeling *et al.* (2019) investigated online websites' compliance with the data collection requirements imposed by the GDPR. They found that while most websites adjusted their privacy policies and implemented consent mechanisms in the months immediately following GDPR enforcement, some had not complied and did not provide users with means to meaningfully consent to tracking. In an empirical investigation of intermediaries in the online travel industry, Aridor *et al.* (2020) found that the total number of consumers observed by the intermediary decreased by 12.5% after the GDPR, suggesting that a significant number of consumers decided to opt-out. The authors also found that the remaining set of consumers who decided to not opt-out were more persistently identifiable. Finally, they observed a drop in ad interactions across their data-set, along with an increase by advertisers in the average bids for the remaining observable consumers, leading to a smaller overall decline in revenue. Sanchez-Rola *et al.* (2019) investigated the use of opt-out options by users and 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 and found that many elements can be used to nudge users to accept tracking. In an empirical study, Godinho de Matos and Adjerid (2021) found that user opt-in for the disclosure of different data types increase if GDPR-compliant consent was used. As our longitudinal dataset includes both websites' responses and downstream outcomes, we can document not just the evolution over time of content providers' responses to the regulation, but also the downstream impact of regulation and websites' responses on the latter's outcomes.

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; by so doing, they aim at attracting more advertisers willing to buy ad space on their websites, targeted to specific audiences (Anderson and Gabszewicz, 2006). Empirically, Monic and Feng (2013) investigated changes in the quality of blogs' posts after the implementation of ad-supported business models. They found that the quality of blogs' posts tended 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, in turn, increases the 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 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 Theoretical Framework

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). This scope is uniquely expansive in both the types of data it covers and its territorial reach. The GDPR defines 'personal data' to be any data that relates to 'an identified or identifiable natural person' (GDPR Article 4(1)). Due to the way personal data is used to facilitate behavioral advertising, industry groups and EU governmental bodies have considered that data to be within the scope of the regulation (UK Information Commissioner's Office (2019) International Association of Privacy Professionals (2020)). Since the GDPR is extraterritorial in its scope, applying to any entity that handles the personal data of EU data subjects independent of location (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.

Article 6 of the GDPR provides the primary mechanism by which the behavioral advertising practices of websites would be impacted. It establishes six lawful bases for data collection. Any website that collects the personal data of EU data subjects (acting as a 'data controller') must justify that collection under one of the six bases. In the context of online content providers, two of these 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 it is being collected (GDPR Article 6(1)(a)). For example, when a user browses a website owned by a publisher, the publisher must request the user's explicit permission to allow cookies to be set on the user's machine and, if so, whether she would also allow tracking cookies by third parties. This differs from the pre-GDPR *de facto* standards for most websites across the world: in absence of regulatory obligations, websites typically track users' behaviors by default, in some cases merely informing users that they implicitly consent to tracking by virtue of accessing the website. In addition to requiring that organizations provide a lawful basis for data collection and processing, the GDPR establishes steep financial penalties for organizations that do not comply. Both the fine imposed on Google by the French privacy regulatory authority (CNIL) (see Section 1) for lack of valid consent for the personalization of advertising, and advice such as that released by the UK Information Commissioner's Office,[4] suggest that consent has emerged as one of the primary basis for enabling data processing for behavioral advertising under the GDPR.

Under the second possible justification, data controllers (such as websites) can collect and process data if it is necessary 'for the purposes of the legitimate interests

[4] `https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf`

pursued by the controller' (GDPR Article 6(1)(f)). In the case where websites use legitimate interest to justify data collection, these interests must be communicated to the data subject (GDPR Article 14(2)(b)). Typically, this is achieved by including verbiage referring to legitimate interest on the website's privacy policy.

In the following subsections, we first establish why the requirements of the GDPR may result in downstream economic effects for publishers. We then break these effects apart into 'ecosystem effects' and 'website-level effects' and describe each in detail. Throughout, we propose three economic expectations that describe how the GDPR may impact the market of online content providers.

3.1 Downstream Economic Impact

Article 6 of the GDPR includes requirements that, on theoretical grounds, may affect downstream outcomes for websites generating some or most of their revenues through advertising. First, the restrictions introduced by the GDPR may reduce the extent to which website visitors are tracked. The reduction in tracking may be due to a number of factors and manifest itself in a number of ways, including: through the decision by data controllers (such as websites) to reduce or altogether stop the tracking of EU visitors; through websites' decision to block altogether traffic from EU visitors (in order to avoid potential fines associated with GDPR violations); or through the adoption of consent mechanisms, which present visitors with options concerning the usage of their data, and therefore may increase the portion of visitors who opt-out from tracking and targeting relative to the status quo pre-GDPR. We note that the reduction in tracking may be associated with decisions made by (and which may, in turn, affect) both individual websites and the online advertising/publishing ecosystem as a whole (for instance, data intermediaries such as advertising networks). In Subsections 3.1.1 and 3.1.2 below we discuss the subtle differences in the way decisions by different stakeholders may affect websites' downstream outcomes—something our empirical analysis attempts to account for.

A reduction in the ability to track users will, in turn, adversely affect websites' ability to target them with personalized ads when they visit. The availability of personal data makes online advertising more efficient and more accountable, in the sense that advertisers can monitor advertising performance and effectiveness through quantitative metrics (Johnson *et al.*, 2017). This creates a self-reinforcing cycle that increases the efficiency of online advertising. Vice versa, curtailing the collection of personal data and the tracking of online users may negatively affect profiling (Goldfarb and Tucker, 2010).

A reduction in the ability to personalize ads may, in turn, negatively affect ads value. Ads that are tailored to visitors' preferences are more valuable and personal information increases targeting efficiency (Tucker, 2012). Besides allowing for more granular targeting, online advertising has significant cost advantages compared to offline advertising (Goldfarb, 2014). Non-targeted impressions may therefore receive lower bid prices 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). In turn, and 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 overall revenues of content providers may decrease (Lambrecht *et al.*, 2014).

Finally, revenue reduction may impact content provision. 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 ultimately not be able to sustain the quantity and quality of output (content) they generated before the regulatory shock Downes (2018).

While we cannot directly capture changes in revenues for a sufficiently large number of websites, we can capture metrics correlated with websites' ability to provide quality content. The metrics include variables used before in related literature (Shiller *et al.*, 2018; Gallea and Rohner, 2021; Ferreira *et al.*, 2021), 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. We discuss downstream economic outcomes in Section 4.2.

As noted above, there are subtle but important differences in terms of the process through which choices by various stakeholders (individual websites vs. other agents in the online advertising/publishing ecosystem) may, in theory, disparately affect websites' downstream outcomes. In the rest of this section, we distinguish between "ecosystem" effects and "website-level" effects. Ecosystem and website-level effects may operate both separately and in combination to affect downstream outcomes. In section 6 we discuss how our empirical strategies attempt to detect and separate these various possible effects.

3.1.1 Ecosystem Effects

By ecosystem effects we refer to the process through which the responses to the GDPR by all the different stakeholders in the online advertising/publishing ecosystem (such as individual websites, ad networks, and so forth) may *collectively* affect the aggregate availability of consumer data within that ecosystem, and how such changes in data availability may, in turn, affect the ability of *individual* websites to target their respective visitors with behavioral ads—*regardless* of that specific website's own GDPR response. As more stakeholders respond to the GDPR by limiting data collection (for instance, by adopting consent mechanisms that facilitate users' opt-outs, or by reducing tracking altogether), less personal information may become available in the ecosystem, and fewer individuals may be precisely profiled for behavioral advertising when they visit any given website in that ecosystem. This, in turn, would affect that website's revenues, and ultimately its content—again, independently of its own distinct GDPR response—through the chain of effects discussed above at the start of Section 3.1.

Put in other terms, ecosystem effects capture downstream economic impacts of the GDPR that arise not as direct consequence of a given website's specific GDPR response, but indirectly, through changes in the overall availability of consumer data across the entire advertising and publishing ecosystem. For instance, following the enactment of the GDPR, the profiles of existing users within ad-tech companies or ad platforms may become less accurate or less up to date, as data becomes more sparse. In turn, the broader advertising ecosystem may see lower bids in ad auctions for users with less detailed profiles (Goldfarb and Tucker, 2011). In addition, the amount of behavioral data associated with a random visitor to a specific website may, on average, decrease. In both cases, we may expect a reduction of websites' advertising revenues.

It stands to reason that the revenue loss due to this ecosystem effect—and therefore the resulting reduction in websites' ability to provide quality content—should not be expected to be the same for EU websites relative to US websites. First, while EU-based data controllers (such as websites and advertising networks) are required to comply with GDPR rules for *all* users and visitors, this is not the case for data controllers based outside the EU, which may choose to apply GDPR-compliant practices only to the share of visitors originating from the EU. We expect this to reduce the overall availability of user data within the EU-based advertising ecosystem more intensely than any reduction within the US-based advertising ecosystem. In addition, we expect EU-based websites to receive a higher percentage of traffic from EU-based visitors (that is, visitors from a EU IP address) than US-based websites; therefore, we expect a higher proportion of traffic to EU-based websites to be less precisely trackable—and therefore less precisely targetable—than traffic to US-based websites. In short, we expect the magnitude of ecosystem effects on website-level downstream outcomes to be moderated both by where websites are based and by the share of each website's traffic that originates from the EU, independently of a given website's GDPR response. Specifically, we expect EU websites, on average, to be more negatively affected by ecosystem effects 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 addition, we expect websites with a higher percentage of EU-based traffic to be more affected than websites with lower percentage of EU traffic, at parity of type of GDPR response by the website. We expect this to be the case regardless of the geographical location of the website: in other words, if both a EU-based website and a US-based website responded to the enactment of the GDPR by similarly reducing tracking and targeting of EU visitors, we would expect the overall impact on downstream outcomes of their response to be larger for the EU-based website, because (in this example) the website with a higher percentage of EU traffic would be curtailing tracking and targeting of a higher percentage of its traffic, relative to the website with a lower percentage of EU traffic.

We test these hypothesis, and attempt to disentangle these ecosystem effects, in Section 6. In Section 4 we discuss nuances associated with identifying EU and US websites based on domain, traffic patterns, and headquarter location.

3.1.2 Website-Level Effects

By website-level effects we refer, instead, 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 (this information may be shared with and sold to other players in the the rest of the ad-tech ecosystem) 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' response 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. We define five categories of website-level responses, and discuss them in the rest of this subsection, starting with arguably the stronger or more aggressive responses (by which we refer to responses highly likely to curtail the websites' access to visitors' data), and ending with, arguably, the most lax.

First, faced with the compliance burden imposed by the GDPR, websites—especially those based outside the EU—have the option of exiting the data and advertising market 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 redirect their former visitors to other websites (if substitutes exist), 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 unilaterally reducing or halting the tracking and targeting of EU visitors while still allowing them to browse their content. We expect this response to also negatively affect a website's advertising revenues, although arguably not as intensely as blocking EU visitors from accessing the website. We will 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. Such implementations of consent mechanisms may, too, 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 (see Section 5.3). Websites which implement consent dialogs that make it easier for users to deny consent for tracking (such as dialogs that only require a single-step to reject tracking) may experience 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 which require multiple steps to reject tracking). We will refer to this response as *"Consent Mechanism"*

Fourth, websites may attempt to minimize the impact of having to implement a consent mechanisms by instituting 'cookie walls.' Cookie walls force users to first 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 that do not wish to be tracked may react to the appearance of a cookie wall by turning away from the website altogether. While 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 (see Section 5). We will 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.[5] Other websites may simply continue to comply with older EU privacy directives, merely displaying 'cookie notices' which often appear as banners at the bottom of websites.[6] Functionally, the effects on tracking and targeting for the websites invoking GDPR legitimate interest and websites not even bothering doing so may be similar: either way, these websites do not engage in changes that are likely to affect the trackability of their visitors (if anything, they may even benefit 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, we expect this category of websites to experience the mildest effect on revenues and thus on downstream outcomes. For our analysis, as all these responses are unlikely to have a significant impact we group them together and refer to them as *"No Response or Legitimate Interest"*.

Note that it is possible for a website to adopt more than one response at the same time (for example, we found in our data websites implementing a consent mechanism while still invoking legitimate interest), as well as different responses at different moments in time. In these cases, for the purposes of our analysis below, we assign a website to the most frequently adopted stronger response, which we define as the response adopted by a website in the majority of the waves after the GDPR that constitute our sample (Section 4.2).

[5] The legality of this justification for tracking is contested, and the compliance risk 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. Earlier this year, IAB Europe published guidance stating that legitimate interest cannot be used as a basis for setting tracking cookies IAB Europe (2021).

[6] 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.

To summarize, we expect the magnitude of website-level effects on downstream outcomes to vary based on how websites choose to respond. Some responses may lead to stronger negative impacts on tracking and targeting, which may result in stronger negative impacts on downstream outcomes. We expect that websites which respond to the GDPR by blocking EU users or curtailing tracking and targeting to experience the strongest effects, followed by websites implementing consent mechanisms with explicit options to reject tracking with a single-step, followed by websites implementing consent mechanisms designed to nudge visitors towards opt-in (for instance, those that requires multiple steps to reject tracking), followed by websites that do not actively respond to GDPR and/or claim legitimate interest. Of course, how a website chooses to respond will be influenced by factors including the percentage of EU traffic they receive. For instance, websites that choose to block EU visitors are likely those that only receive a negligible share of traffic from EU visitors. We account for this endogeneity in our empirical analysis (Section 6).

4 Study Design

In this section we discuss websites sampling strategies (Section 4.1) and metrics periodically captured for each of the websites in the sample (Section 4.2).

4.1 Websites Sample Selection

We constructed a longitudinal panel of websites located in the US and in several EU countries (Germany, France, UK, Italy, Spain, and the Netherlands).[7] While the panel includes multiple categories of websites, in this manuscript we focus exclusively on

[7] The country of a website was determined by the location of its headquarters as reported by SimilarWeb. When this information was not available (for about 26% of websites), we defined the country of a website by using the website's top-level domain country of origin. In the case where neither the country nor the top level domain are available, we assigned the website to the country where most visitors originated from.

content providers (publishers), such as news websites and online magazines, due to their reliance on online tracking and behavioral targeted advertising for revenues.

