The Impact of the General Data Protection Regulation (GDPR) on the Amount of Online Tracking

Karlo Lukic[1], Klaus M. Miller[2], Bernd Skiera[3]

This version: 2021-12-15

Working paper

[1] Karlo Lukic, Doctoral Candidate, Department of Marketing, Faculty of Business and Economics, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany, Phone +49-69-79-834565, lukic@wiwi.uni-frankfurt.de.

[2] Klaus M. Miller, Assistant Professor of Marketing, HEC Paris, Rue de la Liberation 1, 78350 Jouy-en-Josas, France, Research Associate, Hi!Paris Research Center for the Study of Artificial Intelligence in Business and Society, millerk@hec.fr.

[3] Bernd Skiera, Professor of Marketing, Department of Marketing, Faculty of Business and Economics, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany, Phone +49-69-798-34649, skiera@wiwi.uni-frankfurt.de.

Abstract

We determine whether GDPR's enforcement increased consumers' online privacy by decreasing the amount of online tracking. We exploit a difference-in-differences design to evaluate the effect of GDPR's enforcement using 718 websites visited by 5 million EU and US consumers browsing 2.6 billion web pages between April 2018 and December 2019 that target EU and non-EU audiences. We reveal that online tracking, measured by the number of tracker providers and the number of trackers, increased over time; websites that comply with GDPR use 12 tracker providers (19 trackers) on average in the post-GDPR period. Without GDPR, they would have used 13 tracker providers (21 trackers). So, GDPR's enforcement decreased the average number of tracker providers (trackers) on each website by 1 tracker provider or -8% (2 trackers or -9%), having a minor positive impact on consumers' online privacy. We find that the enforcement of the law decreased advertising, analytics, and essential trackers on news, entertainment, and business sites – emphasizing a minor positive impact on consumers' privacy. We conclude that the GDPR's enforcement increased consumers' privacy to a small extent. But the increasing trend of the amount of online tracking after the introduction of the GDPR poses a justifiable privacy concern for consumers.

Keywords: online privacy, online tracking, privacy law, GDPR

1. Introduction

Online tracker (e.g., DoubleClick) is software owned by a tracker provider (e.g., Google), placed on the website (e.g., The New York Times) by the website owner that uses tracking technologies (e.g., cookies) to track consumers. An online tracker (henceforth "tracker") is usually code embedded on the website (hereafter "site") that records and transmits consumers' data to the tracker provider (Karaj et al., 2018a).

The tracker provider – owning one or several trackers – decides whether to share that consumer data with the site owner. While generating consumer insight to a tracker provider, the tracker also supports the site concerning advertising (e.g., personalized ads) or consumer experience (e.g., Facebook's "Like" button). For this reason, 94% of 7.5 million sites used trackers last year (Viscomi, 2020).

While trackers benefit sites, they inherently raise privacy concerns because they often access information like IP addresses and current or last viewed web pages (subsequently "pages"). Such data reveals many insights, especially when collected over time and tied to a particular consumer via a unique identifier. E.g., Mayer & Mitchell (2012) explained that an individual's browsing history alone "can reveal her location, interests, purchases, employment status, sexual orientation, financial challenges, medical conditions, and more" (p.3).

Past initiatives reflected consumers' privacy concerns and tried to limit trackers from collecting and processing their (personal) data. One such initiative is the European Union's (EU) General Data Protection Regulation (GDPR), enforced in May 2018 as a revolutionary privacy law to protect EU consumers' privacy. GDPR is a model privacy law for many other countries worldwide (Lubowicka, 2019).

Our paper aims to tackle the question: Did GDPR increase consumers' privacy by decreasing the amount of online tracking? More specifically, we aim at answering the following two research questions, (1) did GDPR decrease the number of tracker providers that sites use, and (2) did GDPR decrease the number of trackers that sites use?

To answer our research questions empirically, we use the WhoTracks.me data sets generated by 5 million consumers browsing 2.6 billion pages from April 2018 to December 2019 (Karaj et al., 2018a). We differentiate between EU- and US-based consumers accessing 718 sites. Because two consumer bases accessed the same sites, our observation unit is a "site instance." Accordingly, a site instance represents one of the two consumer groups (EU or US) that accessed a site in a specific month. The above time frame captures before- and after-GDPR periods, whereas EU and US consumers' differentiation allows us to separate sites tracking EU consumers. The data enables us to measure consumers' exposure to online tracking.

We use SimilarWeb's data and the site's top-level domain (TLD; e.g., .com, .de) to determine whether each site receives the most traffic from EU or non-EU audience and, thus, targets EU or non-EU consumers. This procedure allows us to approximate whether a site should (not) comply with GDPR. Therefore, our control group consists of US consumers accessing sites that target non-EU consumers – all other site instances represent a treatment group. We set up our empirical study to utilize the difference-in-differences (DiD) design. We measure the amount of online tracking by counting the number of distinct trackers (tracker providers) a site uses.

Our results contribute to the ongoing discussion on the effectiveness of privacy laws at increasing consumers' privacy. EU consumers benefit from our results, as we show the minor positive implications of the law on their privacy. European regulators could use our results to evaluate whether the GDPR accomplished its aim. Likewise, EU regulators could benefit from our findings when finalizing the upcoming ePrivacy Regulation to extend GDPR's requirements in online settings (European Commission, 2021). Countries currently designing their privacy laws (e.g., Chile, New Zealand, Japan) and those that already enforced privacy laws (e.g., Brasil, India, China) based on GDPR could also gain valuable insights from our study. Our results might serve as an indication of the effects of privacy laws in their countries.

2. Description of online tracking

2.1. Definition of online tracking

We define online tracking as collecting data about consumers over time. From a tracker's perspective, tracking consists of two steps, (1) create a unique identifier using any tracking technology (e.g., tracking cookie, browser's local storage, digital fingerprint, tracking pixel, advertising identifier, login, single-sign-on), and (2) assign that unique identifier to a particular consumer, browser, or device (Falahrastegar et al., 2016; Soltani, 2011). Accomplishing both steps allows the tracker to collect consumer data over time across multiple sites, browsers, or devices. Note that the identity of a particular consumer behind the unique identifier might be known or unknown to the tracker (Mayer & Mitchell, 2012).

Tracking, in turn, allows (1) consumer profiling and (2) behavioral targeting of consumers (Lambrecht & Tucker, 2013). For instance, tracking consumers' past pages allows collecting characteristics such as demographics (e.g., female), interests (e.g., sport), purchase intentions (e.g., sports shoes), or brand preferences (e.g., Nike). Having such consumer characteristics allows a tracker provider to create a unique profile of that single consumer (e.g., a female consumer interested in Nike sports shoes) or a pool of consumers with similar characteristics (e.g., sport shoe lovers).

That information profits tracker providers and site owners as they can, e.g., target consumers with personalized ads or customize site content to each consumer's liking (Goldfarb & Tucker, 2011; Laub et al., 2021). At the same time, consumers face a trade-off between (a) a more personalized (free) online experience and (b) a loss of privacy (Tucker, 2012). We note that although consumer profiles may contain an enormous amount of information, this information is neither always accurate nor consistent (Kraft et al., 2021; Neumann et al., 2019).

2.2. Parties in online tracking

2.2.1. Users of online trackers

Users of online trackers use trackers to support them concerning advertising or consumer experience. Our paper focuses on sites as primary users of trackers. However, any internet-connected device (e.g., computer, smartphone, smart TV) or its software (e.g., sites displayed by browsers, browser extensions, mobile apps, smart TV apps) can use trackers for advertising or consumer experience support (Aggarwal et al., 2018; Binns et al., 2018; Kummer & Schulte, 2019).

Suppose that a site xyz.com starts offering free content to consumers. To improve the consumer experience, xyz.com combines its content with audiovisual content from other sites (e.g., YouTube), provides consumer sign-up via the single sign-on (SSO) mechanism of their favorite platform (e.g., Google account), allows consumers to share and "Like" its content via social media platforms (e.g., Facebook), and to comment on its content (e.g., Disqus). As the site receives consumer visits, it likely wants to analyze their characteristics (e.g., demographics, country of origin) and behavior (e.g., where they click, what pages they like most) to improve the offering. The site could use the Google Analytics tracker for that purpose. Over time, xyz.com implements changes based on consumer insights and attracts a larger audience because it can better cater its content to its consumer base. As its customer base grows, the site owner might like to monetize the site's online presence. The site owner could display advertisements to consumers and use, e.g., the DoubleClick (Google Ad Manager) tracker to achieve that.

The above example illustrates the interwovenness of today's sites with trackers (Falahrastegar et al., 2014). It is costly for a site like xyz.com to create an in-house software solution for each case to improve the consumer experience or display advertisements. So, the site relies on existing market solutions provided by trackers. Moreover, such software solutions fulfilled by trackers are likely free and convenient to implement for site owners.

For that reason, sites from various industries use trackers. In general, news sites tend to use many trackers as their revenue model involves selling ad spaces to advertisers and displaying ads to consumers while providing content (Libert et al., 2018; Libert & Nielsen, 2018). Thus, news sites have a high incentive to use advertising trackers (see Section 2.3.1). However, sites from other industries, such as e-commerce, entertainment, adult, governmental, banking, or health, also use trackers for consumer experience support (Macbeth, 2018; Sørrensen & Kosta, 2019). E.g., McCoy et al. (2020) recently showed that COVID-19 sites use trackers on 89% of their pages.