The panel includes both top-ranked and long-tail websites, in order to compare the impact of the GDPR on both major and minor content providers. We used 2018 Alexa data to identify top ranked websites. At the time, Alexa data provided the top 500 websites from various geographical areas and five content categories (News, Sports, Society, Health, and Games).[8]

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 the 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 dataset with the top 500 websites for various top-level domains, such as *.au, .de, .fr, .uk, .it, .es, .nl, .com, .net* and *.us*.

We complemented highly ranked websites with a random sample of websites ranked between 200,000 and 1 million. 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.

The resulting sample included 11,254 websites. We eliminated websites that only got a minor fraction of their visitors from EU countries or the US, despite the fact that they were among the most popular websites in one of our countries of interest.[9] Finally, we verified the content categories of the remaining websites. We noticed that the content categories provided by Alexa were sometimes inconsistent. Thus, we obtained categories

[8] In September 17, 2020 Alexa discontinued its categories ranking. This does not affect our data, as it happened after our sample was created.

[9] For example, the Russian shopping website *avito.ru* was the 52nd most visited website in The Netherlands in May 2019. However, visitors from The Netherlands account for less than 2% of all avito.ru's visitors, while visitors from Russia account for roughly 85%. Therefore, although the website is popular in at least one EU country, it would be unreasonable to assume it will significantly change its behavior due to a European regulation considering that it is a Russian website that gets most of its visitors from Russia.

information using SimilarWeb,[10] which provided a more robust categorization. We excluded from the sample all websites categorized as providing adult content, or not assigned to any category.

The resulting set contains 5,474 websites. Of them, 909 were in the News and Media category. This sample is the focus of our analysis.

4.2 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 4.2.1). The goal of these metrics 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 4.2.2). 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 spans a period of time of at least 19 months for technical variable metrics (from April 2018 to November 2019), and 31 months for downstream economic outcomes (from April 2017 to November 2019).

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

4.2.1 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 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. We analyze these data to extract technical variables that capture websites' behaviors (including tracking, privacy notices, advertising choices, consent mechanisms) and changes in behaviors in response to the GDPR. We discuss these variables below.

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 3rd party cookies. The variable *1st Party Cookies* measures the cookies which 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. In our analysis, we rely on a drop to zero in either advertising or tracking cookies to identify when a website decides to respond to the GDPR by stopping the tracking of EU visitors.

Advertising Content Length: 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.[11] The variable *Ads Length (KB)* 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*[12].

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

[11] 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 by recognizing the advertising tags of the main ad servers and advertising networks. We cross-referenced the data we collected from OpenWPM with these filtering lists (blocklists). (List inside ad-blocker add-on to block unwanted content like advertising.) We rely on two blocklists: Adblock Plus (`https://adblockplus.org/fr/subscriptions`. Last retrieved, February 2020) and Disconnect (a free extension for the web browser responsible for blocking trackers from web pages that the user visits: `https://disconnect.me/`), which establishes identification and classification rules for advertising and tracking entities.

[12] 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 measured the length of advertising instead of hiding it. The list is available at: `easylist.to`

example Figure 12a and Figure 12b). By contrast, cookie banners inform users about cookies, but do not provide them with a way to reject tracking (see Figure 14). 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 13. Finally, we are able to identify US websites which decided to block EU consumers (visitors) by identifying a static page shown to EU visitors informing them that the website is unavailable (see Figure 11). For each of the responses so identified, we create a dummy variable which 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.

4.2.2 Downstream Economic 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).[13] 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 which 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

[13] `gdeltproject.org`

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 content quality. The underlying premise is that, were the quality of the content provided by the website to decrease, users may 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).[14] We use Alexa's *Rank*, a measure of a website's popularity. It is calculated as a combined measure of page views and unique visitors (reach). We use Alexa's *Reach Per Million* as a measure of the number of (unique) users visiting a website. [15] 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.* (2015) 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 1 presents

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

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

the descriptive statistics of the technical variables and the downstream outcomes, for the overall sample, across all waves.

Table 1: Descriptive Statistics - Overall Sample

	Mean	Std. Dev.	Min	Max	N
	Technical variables				
***Tracking*:**					
1st Party Cookies EU Visitor	11.058	8.008	0.0	52.0	11114
3rd Party Cookies US Visitor	33.269	37.620	0.0	272.0	11114
1st Party Cookies US Visitor	12.942	8.288	0.0	50.0	11175
3rd Party Cookies EU Visitor	53.892	50.060	0.0	351.0	11175
***Advertising*:**					
Ads Length (KB) EU Visitor	576.564	1279.351	0.0	111046.5	11107
Ads Length (KB) US Visitor	695.740	932.966	0.0	28571.3	11172
***Website Level Responses*:**					
Blocking EU Visitor	0.023	0.151	0.0	1.0	21798
Stop EU Tracking	0.118	0.322	0.0	1.0	11114
Consent Mechanism EU Visitor	0.133	0.340	0.0	1.0	21798
Cookie Wall EU Visitor	0.017	0.130	0.0	1.0	21798
Cookie Banner EU Visitor	0.101	0.302	0.0	1.0	21798
***Website Visitors*:**					
Share of EU Visitors	0.430	0.420	0.0	1.0	21798
Share of US Visitors	0.395	0.403	0.0	1.0	21798
	Downstream Outcomes				
***Dependent Variables*:**					
Log GDELT URLs	5.014	1.705	0.0	9.6	17588
Page Views Per User	2.032	0.953	0.6	19.1	21797
Rank	65063.040	100954.740	0.0	1832762.9	21797
Page Views Per Million	14.463	61.499	0.0	1451.4	21797
Reach Per Million	243.692	862.081	0.0	18714.3	21797
FB Average Reaction	110.463	466.708	0.0	12476.9	17588

5 Descriptive Statistics

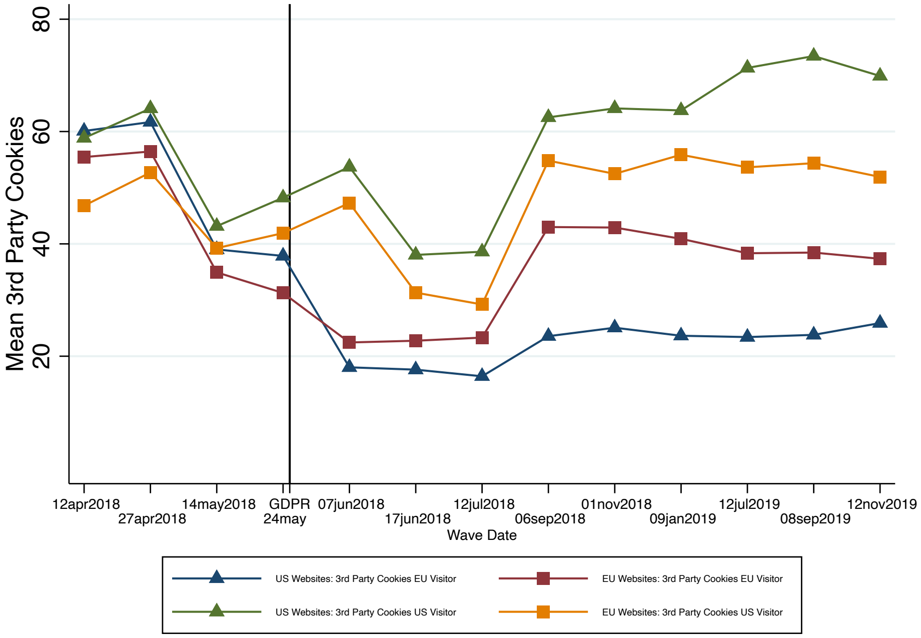
In this section, we leverage the technical variables we collected to describe patterns in website behavior. We first investigate general changes in cookies and advertising patterns; then, we zoom into the five different types of websites responses to the GDPR, as introduced in the previous sections.

5.1 Changes in Cookies and Advertising Patterns

We start by analyzing changes in cookies and advertising patterns for the websites in our sample, before and after the GDPR became effective. We contrast EU vs US based websites, and how the results change if the websites are visited from EU or US based visitors.

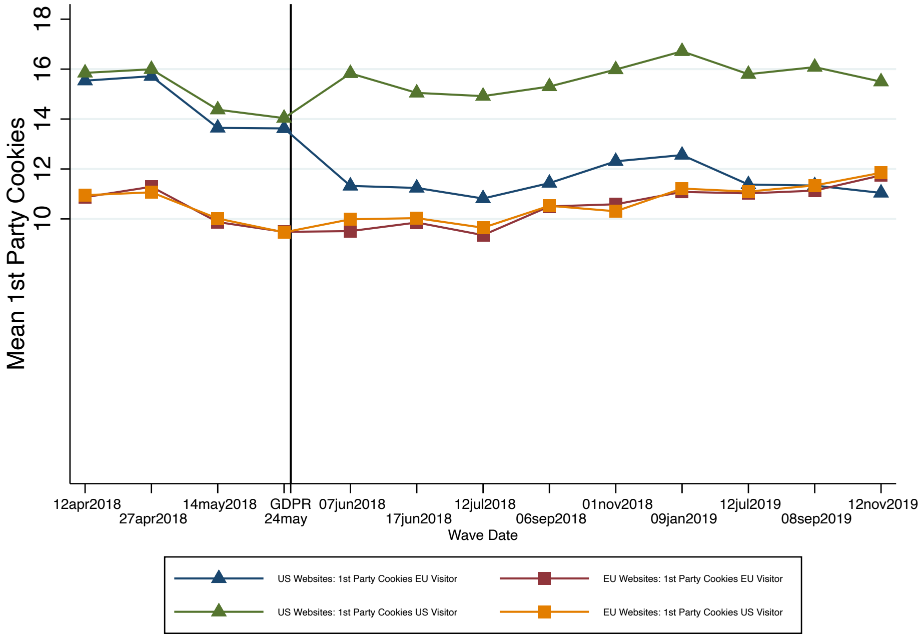
We first consider 3rd party cookies, which are typically used to track users across websites. Our data collection strategy allows us to observe and distinguish four scenarios: how websites based in the US (US websites) treat visitors that originate from the EU (EU visitor) vs from the US (US visitor); and how websites based in the EU (EU websites) treat EU visitors vs US visitors. Figure 1 shows how, before the GDPR, the number of 3rd party cookies used by EU and US websites were similar for both EU and US visitors. Even before the GDPR came into effect, we observe a drop in the number of 3rd party cookies being used in EU/US websites for both EU/US visitors. Right after the GDPR become effective, the sharpest drop happens in US websites for EU visitors, followed by EU websites for EU visitors. However, these drops seem to be short lived, as we observe a rebound in the number of 3rd party cookies set by websites roughly 3 months after the GDPR became effective. Notably, the rebound is not the same for all websites and visitors. US websites continue to set, for EU visitors, a much lower number of 3rd party cookies than before the GDPR. In the case of EU websites visited from the EU, however, the number of 3rd party cookies rebounds to pre-GDPR levels.

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



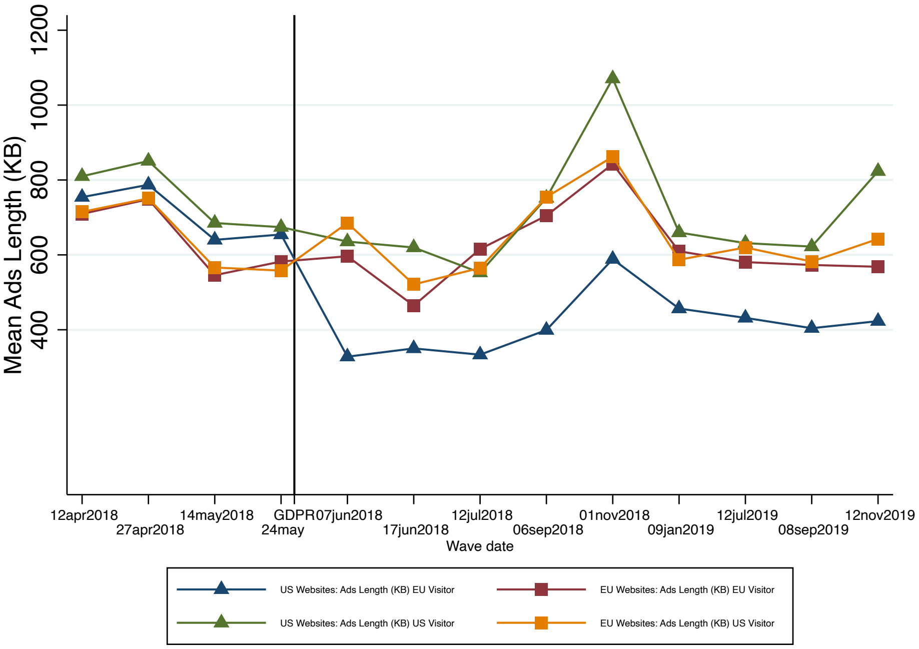
Next, we examine whether the number of 1st party cookies used by websites changed over time. While 3rd party cookies are typically used to track users across websites, 1st party cookies are typically related to particular websites functionalities. For example, a website may use 1st party cookies to remember visitors' login information, products they have browsed, or news articles they have read. However, since 1st party cookies can also be used for advertising purposes, we are interested in examining whether 3rd party cookies are being replaced by 1st party cookies for that purpose. Indeed, such an option was introduced by Facebook in 2018 (Flynn, 2018). Figure 2 suggests that the number of 1st 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. From the figure it is also clear that EU websites seem to set, on average, fewer 1st party cookies than US websites.

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



While 3rd party cookies are typically used by advertising technology firms to track users across websites, 3rd party cookies 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, we explore how advertising length (as defined in section 4.2.1) evolved over time. In Figure 3, we observe interesting changes in advertising length, following the GDPR enactment. EU websites experience a drop right before the GDPR, followed by a number of fluctuations, which, overall, do not seem to signal a significant change in advertising length's levels. For US websites, the response is more nuanced. While over the long-term advertising length for US visitors seems to return to pre-GDPR levels, it remains at a much lower level for EU visitors.