2.2.2. Providers of online trackers

Providers of online trackers are firms that develop one or several trackers. They offer such trackers as software solutions (i.e., software as a service, SaaS) to support users of trackers concerning advertising or consumer experience. Tracker providers benefit significantly from users of their trackers. They can collect more consumer data from different users (e.g., sites, mobile apps, smart TV apps) across devices (e.g., computers, smartphones, smart TVs) and combine such consumer data by purchasing it from other tracker providers (Bleier et al., 2020). Note, however, that just because a tracker provider owns multiple trackers does not mean it combines collected consumer data (Lerner et al., 2016). Having access to a large pool of consumer data allows tracker providers to improve their (tracking) services or target personalized ads more effectively if they offer their users advertising-related services. They can also sell gathered consumer data to other tracker providers (Bergemann & Bonatti, 2015; Lambrecht et al., 2014). Likewise, tracker providers can combine consumers' online data with offline data (e.g., store purchases), buying it from other tracker providers, or by owning such trackers (e.g., in location-based mobile apps) (Andrew & Baker, 2021; Exactag, 2021).

In the previous example, xyz.com used six trackers: YouTube, Google's SSO, Facebook's "Like" Button, Disqus, Google Analytics, and DoubleClick. In this case, Google — the tracker provider — owns four (YouTube, Google SSO, Google Analytics, DoubleClick) out of six trackers that the site uses for different purposes. However, other sites like xyz.com might enter the market with similar needs as xyz.com. The other sites might start using the same trackers as xyz.com, especially if they are free to use. Such an increase in tracker usage across different sites greatly benefits tracker providers. Nevertheless, the site owner might not be aware that by including a tracker to improve the consumer experience or display ads, the tracker can observe the behavior of its consumers and the consumers of any other site that uses the same tracker.

In the online advertising industry, any of the following firm types can act as tracker provider: advertising firms or agencies, advertising networks or exchanges (demand-side platforms, DSPs), data management platforms (DMPs), content management platforms (CMPs), analytics providers, retargeting providers, data providers, data aggregators, and others (Evidon, 2021; Mayer & Mitchell, 2012).

2.3. Categories of online trackers

2.3.1. Online trackers supporting advertising

Advertising trackers support sites in monetizing their consumers' data by better displaying advertisements. Such trackers offer advertising-related services like data collection, targeting, and retargeting to sites. Sites can provide content "free of charge" to their consumers using advertising trackers because they allow sites to sell ad space to advertisers and generate revenue (Deighton & Kornfeld, 2020). However, sites' consumers do not receive "free" content from the sites they visit, as they pay either with their data or willingness to see ads. E.g., Google Ads, Facebook Ads, Google Marketing Platform, DoubleClick (Google Ad Manager), BlueKai, or Datalogix, are trackers that provide advertising-related services to sites.

Advertising trackers benefit sites (e.g., in monetizing online presence), advertisers, ad networks, or ad exchanges (e.g., targeting and retargeting consumers, granular level of ad measurements, limiting consumer exposure to an ad, attribution modeling). Consumers also benefit from advertising trackers.

They can view more relevant ads and receive "free" site content. However, advertising trackers also pose privacy concerns for consumers.

E.g., irrespective of the privacy law(s), a site that allows consumers to conduct self-depression tests can use a single advertising tracker that employs programmatic advertising and real-time bidding (RTB) to display a different ad to each consumer. Using such a tracker, however, the site risks sharing consumer data (e.g., self-depression test answers) with hundreds of other trackers that are part of that RTB ecosystem, which might be unknown to the consumer and the site owner who placed the original tracker (Privacy International, 2019).

So, advertising trackers can invite other advertising trackers to the site. This practice is known as "piggybacking" as it describes the process of an original tracker giving access to other "piggybacking" trackers that were not initially placed on the site by the site owner (Hanson et al., 2018). In general, an advertising tracker adds another advertising tracker to the site if the original tracker delivers an ad to a consumer in an automated way from the second, third, fourth, and so on, tracker's server (i.e., its partners).

2.3.2. Online trackers supporting consumer experience

Consumer experience trackers support sites in enhancing consumer experience by offering a wide variety of services. E.g., consumer experience trackers provide services to sites such as analytics, social media or commenting integration, integration with external content, customer interaction, or hosting (Karaj et al., 2018a; Mayer & Mitchell, 2012). Trackers offering consent notice services that help sites comply with privacy laws also fall into this tracker category. The latter four kinds of trackers can be considered "essential" for a modern site (Karaj et al., 2018b). Some of the trackers supporting consumer experience on sites are Google Analytics, Facebook's "Like" or "Comment" widgets, Disqus widget, Weather Channel widget, OneTrust, or Amazon's content delivery network (CDN).

Trackers supporting consumer experience benefit sites as they allow them to measure their reach (e.g., count unique visitors), make improvements (e.g., change interface), or analyze consumers' characteristics (e.g., female consumers read content longer than male consumers). Such information supports sites in driving consumer engagement which might attract more consumers to the site or advertisers willing to target such consumers of the site. Consumer experience trackers also benefit consumers: they expose them, among others, to more relevant content (e.g., personalized news or product recommendations), allow content-sharing on their favorite social media platform (e.g., Twitter), or the convenience of signing up to different sites via SSOs.

On the other hand, consumer experience trackers violate consumers' privacy. Consumers might not realize that a site's content originates from a tracker. Thus, consumers might not know that a site shares their data with trackers. This lack of data sharing awareness usually implies that consumers did not permit a site to share their data with trackers or that consumers were aware of such data sharing but could not prevent it in any way. Moreover, consumers might not know if – and how – the trackers will use their data. EU regulators enforced the GDPR for such reasons as to increase consumers' online privacy.

3. Description of GDPR

3.1. Aim of GDPR

EU regulators enforced GDPR on May, 25th 2018 (European Commission, 2016). GDPR is essentially a privacy law applicable to all European Union (EU) member states. The regulation aims (1) to increase consumer privacy by strengthening consumers' control over their personal data and (2) to harmonize EU member states' existing national privacy laws via one regulation for all EU member states. GDPR achieves these aims by defining consumers' rights concerning their personal data (e.g., the right to access, edit, or remove data) and imposing obligations on firms that process such consumer data (i.e., tracker providers).

3.2. Area of GDPR applicability

Unlike previous EU privacy laws, which only affected EU firms, GDPR applies to EU firms and firms outside the EU that process EU citizens' personal data. The only case in which GDPR treats EU and non-EU firms differently is the processing of non-EU citizens' personal data; in that case, GDPR applies to EU firms, but it does not apply to non-EU firms. So, a non-EU site that receives traffic and processes data from EU and non-EU consumers does not have to comply with GDPR for non-EU consumers, while it must obey GDPR when processing data of EU consumers.

3.3. How GDPR affects online tracking

GDPR expanded the concept of personal data. European Commission (2016) define personal data as "any information relating to an identified or identifiable natural person" (Article 4, Definitions). As mentioned in Section 2.1, online tracking includes creating and linking a unique identifier to a particular individual. Thus, any online identifier (e.g., IP address, cookie identifier) makes an individual "identifiable" and is considered "personal data" under GDPR.

The law granted EU consumers more control over their personal data by asking any site worldwide that tracks them to (1) display a cookie consent banner to a consumer asking opt-in to tracking, and (2) clearly explain tracking purposes in its privacy policy. GDPR also introduced high non-compliance fines (€20 million or 4% of the site's global annual revenue, whichever amount is higher) and made the site and its tracker providers jointly responsible for tracking consumers (European Commission, 2016).

Such changes introduced by GDPR could negatively impact trackers by giving consumers a choice to "reject all tracking." Further, the negative impact of GDPR on trackers could reflect the fear of regulatory fines from sites or tracker providers (Haddon, 2018). However, the regulation could also positively impact trackers if consumers prefer personalized ads, reveal a tendency to opt-in to tracking, or ignore cookie banners, thereby "accepting all tracking" (Data & Marketing Association (DMA), 2019; Utz et al., 2019). Similarly, GDPR could positively affect trackers if sites do not offer consumers real opt-out choices (Degeling et al., 2019; Sanchez-Rola et al., 2019). Lastly, the law could not affect trackers. E.g., if the positive or negative effect of GDPR on trackers is only short-term or if the regulators do not enforce the law (Johnson et al., 2021; Peukert et al., 2021; Ryan, 2020).

4. Existing knowledge on the impact of GDPR on the amount of online tracking

In Table 1, we provide a brief overview of studies focusing on the impact of GDPR on the amount of online tracking.

Table 1: Existing knowledge on the impact of GDPR on the amount of online tracking

author(s)	main data source	method	number of months	consumer base control[1]	consumer exposure to trackers	tracker provider control[2]	differences in the effect of GDPR	key result	GDPR affected trackers
Degeling et al. (2019)	crawl	before/after	12	✗	✗	✗	✗	There was no significant change in the usage of trackers.	✗
Sørensen & Kosta (2019)	crawl[4]	before/after	21	✗	✗	✗	✗	Although usage of trackers decreased, the authors are not attributing this to GDPR.	✗
Sakamoto and Matsunaga (2019)	crawl	before/after	2	✗	✗	✗	✗	There was no significant change in the usage of trackers.	✗
Urban et al. (2020)	crawl	before/after	10	✓	✗	✗	✗	Usage of trackers decreased in the short term but increased in the long run.	✓
Johnson & Shriver (2021)	crawl	panel differences estimator	8	✓	✗	✗	✓	Usage of trackers decreased in the short term but increased in the long run.	✓
Peukert et al. (2021)	crawl	panel differences estimator	18	✓	✗	✗	✓	Usage of trackers decreased in the short term but increased in the long run.	✓
Our study	consumers	DiD	21	✓	✓	✓	✓	Usage of trackers and tracker providers decreased for the treatment group compared to the control group in the short term but increased in the long run.	✓
									∅ = 4

[1]We control which consumers (EU or US) accessed sites and count them as site instances. [2]We rule out whether the tracker owned by the tracker provider is a third-party present on a site (e.g., we do not count trackers owned by Amazon and present on Amazon sites as trackers). [4]We define a "crawl" as automatically visiting sites via software without interacting with the site.

Degeling et al. (2019) measured the law's impact on the number of trackers (tracking services). Although the authors found that the number of first-party cookies decreased by 18% one month after the introduction of the GDPR, they measured no significant difference in the sites' use of trackers (Degeling et al., 2019). Sørrensen & Kosta (2019) was cautious about attributing a decreasing number of trackers to GDPR. In contrast, Sakamoto & Matsunaga (2019) found no difference in the number of ad agencies (tracker providers) that offer consumers opt-out of tracking before- and after-GDPR. Urban et al. (2020) measured the amount of third-party cookie syncing, concluding that the regulation decreased third-party connections when GDPR came into effect. This relation slightly increased over the long run (Urban et al., 2020). But the law did not have a lasting impact on the number of trackers sites used.

Johnson et al. (2021) looked at the market concentration of trackers (vendors) while measuring the number of trackers sites used. Johnson et al. (2021) revealed that the regulation decreased the number of trackers in the first week after GDPR's enforcement, but the effect disappeared until the end of 2018. Although the aggregate market concentration mirrored that trend, the relative market concentration increased because "websites were more likely to drop smaller tracker providers" (Johnson et al., 2021, p.34).

Like Johnson et al. (2021), Peukert et al. (2021) assumed that the regulation changed the interplay between sites, trackers, and consumers, affecting the market structure and competition. They showed that sites' usage of trackers decreased in the short term but increased over the long run (Peukert et al., 2021). This result suggests the wearing-off effect of GDPR over time.

In general, there is mixed empirical evidence of GDPR's effectiveness on the amount of online tracking. Excluding our study, three out of six papers revealed that the regulation impacted the amount of online tracking by decreasing the sites' tracker usage. All prior studies relied on synthetic crawls as their primary data source, using OpenWPM, WebXRay, or manual implementations that simulate how sites behave when actual consumers visit them (Englehardt & Narayanan, 2016; Libert, 2015; Sanchez-Rola et al., 2019).

Such approaches, however, are (1) limited in capturing the differences in consumer environments (e.g., desktop consumers using various browsers, accessing sites from different regions), (2) have restricted access to "walled gardens" like facebook.com, and (3) can detect higher numbers of trackers (Karaj et al., 2018a; Zeber et al., 2020). For that reason, we use data generated by actual consumers who interacted with the sites, and those sites exposed them to trackers.

Three studies accounted for site visits from within- vs. outside-EU by controlling the consumer base. As Eijk et al. (2019) explain, such control is essential because sites with .com TLD treat EU and US consumers differently with a cookie banner. Samarasinghe & Mannan (2019) and Macbeth (2017) asserted that online tracking varied by consumer's (geo)location, and Fruchter et al. (2015) confirmed that it significantly differed between EU and US consumers. Similarly, few studies use (regression) model estimations, relying on before-/after-GDPR comparisons that do not accurately capture the causal impact of GDPR.

Lastly, previous studies often count a tracker if its domain differs from the domain of a visited site. E.g., if amazon.com (first-party domain) uses amazon-adsystem.com (third-party domain), amazon-adsystem.com is counted as a tracker. We control for such cases to avoid over-counting the number of trackers by looking at the tracker provider of each tracker: e.g., as amazon-adsystem.com is a tracker owned by Amazon, amazon-adsystem.com is not counted as a tracker on amazon.com.

Based on the literature review, we make the following insights and contributions to the literature stream on online privacy:

1. We use a sample of data from the most extensive data set on online tracking, generated by 5 million consumers who browsed about 2.6 billion pages.

2. We differentiate between EU- and US-based consumers accessing 718 sites from April 2018 to December 2019 and differ in the amount of online tracking due to different browsers, browser configurations, and regions.
3. We determine the target audience of each site and – depending on the location of the consumers who accessed it – whether it has to obey the GDPR or not.
4. We utilize the DiD design to evaluate the causal impact of the GDPR on the amount of online tracking.
5. We use two measures for the amount of online tracking, (1) the number of trackers and (2) the number of tracker providers – which captures the firm owning one or multiple trackers.
6. We reveal how the effect of GDPR on the amount of online tracking develops over time and how it differs between different users of trackers (tracker providers).
7. We use three robustness checks that support the results of our primary analysis.

5. Setup of empirical study

5.1. Description of data

As our first data source, we use WhoTracks.me data sets and their monthly information on sites' usage of the amount of online tracking from April 2018 until December 2019. Consumers using the (1) Cliqz browser, (2) Cliqz browser extension for Firefox, and (3) Ghostery browser extension for Firefox, Safari, Chrome, Opera, and Edge browsers generated the data (WhoTracks.me Privacy Team, 2017b).

German consumers primarily used Cliqz, whereas worldwide consumers used Ghostery (WhoTracks.me Privacy Team, 2018). WhoTracks.me combined Cliqz with Ghostery data in February 2018, and this merge of the two data sources caused a slight decrease in the number of trackers on sites in April 2018's data release. This decrease happened because Ghostery consumers blocked more trackers than Cliqz consumers (WhoTracks.me Privacy Team, 2018, 2018).

Consumer distinction per region and longitudinal nature of the WhoTracks.me data sets allow us to compare how online tracking varies between EU and US regions and over time. After merging and cleaning publicly available data sets from GitHub, we created a balanced panel of 718 sites that EU and US consumers accessed each month from April 2018 to December 2019 (T = 21 months).

As EU and US consumers access each site, our observation unit is a site instance. We reduced our observation period until December 2019 as the US state of California introduced California Consumer Privacy Law in January 2020, which might change the amount of tracking for US consumers from that period onward.

As our second data source, we use SimilarWeb. It is challenging to determine whether a site is an "EU" or a "non-EU" firm under GDPR as it is open to consumers worldwide (see Section 3.2). We use the site's TLD to determine whether it targets EU or non-EU consumers and hence, whether it can be considered an "EU" or a "non-EU" firm under the GDPR. However, some sites (e.g., apotheke.com) use international TLDs while still targeting EU consumers (Eijk et al., 2019). To account for cases of sites with international TLDs targeting EU consumers, we combine the site's TLD with the SimilarWeb's data.

SimilarWeb's data allows us to observe the share of traffic from an EU or a non-EU region for a particular site at one moment in time (October 2021). So, we added SimilarWeb's data from October 2021 to our balanced panel to find whether each site received the most traffic from the EU or non-EU region that month. Thus, SimilarWeb's data, combined with the site's TLD, allows us to determine the site's target audience and whether the site can be considered an "EU" or a "non-EU" firm under the GDPR (see Section 5.3).

5.2. Two measures for the amount of online tracking

We are interested in whether GDPR increased consumers' privacy. We measure consumers' privacy using the amount of online tracking: where an increase in the amount of online tracking would mean a decrease in consumers' privacy. We use two measures for the amount of online tracking (1) the number of tracker providers and (2) the number of trackers.

WhoTracks.me differentiates between a (1) site (collection of pages) and a (2) page (a single component of a site). There are three critical considerations how the WhoTracks.me data generated our two measures for the amount of online tracking. First, a single page can contain trackers that lack reach, and Ghostery/Cliqz would not count them. According to WhoTracks.me Privacy Team (2017a), such trackers are present on less than ten different sites or do not track consumers via tracking technologies. So, we would not observe such trackers on sites in the WhoTracks.me data.

Second, Ghostery/Cliqz consumers might have installed additional browser extensions during monthly measurements. Such software could have blocked trackers on pages before Ghostery/Cliqz browser extension would detect them: lowering the detected number of trackers per page and leading to an underestimation in the amount of tracking. We assumed such consumers are equally distributed between the two groups within our research design (see Section 5.3).

Third, WhoTracks.me also reports trackers detected from other browser extensions (e.g., Kaspersky Labs or Adguard) that the consumers installed under the "extensions" tracker category. We removed all observations from trackers categorized as "extensions" as sites do not use such trackers, and we might overestimate the amount of online tracking. Instead, consumers' browser extensions "inject" such trackers into the count of trackers that WhoTracks.me reports.