Fig. 3 *Advertising Length EU/US Websites for EU/US Visitors*



Looking at the evolution of technical variables related to cookies and advertising reveals that websites respond significantly to the enactment of the GDPR. EU and US websites respond differently, and the response is influenced by the location of the visitor. US websites drop the level of both 3rd party and 1st party cookies for EU visitors, while cookies for US visitors tend to stay stable. This is also reflected in a drop in the amount of advertising showed to EU visitors. EU websites respond by initially dropping 3rd party cookies for both EU and US visitors; but the level increases again after a few months, in particular for US visitors. The level of 1st party cookies does not seem to change, following the enactment of the GDPR. Similarly, the advertising length experiences some fluctuations, but does not seem to significantly change. In the next subsection we study the prevalence of the 5 stylized responses to GDPR we identified in section 3.1.2 over time and characterize the different websites using them.

5.2 Website-Level Responses to GDPR

Table 2 shows descriptive statistics of pre-GDPR website level characteristics, for the different types of website-level responses we identified in section 3.1.2.[16] In the following subsections, we analyze the features of websites in each response category.

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	81331.78	35102.78	82168.56	28077.95	103426.14	104531.72	23557.45	2193.72	59264.68
	(122952.81)	(74224.19)	(59092.21)	(40600.39)	(74799.50)	(121985.74)	(59560.49)	(2683.45)	(88213.71)
Share of EU Visitors	0.77	0.79	0.95	0.83	0.01	0.04	0.07	0.04	0.03
	(0.29)	(0.23)	(0.05)	(0.20)	(0.01)	(0.05)	(0.12)	(0.01)	(0.05)
Share of US Visitors	0.03	0.04	0.00	0.03	0.91	0.75	0.65	0.69	0.81
	(0.10)	(0.07)	(0.00)	(0.10)	(0.04)	(0.23)	(0.22)	(0.09)	(0.17)
Ads Length (KB) EU Visitor	221.27	817.38	201.80	780.27	1135.40	517.56	681.60	1025.68	797.51
	(390.67)	(802.11)	(362.33)	(651.87)	(715.43)	(553.01)	(547.89)	(651.53)	(649.79)
Cookie Banner EU Visitor	0.11	0.11	0.03	0.11	0.00	0.02	0.02	0.07	0.01
	(0.32)	(0.31)	(0.16)	(0.31)	(0.00)	(0.12)	(0.15)	(0.25)	(0.10)
Privacy									
3rd Party Cookies EU Visitor	10.94	58.09	16.68	55.21	58.91	36.05	60.54	92.16	56.81
	(18.40)	(47.07)	(27.97)	(45.87)	(29.61)	(39.87)	(44.26)	(62.12)	(42.15)
1st Party Cookies EU visitor	6.22	12.22	5.52	11.56	10.29	12.41	16.02	15.95	17.56
	(4.81)	(6.44)	(4.55)	(6.80)	(5.67)	(8.01)	(9.49)	(3.36)	(9.27)
Obs.	1,468	2,805	240	2,445	643	2,456	1,019	75	2,473
Unique websites	98	187	16	163	43	164	68	5	165

5.2.1 Blocks EU Visitors

A number of websites (43) decided to exit the EU market by blocking EU visitors from accessing their websites altogether. The websites in our sample that implement such response are all US based and, in general, the overwhelming majority of their visitors are US visitors. This type of response was quickly implemented after the GDPR, and

[16] As a reminder, "most prevalent response" refers to the response that we observe most frequently for each website across all the waves of data collection.

the share of US websites using this strategy remains fairly constant across all our waves of data.

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 $72,5\%$ of their traffic from US visitors. In terms of traffic, they rank lower than other websites (and therefore receive less traffic). In terms of reliance on advertising, they seem to rely on advertising to a greater extent than other websites. In other words, US websites that block EU visitors are smaller content providers that receive a smaller proportion of traffic from EU sources but rely heavily on advertising. Their response is probably driven by gains they would obtain from the few EU visitors being very small, compared to the potential costs of compliance to the GDPR requirements and/or potential liability.

5.2.2 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 decrease their number of advertising and (tracking) 3rd party cookies to zero (note that we also include in this group websites that, before the GDPR, were not using 3rd party cookies and continue to not do so after the GDPR)[17]. In our sample, 164 out of 445 US websites stop tracking EU visitors, and only 98 out of 464 EU websites do so. Looking at Table 2, we notice that 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 to not track have a considerable share of EU visitors, but their average advertising length is much lower than EU websites that respond in other ways. These patterns suggest that "Stops EU Tracking" is a plausible response for websites that seem not to rely much on advertising.

[17] We define the advertising and tracking cookies using the same method as the Advertising Length, i.e. using the ad blocking lists

5.2.3 Consent Mechanisms

Before the GDPR, we find that almost no US website implemented consent mechanisms (where visitors were given an option to reject tracking cookies), while about 16.8% of EU websites did. Over time, we observe that the presence of consent mechanisms increases sharply for EU visitors for EU and US websites right before the GDPR became effective, and continues to rise until reaching a stable level with about 58.32% of EU websites in our sample using them (see Figure 17 in Appendix C). The 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 advertisements lengths on their websites.

5.2.4 Cookie Wall

The category cookie wall includes websites that implement consent dialogs which force users to consent to tracking in order to access the website's content. This strategy can have unintended effects: if visitors consent to the tracking, the website may not experience changes in its tracking ability; if instead visitors do not wish to be tracked, they may leave the website altogether. Among EU websites, about 3.4% implement a cookie wall. The proportion falls to 1% for US websites. EU websites that fall in this category have a large proportion of EU visitors, but tend to be smaller websites (in terms of 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 have a great reliance on advertising, as suggested by the average advertisements lengths on their websites.

5.2.5 No Response or Legitimate Interest

The last response category includes websites that claim legitimate interest (and therefore continue to collect/use data as before the GDPR) or 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 a portion of the websites that invoke legitimate interest by collecting and analyzing websites' privacy policies. More specifically, we are able to collect privacy policies for about 45% of the observations in this category; among those, about 18.7% include language suggesting the website reliance on legitimate interest.

For the purpose of our analysis, we combine these two types of responses together (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: we do not expect any website-level effect for this group of content providers. If the websites do not actively change anything following the GDPR (either because they are legitimately doing so or because decide to not act) their actions cannot be the driver of any changes in outcomes we may observe. Instead, if they experience any impact from GDPR, this should be attributed to ecosystem-level effects. While these websites didn't adopt any action to curtail tracking, they may still be impacted if there is overall fewer data available about their users because other websites in the ecosystem have reduced their tracking.

EU websites that fall in this category perform better in terms of ranking (have more traffic), compared to the other websites; and still rely heavily on advertising. US websites that fall into this category also rely considerably on advertising but they rank lower compared to US websites that decide to implement a consent mechanism; and they rank higher than US websites that decide to not track or completely block EU visitors.

5.3 Evolution of Website-level Responses over Time

Website responses to the GDPR evolved over time. Figure 4 and Figure 5 summarize the evolution of responses over time for US websites and EU websites 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. Figure 4 shows how responses evolve for EU websites. What we notice is that the number of EU websites identified in the group *No Response or Legitimate Interest* decreases from 178 to 119 websites over the three periods. Most of this drop is explained by an increasing number of websites using consent mechanisms in response to the GDPR (the number increases from 142 to 243). The number of websites that decide to halt EU Tracking altogether seems instead stable over time. A similar pattern is observed for websites that decide to introduce a cookie wall. As for the US websites, Figure 5 illustrates that their responses tend to be more stable over time.

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

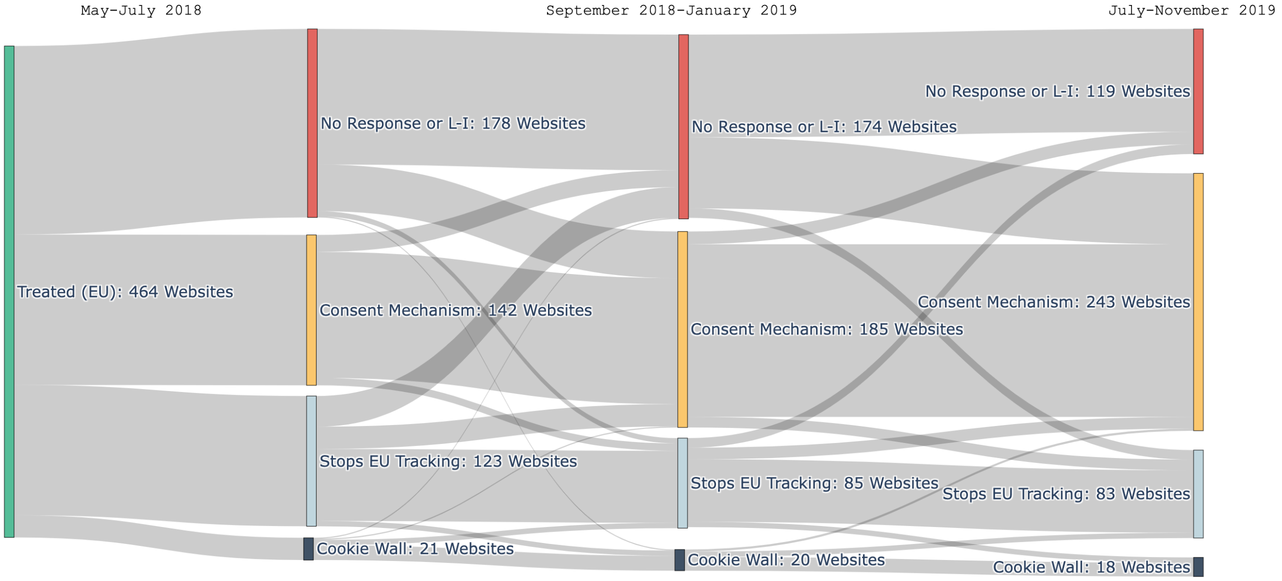
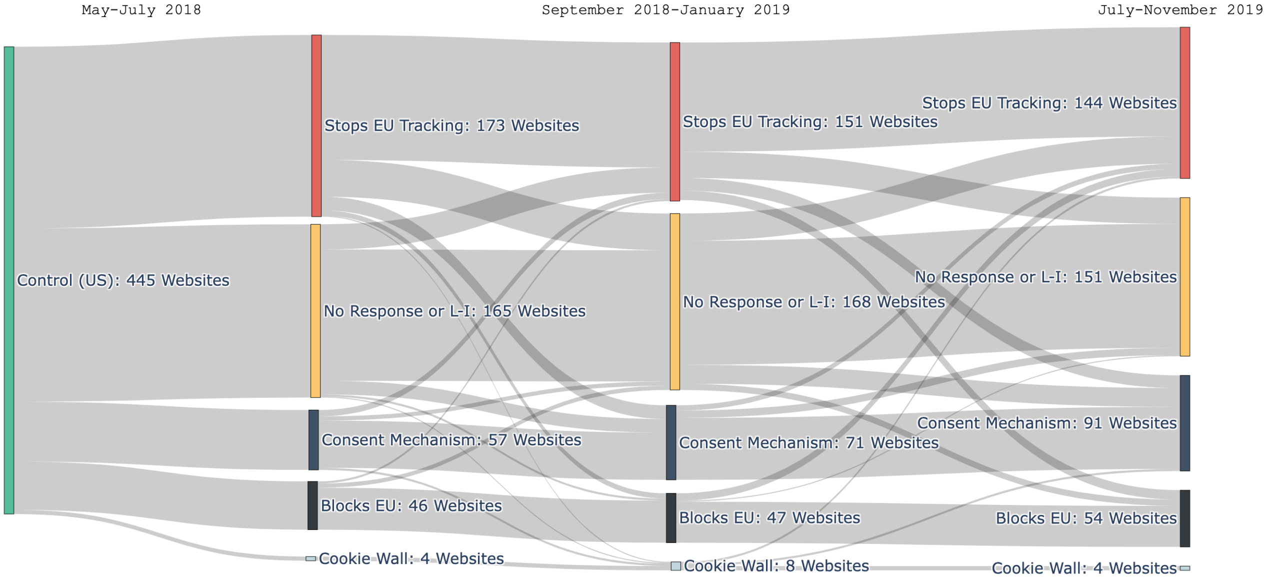


Fig. 5 *Sankey Diagram: Evolutions of US Website-Level Responses over Time*

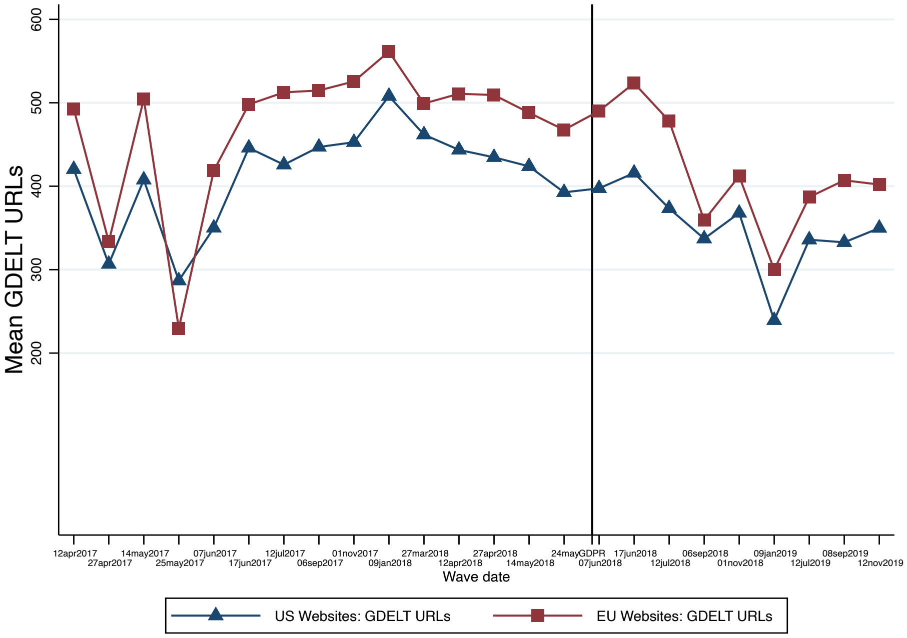


5.4 Changes in Downstream Outcomes

5.4.1 Content Quantity

Next, we focus our attention on websites' downstream outcomes and analyze changes in their trends. 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%, we suppose that the generalized decline is likely either seasonal or due to factors (such as competition from streaming services) other than the GDPR. The number of new URLs starts recovering for both EU and US websites a few months following the enactment of the GDPR.