5.2.1. Number of tracker providers

We use the maximum number of tracker providers as our first measure for the amount of online tracking. This measure allows us to report all tracker providers consumers encountered while browsing a particular site during a specific month.

E.g., Ghostery extension records the different number of tracker providers detected on each xyz.com page, and WhoTracks.me reports all distinct tracker providers seen on a site (e.g., during April 2018, a total of 10 particular tracker providers were detected on the site xyz.com). This statistic is an aggregate measure reporting the maximum number of distinct tracker providers seen across xyz.com pages, generated by about 5 million Ghostery/Cliqz consumers that browsed xyz.com that month. We use this aggregate statistic as the first measure for the amount of online tracking.

5.2.2. Number of trackers

We use the maximum number of trackers for our second measure of the amount of online tracking. So, we observe all trackers that consumers encountered while browsing a particular site during a specific month.

As one tracker provider can own single or multiple trackers, this measure allows us to detect more granular changes in the amount of online tracking. Further, WhoTracks.me offers a categorization of trackers. So, we can observe how the structure of the trackers' market changes over time and what impact GDPR has on different types of trackers that sites use. We explore that question in Section 6.4.3 and describe tracker categories in Section 8.2.

5.3. Description of treatment and control group

As described in Section 3.2, the only case when GDPR does not apply is if a non-EU site processes data of non-EU consumers – in all other cases, a site must comply with the GDPR. Our data allows us to easily distinguish between the EU and non-EU (i.e., US) consumers accessing sites. However, as mentioned, it is challenging to determine whether a site can be counted as an "EU" or a "non-EU" firm because it is accessible to consumers globally.

Nevertheless, we can use a couple of proxies to determine when we should count a site as an "EU" or a "non-EU" firm, such as the site's (1) target audience, (2) (country-code) TLD, (3) cookie banner display, (4) server location, and others. As mentioned in Section 5.1, we use a combination of two proxies, the site's target audience and TLD, for our primary analysis. They account for most wrongly categorized cases of sites (i.e., false positives/negatives). However, Section 6.4.4 illustrates that our results are robust to using cookie banner display and the site's server location as proxies to determine whether a site is an "EU" or a "non-EU" firm under the GDPR.

We consider that a site targets EU consumers if (1) the site uses an EU TLD or (2) the site received the most traffic from the EU region in at least one period. So, if a site uses an international TLD but has received the most traffic from an EU region, it targets EU consumers and can be considered an "EU-firm" under the GDPR. We use stricter criteria to label sites targeting non-EU consumers. A site targets non-EU consumers if (1) the site uses a non-EU TLD and (2) the site received the most traffic from a non-EU region period in at least one period. In comparison to counting "EU" firms, both (1) and (2) must be valid for counting "non-EU" firms.

So, if the site's target audience are non-EU consumers and US consumers visited that site, then the site does not have to obey GDPR. We consider that case a "control group" and all other site instances "treatment group." Table 2 illustrates assigning a site instance to a treatment/control group per cell.

Table 2: Description of treatment assignment framework

	base of consumer	
website target audience	**EU**	**US**
EU[1]	treatment group	treatment group
non-EU[2]	treatment group	control group

[1]Website targets EU consumers if (1) the website uses an EU top-level domain (e.g., .de) or (2) the website received the most traffic from the EU region in at least one period. [2]Website targets non-EU consumers if (1) the website uses a non-EU top-level domain (e.g., .com) and (2) the website received the most traffic from a non-EU region in at least one period.

6. Results of empirical study

6.1. Descriptive results for the sample

First, we describe our sample that contains 718 sites visited by 2 consumer groups (1,436 site instances) over 21 months (N = 30,156). We show the number of site instances per cell in Table 3.

Table 3: Distribution of site instances across audiences and consumer bases

	base of consumer	
website target audience	**EU**	**US**
EU[1]	1,890 (6%)	1,890 (6%)
non-EU[2]	13,188 (44%)	13,188 (44%)
∑	15,078 (50%)	15,078 (50%)

[1]Website targets EU consumers if (1) the website uses an EU top-level domain (e.g., .de) or (2) the website received the most traffic from the EU region in at least one period. [2]Website targets non-EU consumers if (1) the website uses a non-EU top-level domain (e.g., .com) and (2) the website received the most traffic from a non-EU region in at least one period.

Notes: This table shows the number and percentage of observations in our sample per cell. Cell belonging to the control group is colored blue. We colored the treatment group's cells orange. In total, 56% (N = 16,968) of all observations (N = 30,156) make the treatment group and 44% (N = 13,188) the control group.

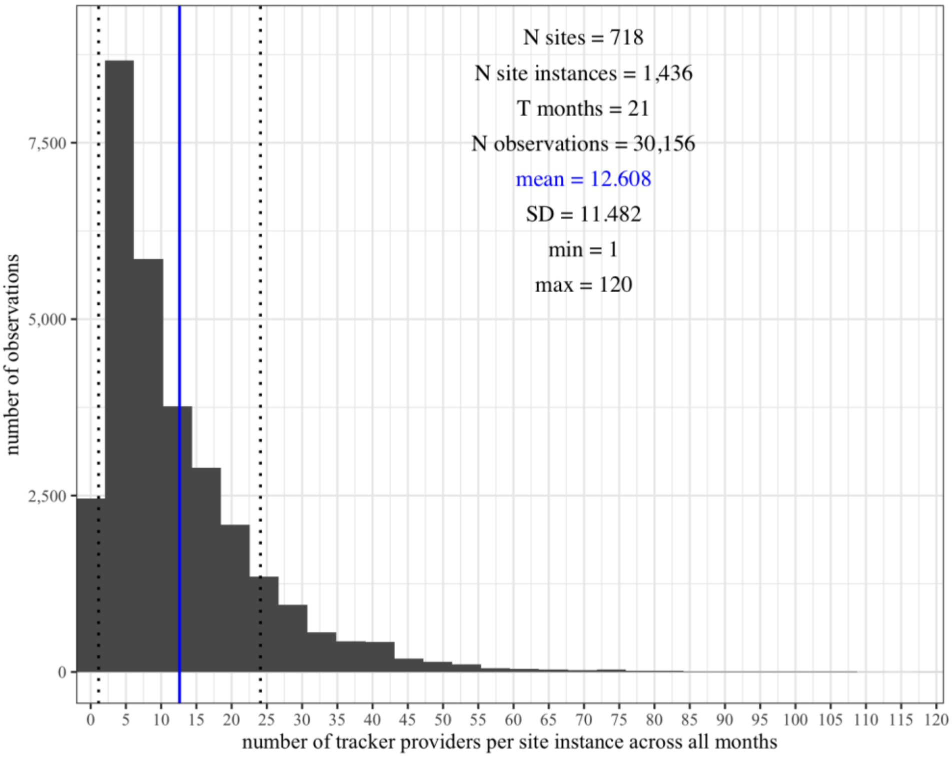
In total, 56% (N = 16,968) of all observations (N = 30,156) make the treatment group and 44% (N = 13,188) the control group. Thus, the share of sites instances between treatment and control groups in our sample are roughly balanced.

6.2. Descriptive results for the amount of online tracking

6.2.1. Descriptive results for the number of tracker providers

Next, we show the descriptive results for our first measure of online tracking, the number of tracker providers, in Figure 1.

Figure 1: Distribution of the number of tracker providers per site instance across all months

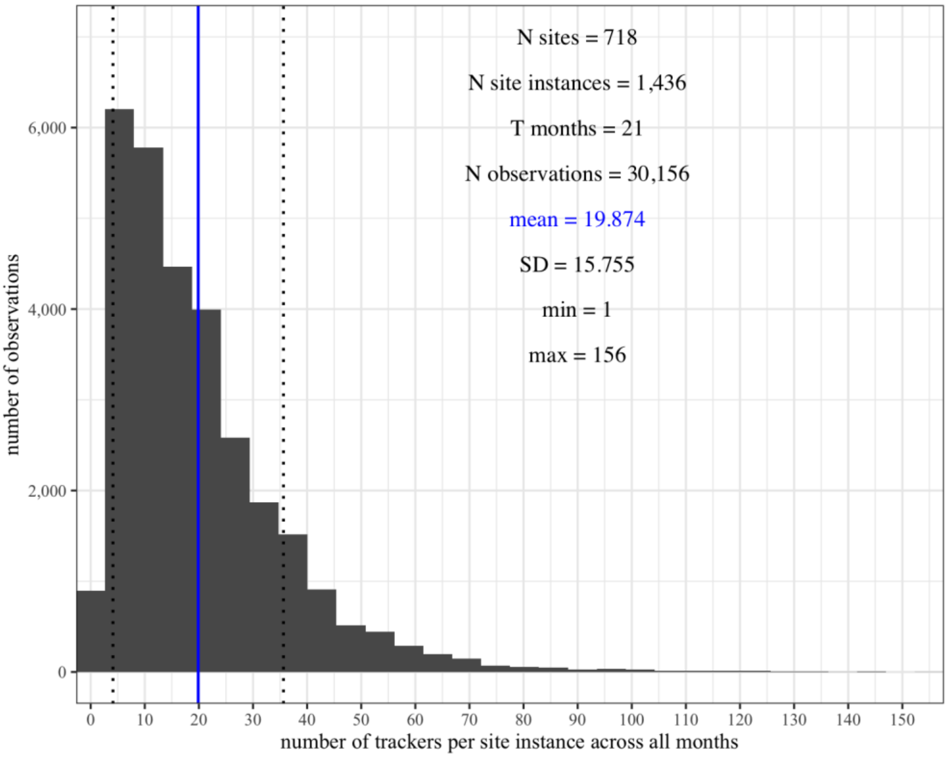


As seen from Figure 1, the distribution of tracker providers is unimodal and right-skewed. So, most site instances typically use from 1 to 5 tracker providers, and a few use many tracker providers. E.g., makeuseof.com, categorized as a news site, had a maximum of 120 different tracker providers from the site instance of US consumers in March 2019. As of December 2021, makeuseof.com does not display a cookie banner to EU consumers as it targets non-EU consumers and has about 13 million unique visitors every month (Makeuseof.com, 2021). The mean number of tracker providers across all site instances and periods equals 12.608 tracker providers (*SD* = 11.482).

6.2.2. Descriptive results for the number of trackers

As before, we show the descriptive results for our second measure of online tracking, the number of trackers, in Figure 2.

Figure 2: Distribution of the number of trackers per site instance across all months



This distribution is also unimodal and right-skewed. As a single tracker provider owns one or several trackers, the maximum number of trackers (*max* = 156) is higher than the maximum number of tracker providers (*max* = 120) reported in Figure 1. The mean number of trackers across all site instances and periods equals 19.874 trackers (*SD* = 15.755). So, site instances of EU and US consumers typically encountered 20 trackers with significant variations when visiting sites in our sample over 21 months. Because Ghostery/Cliqz detects tracker providers on sites containing at least 1 tracker provider, the minimum number of trackers and tracker providers is always 1.

6.3. Descriptive results for the users of online trackers

Next, we show the distributions of trackers between various users of trackers (i.e., site categories) per group in Table 4.

Table 4: Distribution of the number of trackers per site instance in treatment and control groups across users of trackers and all months

	treatment group						control group						difference
users of trackers	**N sites**	**N site instances**	**mean n trackers**	**SD n trackers**	**min n trackers**	**max n trackers**	**N sites**	**N site instances**	**mean n trackers**	**SD n trackers**	**min n trackers**	**max n trackers**	**Δ mean n trackers**
News & Portals	102	2,625	30.276	19.978	1	144	79	1,659	35.937	25.156	1	156	-5.661
Recreation	10	294	20.602	11.086	2	61	6	126	27.183	8.369	2	52	-6.580
Business	247	5,502	19.791	12.306	1	81	232	4,872	20.560	13.521	1	118	-0.770
Entertainment	216	4,851	18.529	13.407	1	100	201	4,221	20.287	16.736	1	137	-1.758
Reference	51	1,218	13.947	9.626	1	79	44	924	16.224	11.297	1	94	-2.277
Adult	92	2,478	9.482	4.685	1	32	66	1,386	7.730	4.163	1	24	1.752
Total	718	16,968	19.141	14.464	1	144	628	13,188	20.818	17.230	1	156	-1.677

Notes: Differences in the distribution of online trackers between treatment and control groups across users of trackers and all months (T = 21 months). We abbreviated the number of trackers to "n trackers." We ordered the table from highest to lowest mean number of trackers for users of trackers in the treatment group.

Most site instances in our observation period contain consumers visiting business sites to purchase items or services online (e.g., amazon.com, nike.com). Following that industry, EU and US consumers also visited entertainment- (e.g., 9gag.com, facebook.com) and news-related (e.g., bbc.com, accuweather.com) sites.

As seen from Table 4, news-related sites tend to use, on average, most trackers in both groups. The maximum number of trackers is also the highest for such users of trackers. We explained why that is the case for news sites in Section 2.2.1. However, the treatment group has 5.661 trackers less on news-related sites than the control group. On average, the treatment group uses 1.667 trackers less than the control group. All users of trackers in the treatment group – except adult sites – tend to use fewer trackers than those in the control group.

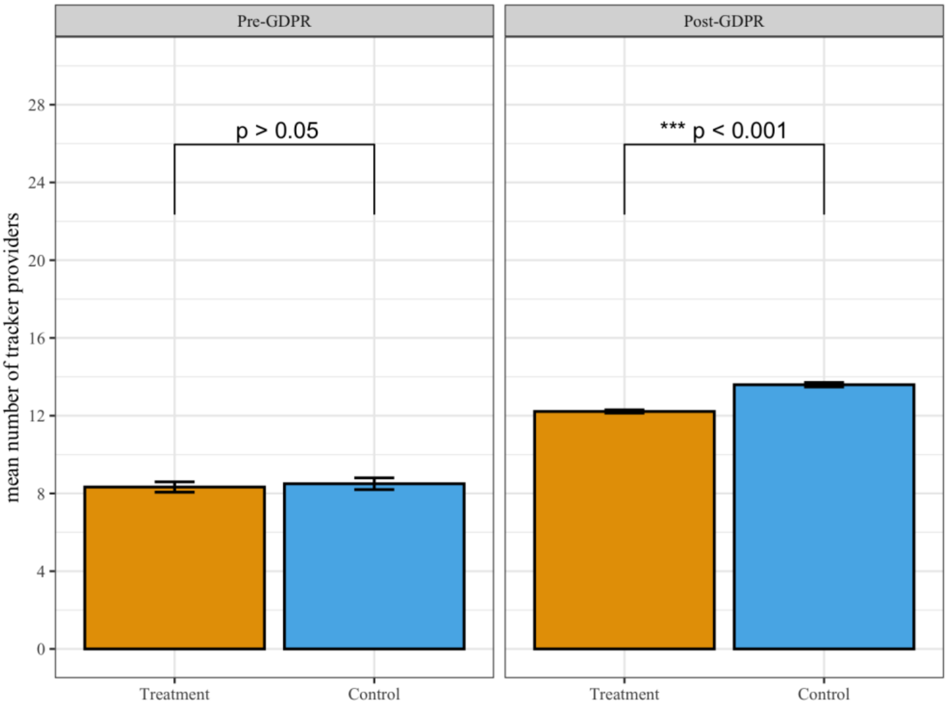
6.4. Effect of GDPR on the amount of online tracking

6.4.1. Effect of GDPR on the number of tracker providers

6.4.1.1. Effect of GDPR on the number of tracker providers pre- and post-GDPR

To investigate the effect of GDPR on the number of tracker providers, we first show how the mean number of trackers differs between the two groups in the pre- and post-GDPR period. We run an independent samples t-test to check whether the group means differ significantly in each of the two periods, and illustrate that in Figure 3.

Figure 3: Number of tracker providers between treatment and control groups pre- and post-GDPR



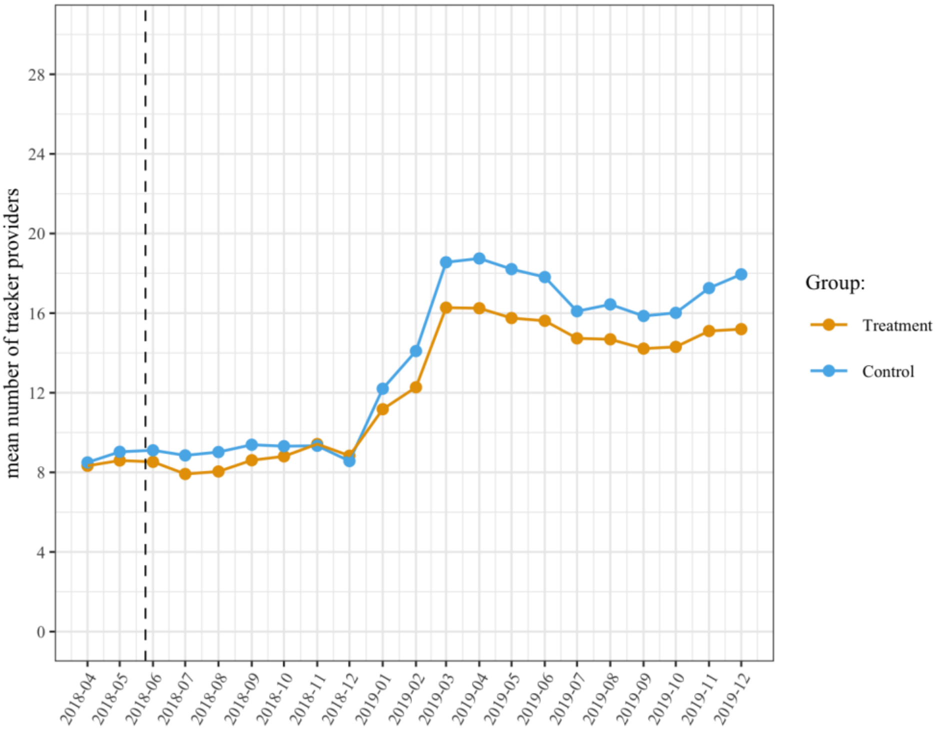
Notes: Independent t-test comparisons between group means in pre- and post-GDPR periods.