Fig. 6 *New URLs (GDELT URLs)*



5.4.2 User Engagement

As proxies for content quality, we measure changes in user engagement via commonly used websites' metrics (*Page Views Per User*, *Page Views Per Million*, *Reach Per*

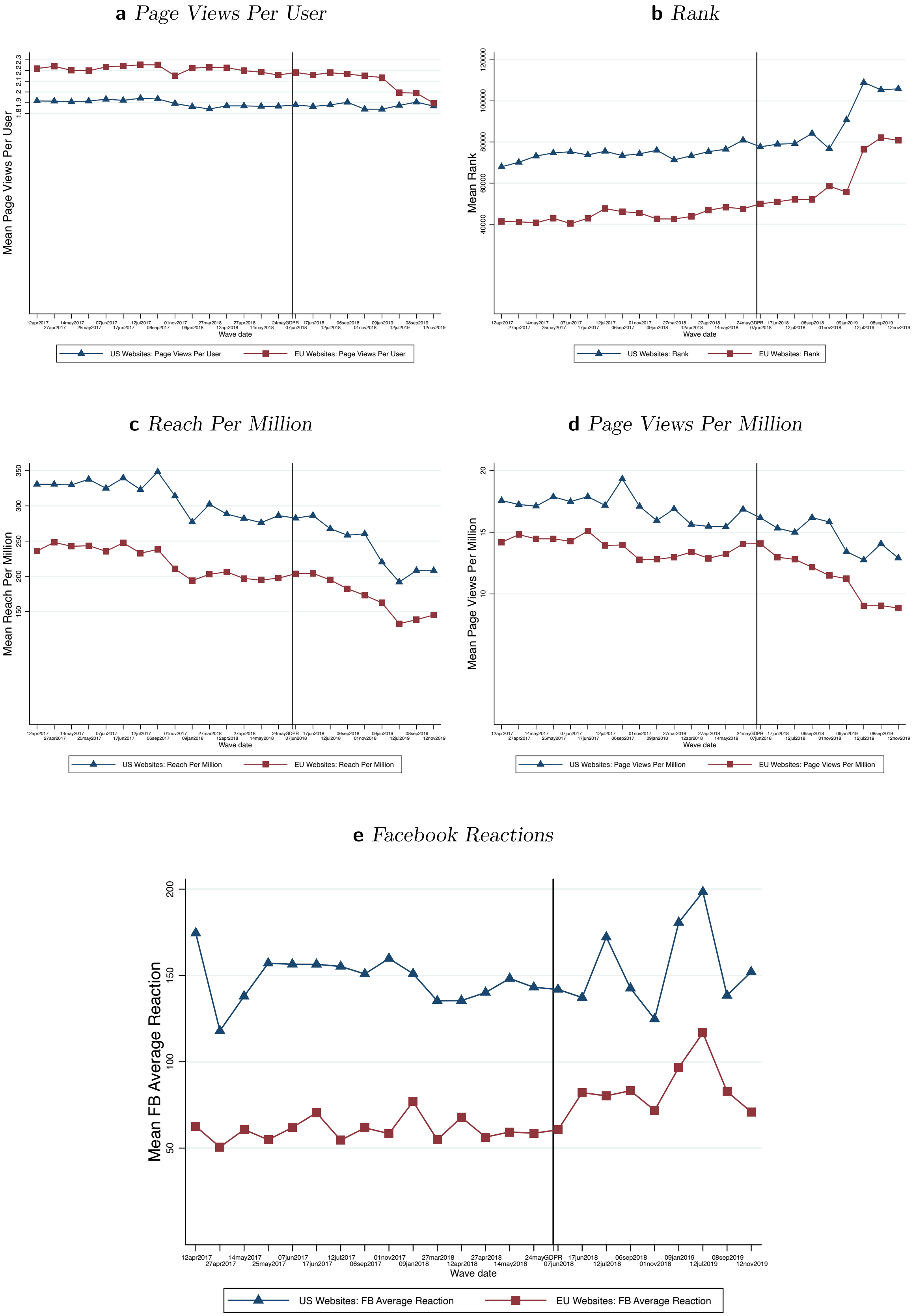
Million, *Rank*) and social media metrics, in specific, the number of Facebook media "reactions" to new content published on the websites in our sample (*FB Average Reactions*).

Figure 7a shows the number of page views per user. While the trend seems stable for US websites, there appears to be a small downward trend for EU websites near the end of the period of observation. A possible explanation for this is that the implementation of consent mechanisms, which we observe is the response adopted by the majority of websites in the EU over the long term, may have a negative effect on users' engagement. This could happen if, for example, consent pop-ups are considered obtrusive and time consuming, and some users turn away when they encounter them. Additionally, we observe, from 7, that both Reach and Page views per Million seem to experience a general decline after GDPR, for both EU and US websites. Nevertheless, reach for EU websites seems to start increasing towards the very end of the period of observation, while page views per million seem to stabilize. These combined patterns could therefore lead to a significant decrease in page views per user for EU websites, when compared to (the stable pattern for) US.

The general declines in reach and page view per million explain the increasing trends in rank (Figure 7b), we observe for both EU and US websites. As a reminder, rank is computed by Alexa as combination of reach and page view per million; an increase in rank suggests that websites are moving towards *lower* rank positions.

Finally, Figure 7e shows, initially, a stable trend for reactions on Facebook, followed by fluctuations in later periods, which are similar for both EU and US websites.

Fig. 7 *User Engagement*



a *Page Views Per User*

b *Rank*

c *Reach Per Million*

d *Page Views Per Million*

e *Facebook Reactions*

5.4.3 Website Survival

A possible unintended consequence of the GDPR may be the interruption of services by content providing websites. Only a very small fraction of websites stopped producing content or completely shut down during the period of observation (around 1%). Based on GDELT data, we find that 4 websites in the EU sample and 6 websites in the US sample were no longer posting new content URLs as of November 2019.

6 Empirical Analysis

6.1 Identification Strategy

The descriptive evidence we have presented so far suggests significant changes in websites' handling of visitors' data following the GDPR, and also nuanced and complex variations in websites' responses. In this section, we attempt to provide an estimate of the impact of the GDPR on content providers, by analyzing the impact on relevant downstream outcomes, and to provide evidence on how the different websites responses described above may affect such outcomes.

Our empirical approach relies on the fact that the GDPR can be thought of as an exogenous shock which, given its scope, should impact (more) certain websites than others (Section 3.1). This makes it a good setting to implement a difference-in-differences approach, where we compare changes in outcome(s) for websites that have been (exogenously) impacted by the GDPR to the changes in outcome(s) for websites that should have not been impacted (or not impacted as much). Nevertheless, the general scope of the GDPR and the ubiquitous nature of the world wide web - where visitors from all over the world can browse and consume content from any website on the Internet - makes the context of the analysis challenging, for a number of reasons.

First, although GDPR applies to all EU citizens and thus affect both EU and US websites, as outlined in the previous sections, we would expect it to affect, foremost, EU websites; and, to a lesser extent, US websites with a considerable proportion of EU visitors. We take this into consideration in our analysis, as explained below (Section 3.1.1).

Second, the GDPR enactment will produce the expected changes if websites do respond to it and implement the relevant changes. In fact, the generality of the GDPR, which leaves different options to the individual websites in terms of how to practically implement the regulation, implies that websites may interpret and respond in various ways; and each type of response may have potentially different implications, both in terms of outcome(s) as well as in terms of the strength of the response (website-level effects; Section 3.1.2). Both the decision to respond to the regulation, as well as the decision about how to respond, are endogenous decisions of the individual websites, which can, therefore, be correlated to websites' observable and unobservable characteristics, adding an additional layer of complexity to our analysis.

Our empirical approach attempts to address the outlined challenges as follows. We start with a difference-in-differences (DID) approach aimed at estimating the overall impact of (the enactment) of the GDPR on the outcomes of interest. We begin by using a definition of treatment and control groups based on the geographical location of websites, where all the EU websites in our sample are considered as treated and all US websites in our sample are considered as controls. The effect estimated by this model is an intention to treat effect, since the estimation includes all the websites subjected to the initial treatment assignment. The estimates we obtain from this model give us a measure of what we referred to as ecosystem effects: we measure the impact of the GDPR for EU websites, relative to US websites, regardless of whether they responded to the GDPR or not, and regardless the type of response potentially implemented. We then 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 (details provided in the following section).

The intention to treat analysis does not take into consideration the fact that not all the websites respond to the enactment of the GDPR. In other words, the enactment of the GDPR can create not only ecosystem effects, but also websites-level effects, determined by whether and how websites respond. Given the challenges described above due to the endogeneity of websites' responses, we tackle this analysis using different strategies.

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, regardless of the type of response. Next, 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. Specifically, we zoom 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). Finally, we leverage the richness of the data we collected to explore the existence of heterogeneity in the estimated effects, based on websites' features.

6.2 Difference-in-differences or Intention To Treat

We start with a simple difference-in-differences (DID) model to tease out potential changes in content quantity and user engagement after the GDPR, for EU websites, relative to US websites. Our framework controls for websites' fixed effects and time-specific fixed effects. The specification of our regressions is as follows:

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

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. $Post\ GDPR \times EU\ Websites_{i,t}$ is equal to 1 if the website i is a EU website 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 based in the EU, compared to US websites.

The results are presented in Table 3. Column (1) presents the results for GDELT URLs that we use as proxy for content quantity. We use a logarithmic transformation of the dependant variable, *Log GDELT URLs*, to take into account the fact that our dependent variables is a count of new URLs. Columns (2) to (6) present the results for *Page Views Per User*, *Reach per million*, *Rank*, *Page Views Per Million* and *FB Average Reactions*, which we use as proxies for user engagement.

Additionally, Table 3 is separated into three panels, each presenting the results for the analysis implemented using different definitions of control and treatment groups. The first panel shows the results for our basic specification, where the treatment group includes all EU based websites while the control group includes all US based websites. In the second panel, the treatment group includes EU based websites and US websites with a share of EU visitors greater than 10%. In the last panel, the treatment group includes EU based websites and US websites with a share of EU visitors greater than 5%. The implementation of these different definitions of treatment and control aims at mitigating concerns due to the fact that the broad application of the GDPR can lead some US websites to respond to it, where the degree of response should be correlated to the amount of EU visitors received by those websites.

The results obtained 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 3-6). Finally, we do not find a negative effect in terms of social media engagement (Facebook reactions, Column 5). We do find a small, negative, and statistically significant effect for *Page Views Per User* (Column 2), suggesting that, after the enactment of the GDPR, EU websites experience an average decrease in the number of pages browsed by their visitors by about 0.09 pages, per user. One possible interpretation is that the reduction in the number of pages visited on a given website 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. Another plausible explanation is that when users visit EU websites, they now encounter consent notices or requests, and in some cases even cookie walls. This may lead viewers to leave the page instead of expressing their consent choices.

In summary, the results suggest that the enactment of the GDPR has not relatively affected the amount of content that EU websites are able to publish, or the degree of average social media engagement and interaction with such content, but may have, to some extent, negatively affected page views per user for EU websites (and US websites with high proportion of EU visitors, relative to US websites (with small proportion of EU visitors). Our expectation of these variables is that they are strongly correlated with content quality, considering that the goal of these websites is to attract and retain viewers.

Table 3: Diff-in-diff Estimations

	Content Quantity	User Engagement				
	(1)	(2)	(3)	(4)	(5)	(6)
	Log GDELT URLs	Page Views Per User	Reach Per Million	Rank	Page Views Per Million	FB Average Reaction
Intention to Treat 1: Treatment group base on EU Websites						
EU Websites × Post GDPR	0.006	-0.092***	19.271	2539.370	-0.149	11.980
	(0.041)	(0.033)	(14.099)	(3322.068)	(0.883)	(15.023)
Constant	5.015***	2.049***	239.996***	64575.950***	14.492***	108.380***
	(0.007)	(0.006)	(2.704)	(637.224)	(0.169)	(2.670)
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.	17577	21797	21797	21797	21797	17577
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	0.004	-0.119***	24.430*	2529.713	-0.048	13.988
	(0.041)	(0.032)	(13.906)	(3326.240)	(0.869)	(14.867)
Constant	5.015***	2.054***	239.107***	64588.247***	14.472***	108.081***
	(0.007)	(0.006)	(2.610)	(624.290)	(0.163)	(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.	17577	21797	21797	21797	21797	17577
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	0.004	-0.114***	19.512	3172.219	-0.134	15.511
	(0.041)	(0.032)	(13.969)	(3323.893)	(0.873)	(14.912)
Constant	5.015***	2.053***	240.006***	64463.728***	14.488***	107.797***
	(0.007)	(0.006)	(2.639)	(627.967)	(0.165)	(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.	17577	21797	21797	21797	21797	17577

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 gives us an estimation of the overall ecosystem effect of the GDPR. In this section, we use an instrumental variable (IV) approach to estimate the effect of the GDPR for the websites that do respond to the regulation, regardless of the type of response. As a consequence, the effect estimated represents a local average treatment effect (LATE) or average effect for those websites which decide to respond to the GDPR. This effect corresponds to a combination of website-level response and ecosystem effects. While the decision to respond to the GDPR is an endogenous decision of the websites, we 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, an approach discussed by Angrist and Imbens (1995). To implement this approach, we need to identify if 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 which is either clearly visible to the visitors and able to induce changes in visitors' behavior and targeting abilities (this includes websites that implement a consent mechanism; that implement a cookie wall or a cookie banner; and websites that block EU users (applies to US websites only); or a response which is clearly detectable and able to induce changes in a website's tracking capability (this includes websites that stop tracking EU visitors).

In our first stage specification, we estimate the probability of a website to respond to the GDPR as function of the enactment of the the GDPR.

$$GDPR\ Response_{i,t} = \alpha_0 + \alpha_1 Post\ GDPR \times EU\ Websites_{i,t} + \omega_t + \mu_i + \zeta_i \qquad (2)$$

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; $Post\ GDPR \times EU\ Websites_{i,t}$ is our instrument; ω_t is a vector of time fixed effects, and μ_i is a vector of website fixed effects and ζ_i is the error. In the second stage, we use the residuals obtained from the first stage to estimate the impact of responding to the GDPR, on our outcomes of interest: *Log GDELT URLs*, *Page Views Per User*, *Reach per Million*, *Rank*, and *FB Average Reactions.*

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

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

Table 4 presents the results of our IV approach. Similarly to the DID analysis, we use three different definitions of treatment and control (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.

Table 4: LATE Estimations

	Log GDELT URLs		Page views per user		Reach per Million		Rank		Page Views Per Million		FB Average Reaction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	GDPR Response	Log GDELT URLs	GDPR Response	Page Views Per User	GDPR Response	Reach Per Million	GDPR Response	Rank	GDPR Response	Page Views Per Million	GDPR Response	FB Average Reaction
LATE 1: Treatment group base on EU Websites												
EU websites × Post GDPR	-0.005		0.020		0.020		0.020		0.020		-0.005	
	(0.021)		(0.019)		(0.019)		(0.019)		(0.019)		(0.021)	
GDPR Response		-0.975		-4.663		949.442		1.37e+05		-8.008		-2490.233
		(9.370)		(4.734)		(1152.051)		(2.08e+05)		(44.518)		(11466.416)
Constant	0.001		0.000		0.000		0.000		0.000		0.001	
	(0.006)		(0.005)		(0.005)		(0.005)		(0.005)		(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		0.688		0.688		0.688		0.688		0.688		0.688
Share of response inside Treatment		0.828		0.828		0.828		0.828		0.828		0.828
Underidentification (LM)		0.050		1.144		1.144		1.144		1.144		0.050
P-value (LM-Stat)		0.822		0.285		0.285		0.285		0.285		0.822
Weak identification		0.050		1.143		1.143		1.143		1.143		0.050
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: Treatment group base on EU Websites + Websites with more than 10% of EU Visitors												
EU Websites and > 10% EU visitors × Post GDPR	-0.005		0.019		0.019		0.019		0.019		-0.005	
	(0.021)		(0.019)		(0.019)		(0.019)		(0.019)		(0.021)	
GDPR Response		-0.751		-6.286		1289.060		1.33e+05		-2.558		-2841.851
		(8.790)		(6.485)		(1483.486)		(2.17e+05)		(45.858)		(12630.292)
Constant	0.001		0.000		0.000		0.000		0.000		0.001	
	(0.006)		(0.005)		(0.005)		(0.005)		(0.005)		(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		0.690		0.690		0.690		0.690		0.690		0.690
Share of response inside Treatment		0.829		0.829		0.829		0.829		0.829		0.829
Underidentification (LM)		0.053		1.039		1.039		1.039		1.039		0.053
P-value (LM-Stat)		0.817		0.308		0.308		0.308		0.308		0.817
Weak identification		0.053		1.038		1.038		1.038		1.038		0.053
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: Treatment group base on EU Websites + Websites with more than 5% of EU Visitors												
EU Websites and > 5% EU visitors × Post GDPR	-0.006		0.018		0.018		0.018		0.018		-0.006	
	(0.021)		(0.019)		(0.019)		(0.019)		(0.019)		(0.021)	
GDPR Response		-0.691		-6.179		1057.286		1.72e+05		-7.244		-2643.287
		(7.305)		(6.571)		(1329.547)		(2.48e+05)		(47.739)		(9909.927)
Constant	0.001		0.000		0.000		0.000		0.000		0.001	
	(0.006)		(0.005)		(0.005)		(0.005)		(0.005)		(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		0.690		0.690		0.690		0.690		0.690		0.690
Share of response inside Treatment		0.828		0.828		0.828		0.828		0.828		0.828
Underidentification (LM)		0.076		0.981		0.981		0.981		0.981		0.076
P-value (LM-Stat)		0.783		0.322		0.322		0.322		0.322		0.783
Weak identification		0.075		0.980		0.980		0.980		0.980		0.075
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

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

We suspect that the extent to which a website is going to respond and the effect of such response may vary with the website's share of EU visitors. Stated differently, websites with a larger share of EU visitors should have more incentive to respond to the regulation and might be more impacted since his response concerns a larger share of consumer. To explore this idea, we repeat our IV analysis by interacting the website response with the website's share of EU visitors. Results are reported in Table 5. The findings, still consistent across the different definitions of treatment and control, suggest that websites that respond to the GDPR do not experience a decrease in content quantity or user engagement, with the exception of a small but statistically significant coefficient for the interaction term, for *Page Views Per User.* The result would suggest that websites that do respond to the GDPR experience a decrease in page view per user which tends to be larger (in magnitude) the larger the website's share of EU visitors.

Table 5: LATE Estimations Accounting for Share of EU Visitors

	Log GDELT URLs		Page views per user		Reach per Million		Rank		Page View Per Million		FB Average Reaction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	GDPR Response × Share EU Visitors	Log GDELT URLs	GDPR Response × Share EU Visitors	Page Views Per User	GDPR Response × Share EU Visitors	Reach Per Million	GDPR Response × Share EU Visitors	Rank	GDPR Response × Share EU Visitors	Page Views Per Million	GDPR Response × Share EU Visitors	FB Average Reaction
LATE 1: Treatment group base on EU Websites												
EU websites × Post GDPR × Share EU Visitors	-0.143***		-0.102***		-0.102***		-0.102***		-0.102***		-0.143***	
	(0.034)		(0.027)		(0.027)		(0.027)		(0.027)		(0.034)	
‘Post GDPR × Share EU Visitors	0.711***		0.696***		0.696***		0.696***		0.696***		0.711***	
	(0.018)		(0.015)		(0.015)		(0.015)		(0.015)		(0.018)	
EU websites × Post GDPR	0.053***		0.039***		0.039***		0.039***		0.039***		0.053***	
	(0.014)		(0.012)		(0.012)		(0.012)		(0.012)		(0.014)	
GDPR Response		0.654		0.177		-756.499**		-4.76e+04		-30.266**		124.656
		(0.643)		(1.075)		(347.483)		(53946.067)		(13.939)		(342.331)
GDPR Response × Share EU Visitors		0.036		-0.208***		74.369**		9738.414		0.850		14.328
		(0.086)		(0.064)		(34.511)		(7086.062)		(1.832)		(24.886)
Constant	0.001		0.000		0.000		0.000		0.000		0.001	
	(0.004)		(0.004)		(0.004)		(0.004)		(0.004)		(0.004)	
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		0.688		0.688		0.688		0.688		0.688		0.688
Share of response inside Treatment		0.828		0.828		0.828		0.828		0.828		0.828
Underidentification (LM)		12.783		11.957		11.957		11.957		11.957		12.783
P-value (LM-Stat)		0.002		0.003		0.003		0.003		0.003		0.002
Weak identification		2.252		3.544		3.544		3.544		3.544		2.252
J-Stat		0.786		0.152		0.333		2.739		0.023		0.790
P-value (J-Stat)		0.375		0.697		0.564		0.098		0.880		0.374
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: Treatment group base on EU Websites + Websites with more than 10% of EU Visitors												
EU Websites and > 10% EU Visitors × Post GDPR × Share EU Visitors	-0.162***		-0.114***		-0.114***		-0.114***		-0.114***		-0.162***	
	(0.037)		(0.030)		(0.030)		(0.030)		(0.030)		(0.037)	
‘Post GDPR × Share EU Visitors	0.709***		0.695***		0.695***		0.695***		0.695***		0.709***	
	(0.018)		(0.015)		(0.015)		(0.015)		(0.015)		(0.018)	
EU Websites and > 10% EU Visitors × Post GDPR	0.071***		0.050***		0.050***		0.050***		0.050***		0.071***	
	(0.018)		(0.015)		(0.015)		(0.015)		(0.015)		(0.018)	
GDPR Response		0.556		-1.014		-623.297**		-7.00e+04		-27.846**		266.665
		(0.641)		(0.635)		(280.007)		(43528.078)		(11.630)		(380.212)
GDPR Response × Share EU Visitors		0.032		-0.192***		72.588**		10031.457		0.819		18.591
		(0.084)		(0.072)		(31.029)		(7299.844)		(1.782)		(27.582)
Constant	0.001		0.000		0.000		0.000		0.000		0.001	
	(0.004)		(0.004)		(0.004)		(0.004)		(0.004)		(0.004)	
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		0.690		0.690		0.690		0.690		0.690		0.690
Share of response inside Treatment		0.829		0.829		0.829		0.829		0.829		0.829
Underidentification (LM)		15.783		15.697		15.697		15.697		15.697		15.783
P-value (LM-Stat)		0.000		0.000		0.000		0.000		0.000		0.000
Weak identification		3.623		5.092		5.092		5.092		5.092		3.623
J-Stat		1.219		0.199		0.000		4.676		0.201		1.252
P-value (J-Stat)		0.270		0.655		0.998		0.031		0.654		0.263
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: Treatment group base on EU Websites + Websites with more than 5% of EU Visitors												
EU Websites and > 5% EU Visitors × Post GDPR × Share EU Visitors	-0.154***		-0.109***		-0.109***		-0.109***		-0.109***		-0.154***	
	(0.036)		(0.029)		(0.029)		(0.029)		(0.029)		(0.036)	
Post GDPR × Share EU Visitors	0.710***		0.695***		0.695***		0.695***		0.695***		0.710***	
	(0.018)		(0.015)		(0.015)		(0.015)		(0.015)		(0.018)	
EU Websites and > 5% EU Visitors × Post GDPR	0.064***		0.046***		0.046***		0.046***		0.046***		0.064***	
	(0.017)		(0.014)		(0.014)		(0.014)		(0.014)		(0.017)	
GDPR Response		0.648		-0.814		-833.796**		-3.55e+04		-32.239**		286.756
		(0.643)		(0.683)		(363.667)		(57282.186)		(13.477)		(370.224)
GDPR Response × Share EU Visitors		0.035		-0.194***		75.409**		9562.937		0.878		19.321
		(0.086)		(0.069)		(36.571)		(7032.901)		(1.873)		(27.790)
Constant	0.001		0.000		0.000		0.000		0.000		0.001	
	(0.004)		(0.004)		(0.004)		(0.004)		(0.004)		(0.004)	
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		0.690		0.690		0.690		0.690		0.690		0.690
Share of response inside Treatment		0.828		0.828		0.828		0.828		0.828		0.828
Underidentification (LM)		14.171		13.831		13.831		13.831		13.831		14.171
P-value (LM-Stat)		0.001		0.001		0.001		0.001		0.001		0.001
Weak identification		2.851		4.256		4.256		4.256		4.256		2.851
J-Stat		0.996		0.067		0.430		2.186		0.021		1.642
P-value (J-Stat)		0.318		0.796		0.512		0.139		0.886		0.200
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

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

6.4 Lookahead Matching

In both the difference-in-difference and LATE analysis, the estimates we obtain capture both ecosystem and website-level response effects. In this section we attempt to isolate the effect of website-level responses. We focus on the implementation of cookie walls and consent mechanisms because, as we explained in section 3.1.2, these are the website level responses we expect to have a greater impact on websites' outcomes. This analysis is challenging because the response websites implement is likely correlated with its characteristics, many of which may not be observable. To deal with this challenge we implement an analysis following the logic of the Look Ahead Matching methodology (Bapna *et al.*, 2018). The endogeneity problem prevents us from directly comparing websites that adopt a particular response vs. websites that do not, as the differences we observe may be driven by the factors that led adopters to implement the response rather than by the response itself. The look ahead matching methodology is based on comparing websites that have adopted a response with websites that have not adopted a response but will adopt it some time in the future. In this way we isolate ourselves from the endogeneity problem, as we are basing our analysis only on 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.

Specifically, for this analysis we focus on EU websites during the post-GDPR period that adopt a cookie wall or a consent mechanism at some point in time. Of the 465 EU websites in our sample, 37 were using a cookie wall and 316 were using a consent mechanism for at least one of the waves. Considering only the subsample of sites that use a cookie wall (or a consent mechanism) for at least one wave, We estimate linear regressions of the form:

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

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 (a cookie wall or a consent mechanism) 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 6: Look Ahead Matching - Cookie Wall

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

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

The results presented in Table 6 show that the only statistically significant effect of the presence of a cookie wall remains over page views per user. Similarly to specifications presented in previous sections, we do not find evidence of a negative impact of the GPDR on content quantity or other measures of user engagement. In table 7 we repeat the analysis to determine if the use of a consent mechanism had any effect on the websites' outcomes. We did not find any statistically significant effect.

Table 7: Look Ahead Matching - Consent Mechanism

	Log GDELT URLs	Page Views Per User	Reach per Million	Rank	Page Views per Million	FB Average Reaction
Consent Mechanism	0.0347	0.00734	-6.275	-5,741	-0.639	-3.467
	(0.0485)	(0.0281)	(6.410)	(3,981)	(0.537)	(8.513)
Constant	5.253***	2.086***	244.3***	47,395***	16.43***	64.53***
	(0.0441)	(0.0224)	(7.149)	(3,206)	(0.575)	(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	0.710	0.975	0.708	0.978	0.633

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

6.5 Heterogeneous effects

Our findings suggest the GDPR had no impact on EU websites' ability to provide content, relative to their US counterpart, or on traffic and engagement measures, with the exception of a negative effect on page views per user. In this section we revisit these results accounting for website heterogeneity. We also investigate the effect the GDPR on content for the top ranking and bottom ranking EU websites (see Appendix B, table 10).