The amount of online tracking, measured by the number of tracker providers, does not significantly differ between the groups in the pre-GDPR period, *t*(1347) = 0.421, *p* = 0.674. Both groups use about 8 tracker providers, and the difference between the group means equals 0.169 tracker providers in the pre-GDPR period, Δ = 0.169. However, group means differ significantly in the post-GDPR period, *t*(24143) = 9.753, *p* < 0.001. The treatment group uses, on average, 12.217 tracker providers and the control group 13.593 tracker providers in the post-GPDR period. The difference between group means in the post-GDPR period equals 1.376 tracker providers, Δ = 1.376. So, group means differ significantly after the GDPR's enforcement, and we can observe the general increase in the number of tracker providers from pre- to post-GDPR period.

6.4.1.2. Effect of GDPR on the number of tracker providers over time

After looking at the number of tracker providers between groups in pre- and post-GDPR periods, we show how the amount of tracking developed over time in Figure 4.

Figure 4: Development of the number of tracker providers in treatment and control groups



Looking at Figure 4, the increasing trend of the amount of online tracking is observable (Lerner et al., 2016). Overall, the treatment group's amount of tracking is lower than the control group in each period, except in December 2018. Around the time of GDPR's enforcement, the number of tracker providers in the treatment group slightly decreased, whereas the control group remained relatively unaffected. However, until the end of 2018, the treatment group's amount of tracking returned at the same level as the control group. Macbeth (2017) and Samarasinghe & Mannan (2019) show that sites load different tracker providers for two consumer bases which explains why the levels between the groups differ over time.

Anecdotal evidence suggests that many sites removed tracker providers with questionable privacy policies after GDPR's enforcement to avoid the risk of fines. However, site owners reintroduced such tracker providers due to the lack of GDPR enforcement throughout 2018. Thus, the effect of GDPR on the amount of online tracking might only be of short-term nature. Moreover, both groups increased the number of tracker providers from December 2018 onward, reaching a new equilibrium in March 2019. From that point, the amount of tracking decreases for both groups, perhaps due to a €50,000,000 GDPR fine enforced on Google in January 2019 or an increase in overall GDPR fines throughout 2019 (CMS International Law Firm, 2020; Davies, 2019).

6.4.1.3. Difference-in-differences analysis for the number of tracker providers

Though previous results might indicate the regulation's effect on the amount of online tracking, we continue with the DiD method to answer our first research question. DiD design accounts for changes in the treatment and control groups over time, for which any unobserved influences on the treatment group have the same effect on the control group.

First, we manually calculate the DiD coefficient (average treatment effect, ATE)–the difference between the mean differences in the number of tracker providers of both groups–as illustrated in Table 5.

Table 5: Cross-table for the mean number of tracker providers between treatment and control groups pre- and post-GDPR

Group	Pre-GDPR	Post-GDPR	Difference
Treatment	8.329	12.217	3.887
Control	8.498	13.593	5.094
Difference	-0.169	-1.376	-1.207

The treatment group's mean increased by 3.887 tracker providers, whereas the control group's mean increased by 5.094 tracker providers from pre- to post-GDPR. The control group increased the mean number of tracker providers more than the treatment group, from pre- to post-GDPR. The difference between those two numbers illustrates the effect of GDPR. So, the DiD coefficient equals -1.207 tracker providers, suggesting that the GDPR lowered the average number of tracker providers by 1 tracker provider per site instance: an effect of small magnitude.

In other words, the treatment group would have used 13 tracker providers in the post-GDPR period if the regulators did not enforce GDPR. As they implemented it, the treatment group used 12 tracker providers. This change from, on average, 12 to 13 tracker providers in the treatment group's post-period corresponds to a percent change of -8% and represents an additional way to quantify the effect of GDPR on the amount of online tracking.

After calculating the DiD coefficient, we use the OLS (ordinary least squares) regression to evaluate whether this effect is statistically significant. In addition, OLS regression allows us to control for other factors that might influence the DiD coefficient (e.g., differences between individual sites), and we can obtain the effect's confidence intervals. We present the specification for our OLS model:

$$Y_{i,t} = \beta_0 + \beta_1 Treatment_i + \beta_2 Post_t + \beta_3 Treatment_i Post_t + \omega_t + \mu_i + \epsilon_{i,t} \qquad (1)$$

In Equation (1), our dependent variable for a site instance i at time t is $Y_{i,t}$. $Treatment_i$ is an indicator variable describing whether the site instance i was treated ($Treatment_i = 1$) with GDPR or not ($Treatment_i = 0$). As described in Table 2, the same non-EU targeting site can be treated with GDPR ($Treatment_i = 1$) if EU consumers' site instances visit it; if seen by the US consumers' site instances, the site instance is not treated ($Treatment_i = 0$). $Post_t$ indicates pre-GDPR ($Post_t = 0$) and post-GDPR ($Post_t = 1$) period after–and including–May 2018. We control for period fixed effects ω_t and site instance fixed effects μ_i to account for factors constant over time. So, our baseline model is a standard two-way fixed effects model (TWFE). Additionally, we cluster standard errors

$\epsilon_{i,t}$ at the site instance level to account for autocorrelation for the same sites over time. Our coefficient of interest β_3 measures the mean difference in response between both groups over time.

We present the results of the OLS regressions in Table 6.

Table 6: Result of difference-in-differences analysis for the number of tracker providers

Dependent Variable:	Number of tracker providers per site instance and month
Model:	(1)
Post x Treatment	-1.207*** [-1.860; -0.554]
Period FE	✓
Site Instance FE	✓
N Observations	30,156
R^2	0.760
Within R^2	0.001

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001

One-way standard errors are clustered at the site instance level; 95% confidence intervals are reported in brackets.

Notes: The Post and the Treatment coefficients have been removed from model (1) due to collinearity. The total number of observations (N = 30,156) is the product of the number of site instances (N = 1,444) and periods (T = 21).

The DiD coefficient (β_3 = -1.207, p < 0.001, 95% CI [-1.859; -0.554]) is significant and negative for the treatment group in post-GDPR period in our simple DiD model presented in column (1). The size of this DiD coefficient is the same as the calculated DiD coefficient from Table 5. The estimated effect prevails in the model that controls for period fixed effects in column (2) and our baseline model that additionally controls for site instance fixed effects in column (3).

These results confirm that GDPR had a significant impact on lowering the amount of online tracking by 1 tracker provider on average per site instance (-8%). Although this effect is significant, its size is perhaps negligible for the consumers' privacy. If such a tracker provider is a high market-share player like Google, we would assume a significant impact on the tracking ecosystem and consumers' privacy. However, this scenario is unlikely, as past studies indicated that large tracker providers pressured smaller tracker providers out of the market (Johnson et al., 2021; Peukert et al., 2021).

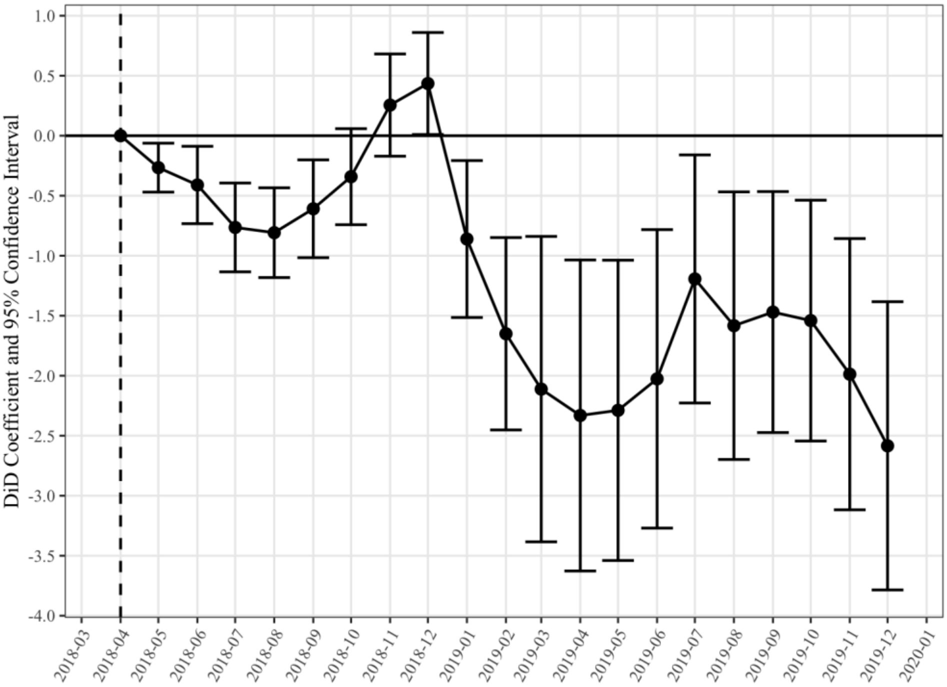
Lastly, we estimate the effect of GDPR on the number of tracker providers over time. For this purpose, we specify a new OLS model:

$$Y_{i,t} = \mu_i + \omega_t + \sum_{j=-m}^{q} \beta_j \, D_{t+j} + \epsilon_{i,t} \qquad (2)$$

In Equation (2), our dependent variable for a site instance i at time t is $Y_{i,t}$. We include site instance fixed effects μ_i and period fixed effects ω_t as in our baseline OLS model in Equation (1). "Leads" of the treatment effect are represented by m and "lags" by q. Our coefficient of interest β_j measures the jth lead m or lag q. D_t is an indicator variable for whether the site instance i got treated in period t (D_t=1) or not ($D_t = 0$). As before, we cluster standard errors $\epsilon_{i,t}$ at the site instance level.