6.5.1 Advertising

We repeat our difference-in-difference analysis for the sub-sample of websites that relied on advertising to monetize their content before the GDPR. We separate our sample in two groups—"low" and "high" ads—respectively representing websites below and above the median advertising length before the GDPR. Table 8 shows that the negative effect we found on page views per user is similar for the low and the high ads group. Similarly to previous results, we find no statistically significant change in the quantity of new content published or in Facebook reactions.

Table 8: Diff-in-diff regressions: For content variables dependent by Low and High advertising subsample

	Log GDELT URLs		Page Views Per User		Reach Per Millon		Rank		Page Views Per Millions		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 base on EU Websites												
EU Websites × Post GDPR	0.060	-0.015	-0.062	-0.135***	40.942	11.333	-52.711	2847.824	0.409	0.223	3.297	11.129
	(0.067)	(0.055)	(0.057)	(0.036)	(27.689)	(14.621)	(5657.050)	(3787.949)	(1.668)	(0.834)	(23.327)	(18.528)
Constant	4.662***	5.288***	2.064***	2.036***	284.757***	195.507***	68011.163***	60787.870***	17.908***	11.008***	128.258***	95.349***
	(0.014)	(0.008)	(0.012)	(0.006)	(5.883)	(2.409)	(1201.870)	(624.157)	(0.354)	(0.138)	(4.729)	(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 base on EU Websites + Websites with more than 10% of EU Visitors												
EU Websites and > 10% EU visitors × Post GDPR	0.052	-0.015	-0.118**	-0.136***	47.111*	14.310	54.909	2972.581	0.444	0.334	10.198	10.096
	(0.066)	(0.055)	(0.058)	(0.036)	(26.872)	(14.564)	(5636.333)	(3791.712)	(1.611)	(0.832)	(23.198)	(18.451)
Constant	4.664***	5.288***	2.074***	2.036***	283.792***	195.027***	67988.702***	60769.582***	17.904***	10.990***	126.940***	95.514***
	(0.013)	(0.008)	(0.012)	(0.006)	(5.512)	(2.389)	(1156.048)	(621.883)	(0.330)	(0.136)	(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 base on EU Websites + Websites with more than 5% of EU Visitors												
EU Websites and > 5% EU visitors × Post GDPR	0.053	-0.015	-0.107*	-0.135***	40.189	11.333	1484.354	2847.824	0.392	0.223	12.512	11.129
	(0.066)	(0.055)	(0.058)	(0.036)	(26.970)	(14.621)	(5618.205)	(3787.949)	(1.624)	(0.834)	(23.259)	(18.528)
Constant	4.664***	5.288***	2.072***	2.036***	285.144***	195.507***	67692.996***	60787.870***	17.914***	11.008***	126.468***	95.349***
	(0.013)	(0.008)	(0.012)	(0.006)	(5.577)	(2.409)	(1161.858)	(624.157)	(0.336)	(0.138)	(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

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

6.5.2 Short vs Long run

The lack of consensus on what it means to be compliant with the GDPR may delay the effect (if any) of the regulation. As we show in Section 5, on the one hand, websites reacted rapidly to the enactment of the regulation by reducing the magnitude of visitor tracking - however, such reduction did not last over time. On the other hand, responses for a majority of EU websites evolved over time. In this section we compare short and long-run effects. We split our sample after the GDPR into two groups. The short-run subsample presents the effect of the GDPR until the beginning of January 2019. In table 9, columns (1), (3), (5), (7), (9), and (11) present the estimation on the short-run group. Columns (2), (4), (6), (8), (10), and (12) present the estimation for the long-run analysis subsample. The long-run subsample excludes from the analysis the time period just after the GDPR, meaning the period from June 2018 to January 2019. In other words, it includes all the pre-GDPR waves in our data (from April 2017 to May 25, 2018) and the latest post-GDPR waves (from July 2019 to November 2019). For the most part, results do not change much from our previous analysis. We find very little evidence of an impact of GDPR on websites' ability to provide content or on visitors' engagement. The only difference is in columns (3) and (4), which suggest that the decrease in page views per user for EU websites compared to US websites may be a long-run effect.

Table 9: Diff-in-diff regressions: For content variables dependent by short and long-run subsample

	Log GDELT URLs		Page Views Per User		Reach Per Millon		Rank		Page Views Per Millions		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												
EU Websites × Post GDPR	-0.006	0.019	-0.023	-0.229***	15.379	27.123	1922.883	3379.643	0.316	-1.081	13.798	10.722
	(0.038)	(0.067)	(0.031)	(0.057)	(10.938)	(21.238)	(3287.165)	(5158.319)	(0.744)	(1.300)	(15.087)	(24.070)
Constant	5.037***	5.041***	2.051***	2.056***	251.933***	247.986***	60798.116***	64149.592***	14.901***	14.755***	106.497***	108.150***
	(0.005)	(0.005)	(0.004)	(0.005)	(1.600)	(1.808)	(480.746)	(439.013)	(0.109)	(0.111)	(2.058)	(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	-0.039	-0.279***	20.648*	32.061	1043.402	5107.264	0.516	-1.180	16.283	11.719
	(0.038)	(0.067)	(0.031)	(0.056)	(10.782)	(20.964)	(3311.108)	(5130.506)	(0.731)	(1.287)	(14.960)	(23.715)
Constant	5.037***	5.041***	2.053***	2.060***	251.227***	247.624***	60930.021***	64011.932***	14.873***	14.761***	106.210***	108.095***
	(0.005)	(0.005)	(0.004)	(0.005)	(1.543)	(1.746)	(473.834)	(427.228)	(0.105)	(0.107)	(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	-0.036	-0.270***	16.686	25.225	2234.922	4650.988	0.431	-1.266	16.944	15.165
	(0.038)	(0.067)	(0.031)	(0.057)	(10.834)	(21.052)	(3302.734)	(5138.329)	(0.735)	(1.291)	(14.994)	(23.822)
Constant	5.037***	5.041***	2.053***	2.059***	251.779***	248.180***	60757.401***	64047.366***	14.885***	14.769***	106.107***	107.831***
	(0.005)	(0.005)	(0.004)	(0.005)	(1.561)	(1.765)	(475.752)	(430.709)	(0.106)	(0.108)	(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

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

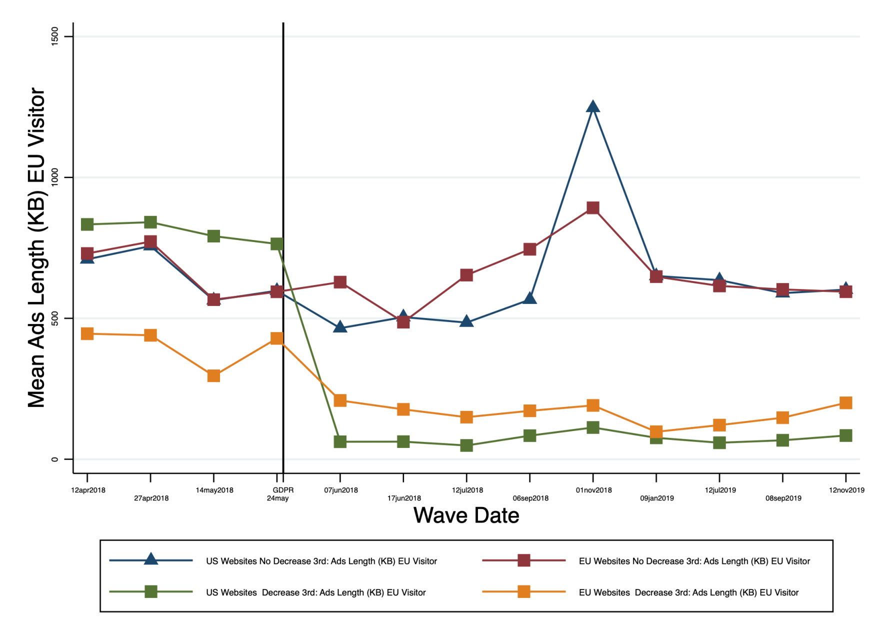
7 Discussion

Based on industry claims and prior research, we expected the GDPR to produce significant long term effects on content providers. Instead, we found very limited ecosystem or website level effects of the GDPR. The ability of EU based outlets to produce content and engage audiences does not seem to have been affected, their overall ranking does not seem significantly changed (as compared to US websites), and more importantly we observe almost no exit of websites. In short, our results suggest that the GDPR did not produce significant changes in downstream outcomes: quality and quantity of content seem to be, on average, stable. This suggests that either overall revenues from ads did not decrease (or at least not substantially so), and therefore websites did not need to adjust their offering; or, perhaps, revenues from ads did decrease, but websites relied on different mechanisms to compensate for this loss. We explore possible alternative mechanisms below.

A first possible explanation for the lack of more significant downstream effects on content provision would be that revenues from ads decreased, but websites' revenues did not (and hence quantity and quality of content did not vary), because EU websites switched to other sources of revenue/business models. While we do not currently have data on changes in subscription models, we do know the number of websites that implemented cookie walls or paywalls. Our data suggest that only a small proportion of EU websites decided to implement cookie walls, and that this number increased from 2.3% and 4.1% during the period of observation, but ended up by the end of that period to levels reverting back to nearly identical to the levels pre-GPDR (roughly 2.5%). In the case of paywalls we don't observe any increase, with the number of websites using them remaining stable at roughly 1.4%. In summary, we find evidence of a few more websites implementing cookie walls following the enactment of the GDPR, but we do not find an increase in the number of EU websites permanently switching to cookie walls.

A second possible explanation would be that EU websites (and in particular, among them, websites that responded to the GDPR in the most forceful manner - for instance by curtailing tracking) attempted to compensate revenue losses due to reduced tracking by increasing ad intensity, even though total number of pages may have stayed the same. However, this also does not seem to be the case. As figure 8 shows, the few EU websites (about 40) that decreased tracking (orange line) did not experience a systematic change in ads length (as we defined it in previous sections): ad length decreased somewhat after the enactment of the GDPR, and picked up again soon after. Even EU websites and US websites that kept tracking constant (red line and blue line, respectively) following the GDPR had somewhat stationary ad intensity (the peak observed for some EU and US websites is associated with the Christmas period). If anything, the ad length for US websites that chose to decrease tracking (green line) seems to decrease, suggesting that a reduction in tracking is correlated with less advertisements, not more.

Fig. 8 *Mean Advertising Length for Sites that Decrease/Don't Decrease 3rd Party Cookies*



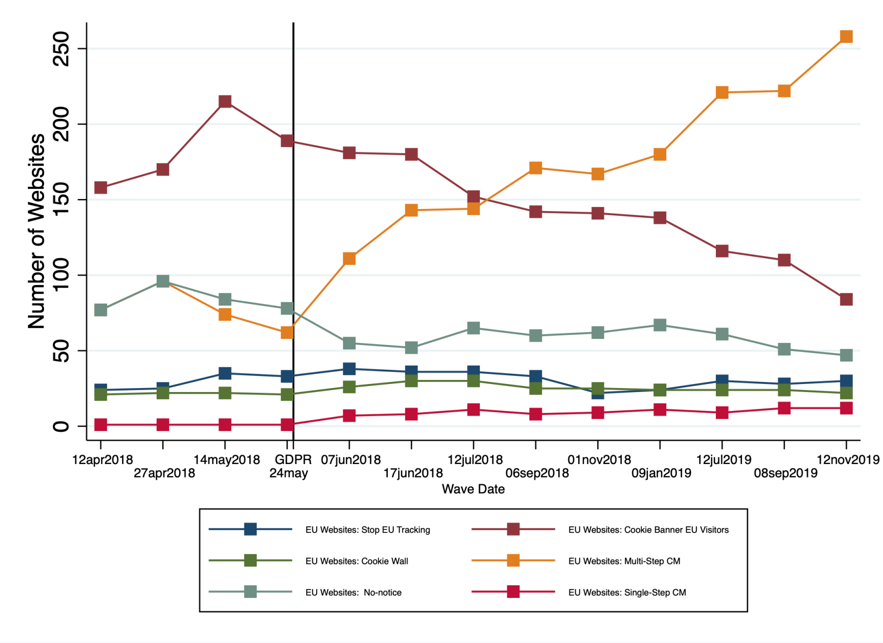
We further confirm the result presented above by examining ad length on EU websites by type of response to the GDPR. While the ad length fluctuates over time (with, again, some decrease after GDPR enactment, followed by an uptick around the Christmas season), by the end of our period of observation the ad length 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 15 in the Appendix).

Another possible set of explanations focuses on the possibility that revenues from ads did not change because the amount of data available for targeting in the EU ecosystem did not ultimately change much. The reasons why data availability may not have changed significantly, especially in the long term, are various and diverse, and anecdotal evidence supports this explanation. For instance, industry reports suggest that a number of third-party players entered the market several months following the enactment of the GDPR and helped publishers manage compliance requirements for websites that, in turn, felt pressure to adopt these platforms as ad buyers placed higher value on inventory with information on user consent (Davies, 2018). Furthermore, various studies have suggested that a number of CMs implemented by websites could have quite effectively nudged visitors towards acquiescence to tracking (see Section 2). Below we consider explore a number of different possible dynamics consistent with the conjecture that the amount of data available in the ecosystem did not change significantly in the long term.

One dynamics that could lead to the availability of data in the ecosystem not changing much is that most visitors may have consented to be tracked or may not have opted-out from tracking. 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, 2021). But in the case of the GDPR, few websites made opt-out choices easily accessible to visitors. Figure 9 shows how the number of websites that introduce a consent mechanism increases overtime. We differentiate between 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 both increase over time following the enforcement of the GDPR, multi-step CMs are much more prevalent and are adopted at a faster rate. These CMs are arguably more likely to dissuade visitors from actually completing the process of opting out of tracking.

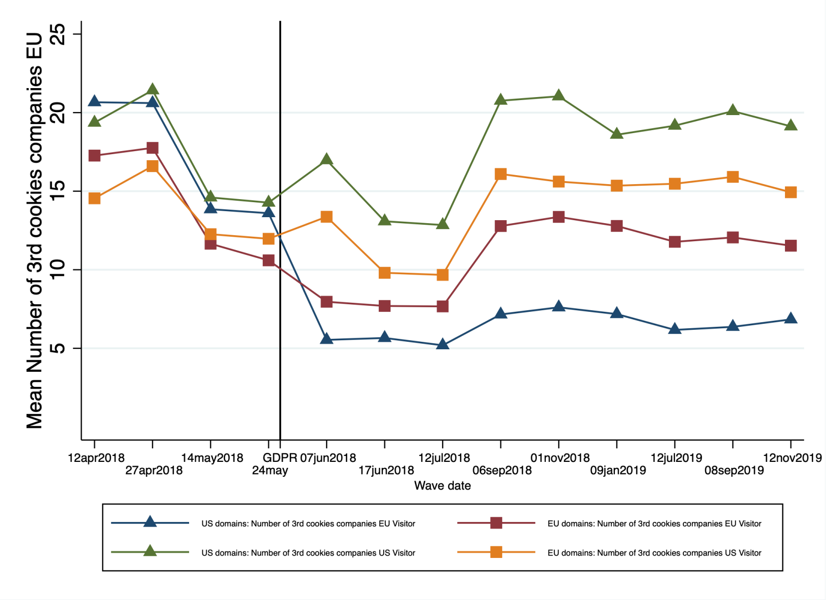
Fig. 9 *Website Level Response to GDPR by EU Websites*



Another, related dynamics that could lead to the amount of data in the ecosystem staying unchanged is that the market of third-party players in the ecosystem (especially trackers) may have become more concentrated, but the actual volume of data collected may not have changed. For instance, following GDPR, more advertising may have started to go through large players (such as Google and Facebook), but the total amount of advertising didn't change much, and publishers ultimately received stable revenues. Put in other terms, there could be as much data as before in the EU ecosystem because large players were able to keep tracking most visitors (because visitors "sign in" into those players' services, such as Gmail or Facebook, and agree to their terms) and now these players run as many ads as the industry collectively did before. The net effect would be that the same amount of ads based on the same data are being served. While we are not able to directly investigate the amount of total consumer data available in the EU ecosystem over time, we can—as noted above—study the extent of tracking that EU websites are engaged in over time. As noted in Section 5, some of the figures we presented do suggest that EU-website level tracking (as measured by the number of third party-trackers) decreased initially, and then picked up again several months after the enactment of the GDPR.

In fact, we can complement those figures from the previous sections in various ways. We can look at how the number of third party trackers on EU websites changed over time, but comparing EU websites that responded to GDPR in different ways. Figure 16 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. Additionally, we can look at whether the market did become more concentrated over time—specifically, we can consider the number of third party trackers unique companies in the market (Johnson and Shriver, 2019). Figure 10 shows a surprising result: the distribution of the number of different 3rd parties companies over time tracks precisely the graph of the number of 3rd parties cookies we presented in previous sections. In other words, while some third party companies left the market, making it more concentrated, they did so temporarily, and then got back into it.

Fig. 10 *3rd Party Cookies Companies EU/US Websites for EU/US Visitors*



A related dynamics we consider is that publishers adopted responses and compliance postures in manners strategically designed not to hurt them (or, evolved their response over time, from one that hurt them initially to one less likely to cause negative outcomes); similarly, advertising technology firms over time may have found ways of making as much money as in the past but while being compliant with GDPR. As a matter of fact, we already considered some supporting (or potentially consistent) evidence: Figure 1 in Section 5 suggests that the amount of tracking at the website level (third-party cookies) decreased and then increased; as well as the Sankey diagrams (also in Section 5) indicating a more dynamic reaction by EU websites over time, compared to their US counterparts. Additionally, in section 6.4 we found evidence that cookie walls, which was a website-level response that grew after GDPR became effective, but was later abandoned by most websites that had adopted it, was the only website-level response we could associate with a negative effect on outcomes (page views per user). On the contrary, we couldn't find any negative effect of the presence of consent mechanisms on outcomes, which is the response to GDPR that grew the most until being adopted by most EU websites in our sample. Furthermore, anecdotal industry reports suggest that the growing popularity of consent mechanism 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 which track a rise in these platforms following GDPR enforcement (Hils *et al.*, 2020). While consent mechanisms may, in theory, reduce the amount of data available to the publishers (by allowing consumers to opt-out), we have noted above how other research has reported high prevalence of dark patterns nudging visitors towards acceptance of tracking in GDPR consent mechanisms (Nouwens *et al.*, 2020); and we have noted how the emergence of CMPs may have facilitated the acquisition of visitors' data across websites - which would explain why, over time, more EU websites switched to CMs. While it is difficult to investigate these dynamics, we have compared, above (Section 6), differences in the impact of GDPR on downstream variables in the short run versus the long run, and we have found that a reduction in page views per user only materializes in the long run.

In summary, we were able to rule out—as likely explanations for a lack of more pronounced negative downstream effects on websites' content—an increase in advertising intensity, or an increase in revenue models that do not rely on tracking. On the other hand, we were *not* able to rule out the possibility that EU websites, after an initial decrease in tracking, over time reached levels of tracking comparable to pre-GDPR levels, or adopted responses and compliance postures in manners strategically designed not to hurt them, and thus were able to maintain revenue levels comparable to pre-GDPR periods.

8 Limitations

Our analysis provides insights into the impact of the GDPR on websites' content quantity and users' engagement as a proxy for content quality. Overall, the GDPR has 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. Furthermore, the enactment of the GDPR may have to some extent negatively affected page views of EU websites, relative to US websites. However, and importantly, the enactment does not seem to have relatively affected the amount of content that EU websites were able to publish, the amount of traffic they received, or the degree of average social media engagement and interaction with such content.

Before concluding, we feel it is important to highlight some limitations of our analysis. While we are using multiple measures to capture content quantity and quality, they are only proxies that may not fully capture the potential effect of the GDPR. Additionally, while we have 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 only capturing a part of the technical changes that are possible.

Finally, despite being over two years into the GDPR, it may still be too early to detect changes in the content produced by publishers. Firms, weighting 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's 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.

9 Conclusion

While previous work has focused on measuring the effects of the GDPR on advertising technologies (such as cookies), the present study attempts to assess the impact of the GDPR on ad-supported content publishers by tracking the potential downstream effects of the regulation. We captured 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.

We examined these variables using multiple identification strategies including DID estimations, LATE models, and a look ahead analysis. The DID analysis examined the changes in our outcomes of interest for US and EU websites viewed from US and EU visitor addresses. For websites viewed from the EU, relative to websites viewed from the US, our results indicate a reduction in the variables often associated with tracking; we also observed some evidence of a negative impact of the regulation on the traffic of EU websites. However, we did not find significant evidence of a negative effect of the regulation on the amount of content that EU websites publish, or the degree of average social media engagement and interaction with such content. The robustness of this result was confirmed by using different methodologies to account for endogeneity concerns. Moreover, we explicitly attempted to identify and distinguish between ecosystem effects (that affect all EU websites regardless of their response) and website-level effects.

References

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.

Angrist, J. and Imbens, G. (1995). *Identification and estimation of local average treatment effects*.

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

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 op onjuiste wijze toestemming voor plaatsen tracking cookies*. `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*. `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 Q.* 42(3), 719–736. ISSN 0276-7783. doi:10.25300/MISQ/2018/13592. Retrievable at `https://doi.org/10.25300/MISQ/2018/13592`.

Beales, H. (2010). The value of behavioral targeting. *Network Advertising Initiative*. 1, 2010.

Brill, J. (2018). *Microsoft's commitment to GDPR, privacy and putting customers in control of their own data*.

Cagé, J., Hervé, N. and Viaud, M.-L. (2015). The production of information in an online world. *The Review of Economic Studies*.

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.

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.

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*.

Davies, J. (2019). *After GDPR, The New York Times cut off ad exchanges in Europe — and kept growing ad revenue*. https://digiday.com/media/gumgumtest-new-york-times-gdpr-cut-off-ad-exchanges-europe-ad-revenue/. Accessed: 2019-05-08.

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.

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. ISSN 2045-2322. doi:10.1038/s41598-021-98628-y.

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. ISSN 0027-8424. doi:10.1073/pnas.2105624118.

Godinho de Matos, M. and Adjerid, I. (2021). Consumer Consent and Firm Targeting After GDPR: The Case of a Large Telecom Provider. *Management Science*. ISSN 0025-1909. doi:10.1287/mnsc.2021.4054.

Goldberg, S., Johnson, G. and Shriver, S. (2019). Regulating privacy online: An economic evaluation of the GDPR. *Available at SSRN 3421731*.

Goldberg, S., Johnson, G. and Shriver, S. (2021). Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic & E-Commerce Outcomes. *Available at SSRN 3421731*.

Goldfarb, A. (2004). Concentration in advertising-supported online markets: An empirical approach. *Econom. Innovation New Tech.* 13(6), 581–594.

Goldfarb, A. (2014). What is different about online advertising? *Review of Industrial Organization*. 44(2), 115–129.

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.

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. ISBN 9781450381383, 317–332. doi:10.1145/3419394.3423647.

IAB Europe (2021). *GDPR Guidance: Legitimate Interests Assessments (LIA) for Digital Advertising*.

IHS Technology (2015). *Paving the way: how online advertising enables the digital economy of the future*. Technical report.

International Association of Privacy Professionals (2020). *Behavioral Advertising*. `https://iapp.org/resources/article/behavioral-advertising-2/`.

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*, –. ISSN 0732-2399. doi:10.1287/mksc.2020.1271.

Johnson, G. and Shriver, S. (2019). Privacy & market concentration: Intended & unintended consequences of the GDPR. *Available at SSRN*.

Johnson, G. A., Lewis, R. A. and Nubbemeyer, E. I. (2017). Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research*. 54(6), 867–884.

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*.

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. doi:10.1007/s11002-014-9310-5.

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*.

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, 213–238. ISSN 0742-1222. doi:10.2753/MIS0742-1222300208.

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.

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. ISBN 9781450367080, 1–13. doi:10.1145/3313831.3376321.

Peukert, C., Bechtold, S., Batikas, M. and Kretschmer, T. (2020). European Privacy Law and Global Markets for Data. *SSRN Electronic Journal*. ISSN 1556-5068. doi:10.2139/ssrn.3560392. Retrievable at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3560392`.

Phelan, D. (2018). *Apple Promotes Powerful Privacy Tools For iPhone, iPad, Mac Users In GDPR Response*.

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. *SSRN Electronic Journal*. ISSN 1556-5068. doi:10.2139/ssrn.3503065. Retrievable at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3503065`.

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).

UK Information Commissioner's Office (2019). *Update report into adtech and real time bidding*. https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf.

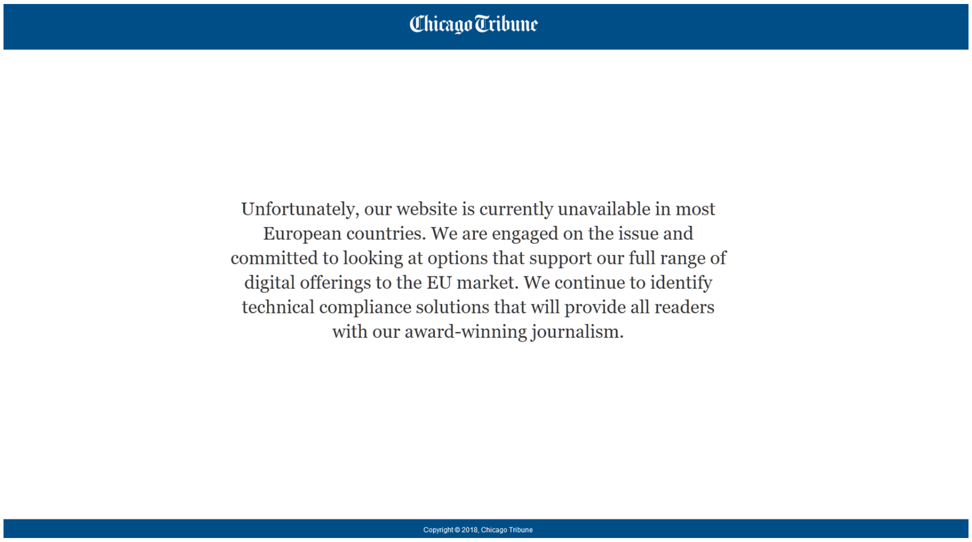
Urban, T., Tatang, D., Degeling, M., Holz, T. and Pohlmann, N. (2020). Measuring the Impact of the GDPR on Data Sharing in Ad Networks. *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security.*

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*. 973–990.

Zhuo, R., Huffaker, B., kc claffy and Greenstein, S. (2021). The impact of the General Data Protection Regulation on internet interconnection. *Telecommunications Policy*. 45. ISSN 03085961. doi:10.1016/j.telpol.2020.102083. Retrievable at https://www.sciencedirect.com/science/article/abs/pii/S0308596120301737.

Supplementary Appendix A:
Types of response

Fig. 11 *Example of Blocks EU Visitors*



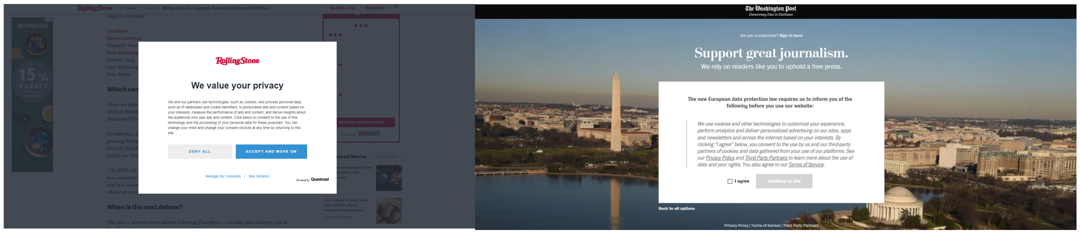
Chicago Tribune

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

Note: This figure presents an example of Blocks EU websites

Fig. 12 *Examples of Types 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.