We specified Equation (2) to estimate the monthly effects of GDPR. We interact all pre- ("leads") and post-GDPR ("lags") period indicators with the treatment indicator while setting the April 2018 period–immediately before GDPR–as a reference period. We plot the OLS regressions' DiD coefficients with 95% confidence intervals in Figure 5.

Figure 5: Development of the difference-in-differences coefficients for the number of tracker providers



Notes: The model includes period and site instance fixed effects. One-way standard errors are clustered at the site instance level. The first period (April 2018) before GDPR enforcement was excluded from the estimation as a reference period.

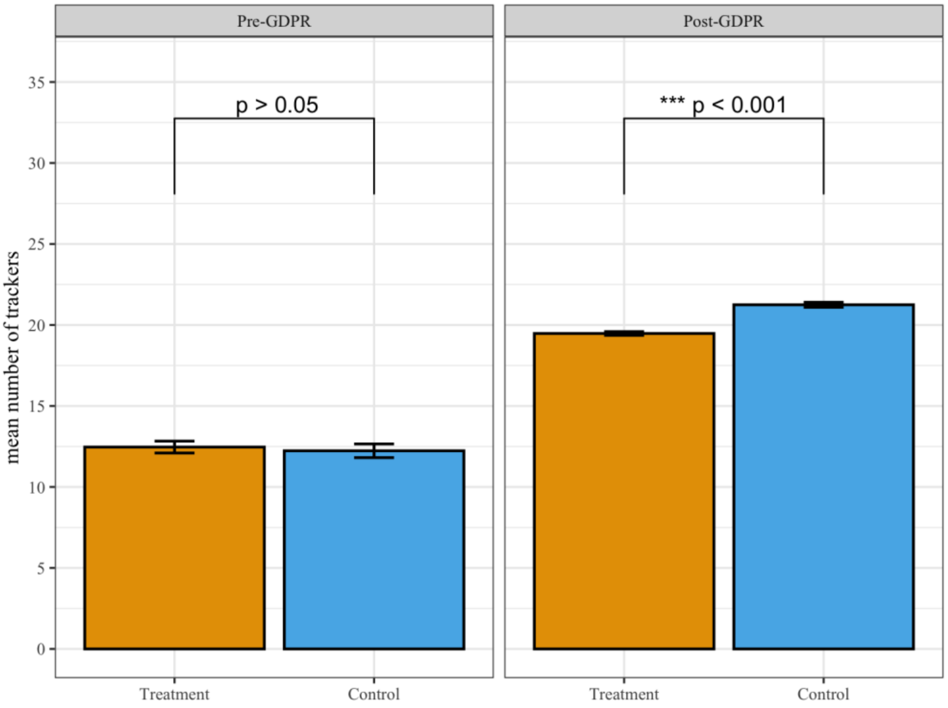
Figure 5 reveals the short-term and long-term effects of GDPR on the number of tracker providers. The level of the short-term impact is highest in August 2018, when it starts to revert. However, from January 2019 onward, the effect becomes prominent again and remains significant. So, the impact of GDPR comes in two waves. As mentioned in Section 6.4.1.2, the development of GDPR fines or the €50,000,000 Google fine in January 2019 could drive such results. We explore the driving force of the GDPR's effect in Section 8.6.

6.4.2. Effect of GDPR on the number of trackers

6.4.2.1. Effect of GDPR on the number of trackers pre- and post-GDPR

To answer our second research question, we determine the effect of GDPR on the number of trackers. As in Section 6.4.1.1, we first determine how the amount of online tracking changed in the pre- and post-GDPR period per group (Figure 6).

Figure 6: Number of trackers between treatment and control groups pre- and post-GDPR



Notes: Independent t-test comparisons between group means in pre- and post-GDPR periods.

The amount of online tracking, measured by the number of trackers, does not significantly differ between the groups in the pre-GDPR period, *t*(1338) = -0.412, *p* = 0.680. Both groups use about 12 trackers, and the difference between the group means equals -0.230 trackers in the pre-GDPR period, Δ = -0.230. However, group means differ significantly in the post-GDPR period, *t*(24353) = 9.193, *p* < 0.001. The treatment group uses, on average, 19.475 trackers and the control group 21.247 trackers in the post-GPDR period. The difference between group means in the post-GDPR period equals 1.773 trackers, Δ = 1.773. So, group means differ significantly after the GDPR's enforcement, and we can observe the general increase in the number of trackers from pre- to post-GDPR period once again.

6.4.2.2. Effect of GDPR on the number of trackers over time

Section 6.4.1.2 shows how the amount of tracking, measured by the number of tracker providers, developed over time. We illustrate such development for the number of trackers in Figure 7.

Figure 7: Development of the number of trackers in treatment and control groups

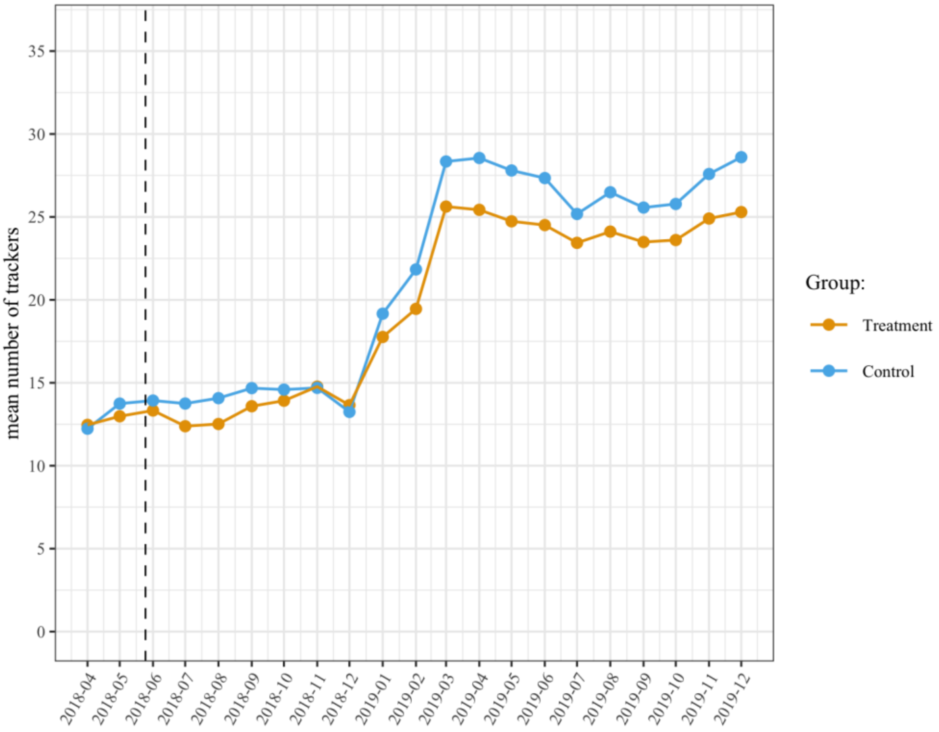


Figure 7 mirrors Figure 4. After GDPR's enforcement, the treatment group's mean decreases. In contrast, the control group's mean remains largely unaffected after regulators enforced the law. By December 2018, however, the treatment group's mean levels with the mean of a control group. The amount of online tracking increased in both groups in the December 2018-March 2019 period, followed by the decrease in the number of trackers.

6.4.2.3. Difference-in-differences analysis for the number of trackers

As in Section 6.4.1.3, we calculate the DiD coefficient for the number of trackers as our dependent variable and present the result in Table 7.

Table 7: Cross-table for the mean number of trackers between treatment and control groups pre- and post-GDPR

Group	**Pre-GDPR**	**Post-GDPR**	**Difference**
Treatment	12.463	19.475	7.012
Control	12.232	21.247	9.015
Difference	0.230	-1.773	-2.003

The treatment group's mean increased by 7.012 trackers, whereas the control group's mean increased by 9.015 trackers from pre- to post-GDPR. The control group increased the mean number of trackers more than the treatment group, from pre- to post-GDPR. So, the differences-in-differences (DiD) co-efficient equals -2.003 trackers, suggesting that the GDPR lowered the average number of trackers by 2 trackers (-9%) per site instance.

We use the OLS regression specification from Equation (1) to calculate the DiD coefficient for the number of trackers instead of tracker providers. We present the results in Table 8.