DENY ALL | ACCEPT AND MOVE ON

Manage My Consents | See Vendors

Powered by Quantcast

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.

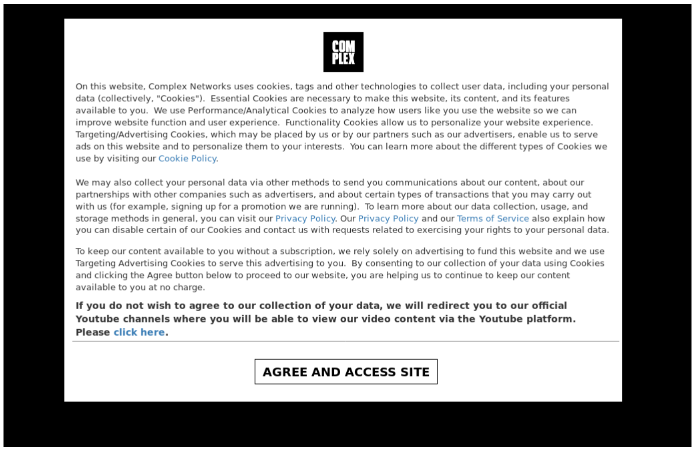
□ I agree | Continue to site

Back to all options

Privacy Policy | Terms of Service | Third Party Partners

Note: This figure presents different kinds of consent mechanisms

Fig. 13 *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.

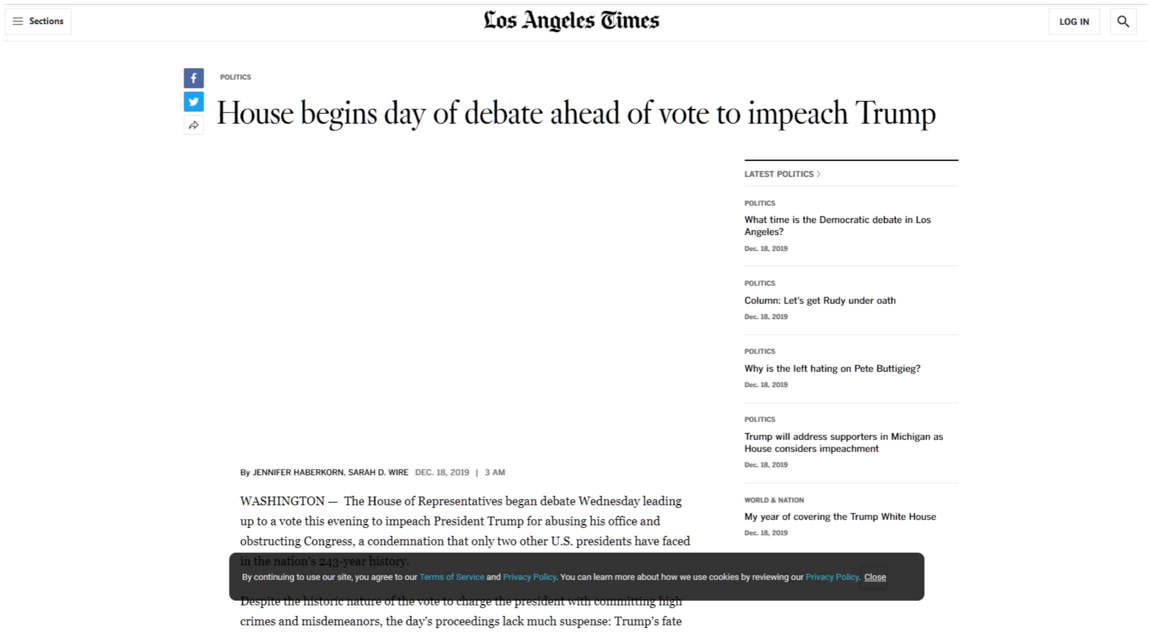
We may also collect your personal data via other methods to send you communications about our content, about our partnerships with other companies such as advertisers, and about certain types of transactions that you may carry out with us (for example, signing up for a promotion we are running). To learn 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.

Note: This figure presents a Cookie wall example

Fig. 14 *Cookies 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

Note: This figure presents different kinds of consent mechanisms

Supplementary Appendix B: Ranking

Given the heterogeneity of websites in terms of size and compliance abilities, we investigate how the GDPR has affected the most prominent websites according to Alexa ranking. We split our sample into two groups: the first group consists of websites ranking in the top 50% of websites in their respective region (EU/US) with respect to websites in our sample; the second group includes the remaining websites. Table 10 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 estimation for the sub-sample of the top ranking websites. The results reported in Columns (3) and (4) suggest that the implementation of the GDPR has a negative and significant effect on page views for both top ranking and bottom ranking EU websites. However, there does not seem to be a significant change in number of GDELT URLs (columns 1 and 2), or Facebook reactions (columns 7 and 8).

Table 10: Diff-in-diff regressions: based on the subsample of the top/bottom websites in EU and in US

	Log GDELT URLs		Page Views Per User		Reach Per Millon		Rank		Page Views Per Millions		FB Average Reaction	
	(1) Top	(2) Small	(3) Top	(4) Small	(5) Top	(6) Small	(7) Top	(8) Small	(9) Top	(10) Small	(11) Top	(12) Small
Intention to Treat 1: Treatment group base on EU Websites												
EU Websites × Post GDPR	0.076	-0.062	-0.107**	-0.092***	45.783	-2.624***	1043.841	4442.115	0.417	-0.139***	17.379	0.502
	(0.059)	(0.059)	(0.054)	(0.035)	(28.110)	(0.952)	(3018.552)	(6009.096)	(1.708)	(0.041)	(28.517)	(7.085)
Constant	5.323***	4.634***	2.151***	1.949***	461.893***	18.287***	13239.239***	1.16e+05***	27.933***	0.913***	170.764***	40.099***
	(0.011)	(0.010)	(0.010)	(0.007)	(5.329)	(0.179)	(572.214)	(1129.430)	(0.324)	(0.008)	(5.223)	(1.176)
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.	9096	7919	10619	10476	10619	10476	10619	10476	10619	10476	9096	7919
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	0.069	-0.059	-0.162***	-0.092***	54.212**	-2.715***	-260.787	5291.371	0.522	-0.141***	21.319	0.805
	(0.059)	(0.059)	(0.053)	(0.035)	(27.532)	(0.957)	(3024.609)	(6011.410)	(1.671)	(0.041)	(28.096)	(7.042)
Constant	5.324***	4.633***	2.161***	1.949***	460.607***	18.298***	13485.054***	1.16e+05***	27.916***	0.913***	170.162***	40.051***
	(0.010)	(0.010)	(0.010)	(0.006)	(5.061)	(0.178)	(555.988)	(1115.519)	(0.307)	(0.008)	(4.989)	(1.152)
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.	9096	7919	10619	10476	10619	10476	10619	10476	10619	10476	9096	7919
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	0.069	-0.059	-0.151***	-0.092***	45.831*	-2.715***	1334.354	5291.371	0.422	-0.141***	24.126	0.805
	(0.059)	(0.059)	(0.054)	(0.035)	(27.790)	(0.957)	(3020.706)	(6011.410)	(1.686)	(0.041)	(28.270)	(7.042)
Constant	5.324***	4.633***	2.159***	1.949***	462.035***	18.298***	13188.566***	1.16e+05***	27.934***	0.913***	169.608***	40.051***
	(0.011)	(0.010)	(0.010)	(0.006)	(5.176)	(0.178)	(562.667)	(1115.519)	(0.314)	(0.008)	(5.085)	(1.152)
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.	9096	7919	10619	10476	10619	10476	10619	10476	10619	10476	9096	7919

Notes: Estimates from the DID estimation for website with top and small rank before the GDPR. Column (1) presents estimation for the *Log GDELT URLs* dependent variable on the top rank subsample. Column (2) reports estimation with *Log GDELT URLs* as dependent variable for the small rank subsample. Column (3) and (4) present respectively *Page Views Per User* as dependent variable for top and small rank subsample. Column (5) and (6) reports estimation with *FB average Reaction* as dependent variable. All estimations include waves and website fixed effect.
Standard errors in parentheses and clustered at the website level.
Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

Supplementary Appendix C: Additional graphics

Fig. 15 *Mean Advertising Length for Sites by type of website-level response to GDPR*

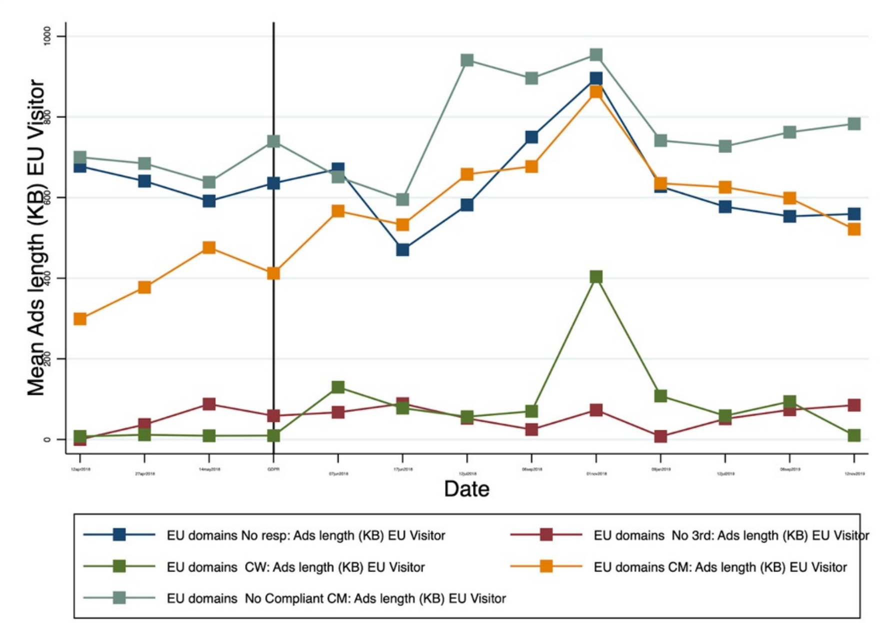


Fig. 16 *3rd Party Cookies By Type of Responses*

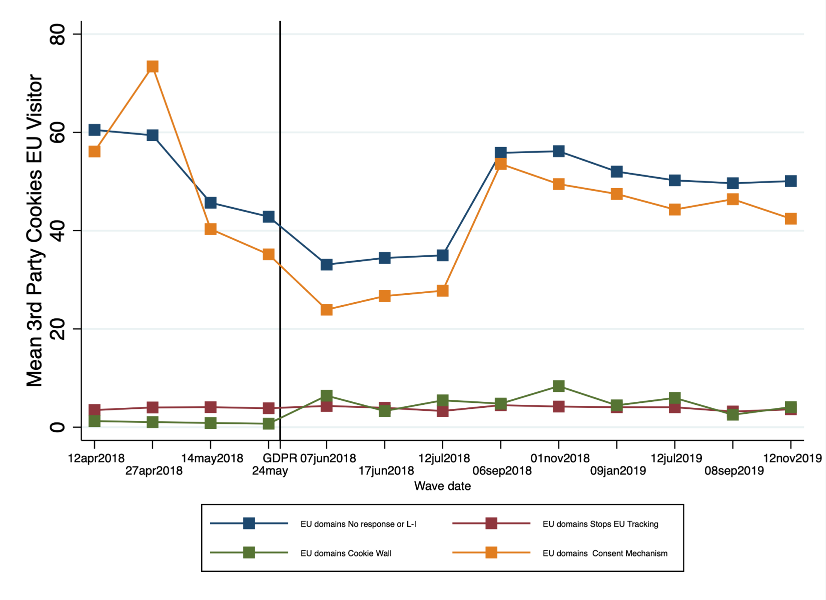
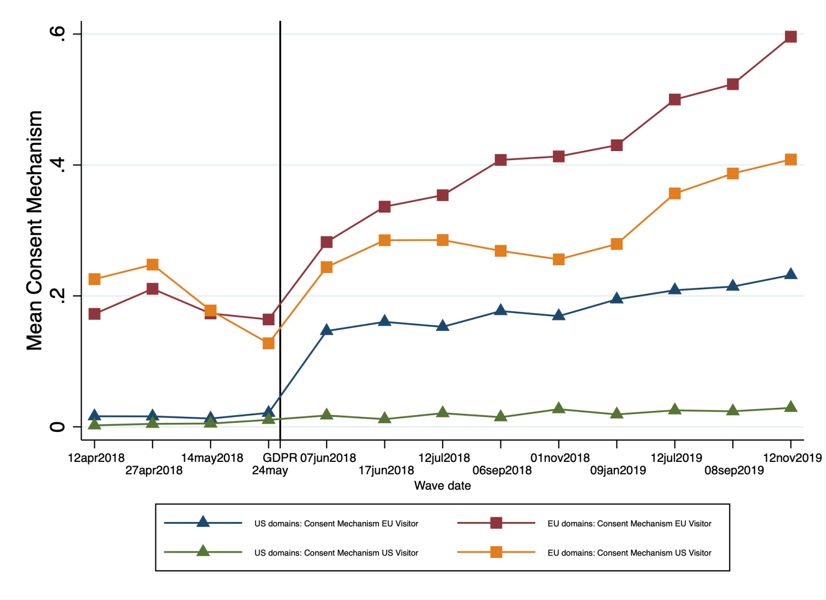


Fig. 17 *Consent Mechanism EU/US Websites for EU/US Visitors*


Mean Consent Mechanism
12apr2018
27apr2018
14may2018
GDPR
24may
07jun2018
17jun2018
12jul2018
06sep2018
01nov2018
09jan2019
12jul2019
08sep2019
12nov2019
Wave date
US domains: Consent Mechanism EU Visitor
EU domains: Consent Mechanism EU Visitor
US domains: Consent Mechanism US Visitor
EU domains: Consent Mechanism US Visitor