Table 8: Result of difference-in-differences analysis for the number of trackers

Dependent Variable:	Number of trackers per site instance and month
Model:	(1)
Post x Treatment	-2.003*** [-2.925; -1.081]
Period FE	✓
Site Instance FE	✓
N Observations	30,156
R^2	0.776
Within R^2	0.001

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001

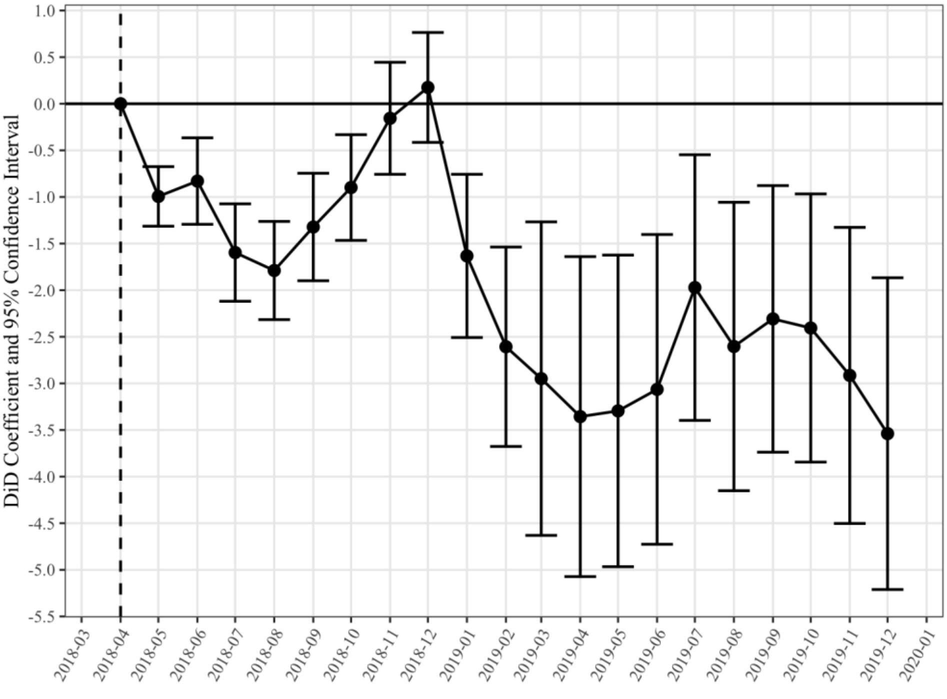
One-way standard errors are clustered at the site instance level; 95% confidence intervals are reported in brackets.

Notes: The Post and the Treatment coefficients have been removed from model (1) due to collinearity. The total number of observations (N = 30,156) is the product of the number of site instances (N = 1,444) and periods (T = 21).

The DiD coefficient (β_3 = -2.003, p < 0.001, 95% CI [-2.925; -1.081]) is significant and negative for the treatment group in post-GDPR period in our simple DiD model presented in column (1). The size of this DiD coefficient is the same as the calculated DiD coefficient from Table 7. The estimated effect prevails in the model that controls for period fixed effects in column (2) and our baseline model that additionally controls for site instance fixed effects in column (3). These results confirm that the GDPR lowered the average number of trackers by 2 per site instance (-9%).

As before, we estimate the development of the GDPR's effect on the number of trackers per month using the model specification in Equation (2). We do so by changing the response variable from the number of tracker providers to the number of trackers. Figure 8 presents the results of these estimations.

Figure 8: Development of the difference-in-differences coefficients for the number of trackers



Notes: The model includes period and site instance fixed effects. One-way standard errors are clustered at the site instance level. The first period (April 2018) before GDPR enforcement was excluded from the estimation as a reference period.

Figure 8 mirrors Figure 5. It, too, reveals the short-term and long-term effects of GDPR on the number of trackers. The level of the short-term impact is highest in August 2018, when it starts to revert.

However, from January 2019 onward, the effect becomes prominent again and remains significant: illustrating the two-wave impact of GDPR once again.

6.4.3. Differences in the effect of GDPR

Lastly, we show how GDPR impacted users of trackers in our sample. Similarly, we show how the effect differs for different categories of trackers, mentioned in Section 2.3. Table 9 reveals the distribution of the GDPR's impact across (1) users of tracker providers, (2) users of trackers, and (3) categories of trackers.

Table 9: Distribution of the effect of GDPR

Across Users of Tracker Providers			Across Users of Trackers			Across Categories of Trackers		
Users of Tracker Providers	**DiD Coefficient**	**95% CI**	**Users of Trackers**	**DiD Coefficient**	**95% CI**	**Category of Trackers**	**DiD Coefficient**	**95% CI**
Adult	-0.925	[-1.472; -0.378]	Entertainment	-2.056	[-3.658; -0.454]	Advertising	-1.256	[-1.838; -0.675]
Entertainment	-1.241	[-2.386; -0.096]	Business	-1.408	[-2.701; -0.115]	Essential	-0.471	[-0.752; -0.189]
News & Portals	-3.047	[-5.865; -0.228]	News & Portals	-3.916	[-7.677; -0.155]	Analytics	-0.202	[-0.358; -0.047]
Reference	-1.180	[-2.984; 0.623]	Adult	-0.938	[-1.892; 0.017]	Other	-0.070	[-0.141; 0.001]
Business	-0.525	[-1.381; 0.332]	Reference	-2.458	[-5.014; 0.098]	Comments	0.007	[-0.013; 0.027]
Recreation	1.229	[-3.691; 6.148]	Recreation	2.290	[-5.049; 9.63]	Social Media	-0.011	[-0.074; 0.052]

One-way standard errors are clustered at the site instance level; 95% confidence intervals are reported in brackets.

Notes: The model includes period and site instance fixed effects. We ordered the table by the upper confidence interval of the model estimate.

Looking at the users of tracker providers, GDPR had an effect on adult- (β_3 = -0.925, 95% CI [-1.472; -0.378]), entertainment- (β_3 = -1.241, 95% CI [-2.386; -0.096]) and news-related (β_3 = -3.047, 95% CI [-5.865; -0.228]) sites. So, the GDPR had the largest effect on news-related sites, lowering their use of tracker providers by 3 tracker providers. We provide the full description of users of trackers in the Section 8.1.

Similar to users of tracker providers, we look at the users of trackers. There, GDPR had an effect on entertainment- (β_3 = -2.056, 95% CI [-3.658; -0.454]), business- (β_3 = -1.408, 95% CI [-2.701; -

0.115]), and news-related (β_3 = -3.916, 95% CI [-7.677; -0.155]) sites. Again, the GDPR had the largest effect on news-relates sites, lowering the number of trackers by 4 trackers.

Lastly, GDPR impacted the site's usage of advertising (β_3 = -1.256, 95% CI [-1.838; -0.675]), essential (β_3 = -0.471, 95% CI [-0.752; -0.189]) and analytics (β_3 = -0.202, 95% CI [-0.358; -0.047]) trackers. However, the levels of GDPR's effect on each of those three kinds of trackers are rather low. As mentioned, we describe such kinds of trackers in Section 8.2.

6.4.4. Insights from the robustness tests

We performed three robustness checks in Section 8 to support the results of our primary analysis. This section summarizes the aim and outcome of each robustness check.

First, in Section 8.3, we broke down the effect of GDPR on different cells in our treatment assignment framework (see Table 2). E.g., we only compared cell one (i.e., EU consumers visiting sites that target EU consumers) to cell four (i.e., US consumers visiting sites that target non-EU consumers) of our treatment assignment framework. We refer to that comparison as the "cleanest" comparison of the treatment/control group. We found a significant negative effect of GDPR on the number of trackers in that case (β_3 = -4.693, $p < 0.001$, 95% CI [-6.445; -2.940]).

Second, in Section 8.4, we assigned a new "treatment" to sites based on (1) whether the site showed a cookie banner to EU consumers and (2) if the site's server location was within the EU region. Treating EU consumers with a cookie banner is another way a site might demonstrate a willingness to comply with GDPR. We found that 4% of all sites visited by EU consumers did not serve EU consumers' cookie banners, even though such sites target EU audiences. On the other hand, we found evidence of the small degree of the Brussels effect – the spillover effect of GDPR on other countries (Bradford, 2020; Peukert et al., 2021; Sanchez-Rola et al., 2019). We discovered that 9% of all sites visited by US consumers served US consumers cookie banners even though they target non-EU consumers and do not have to obey GDPR.

Using a new treatment assignment framework, we still found a significant negative effect of GDPR on the number of trackers (β_3 = -1.438, p < 0.001, 95% CI [-2.119; -0.7571]). Similarly, using a site's server location to assign a "treatment" to each site, we also found a significant negative effect of GDPR on the number of trackers (β_3 = -1.629, p < 0.05, 95% CI [-2.333; -0.9258]).

Lastly, the limitation of the WhoTracks.me data set — which we use for our principal analysis – is that it lacks a longer pre-GDPR period. A more extended pre-GDPR period is vital for a DiD analysis because it allows us to infer the causal effect of GDPR on the amount of online tracking. We used a different sample of the WhoTracks.me data set to provide suggestive visual evidence that parallel-group means in the pre-GDPR period held in Section 8.5. Additionally, we ran our OLS model specification from Equation (1) – with minor adjustment – on this data set and found a significant negative effect of GDPR on the number of trackers (β_3 = -5.871, p < 0.001, 95% CI [-8.866; -2.875]).

7. Summary and conclusion

Our empirical study aimed to investigate if GDPR's enforcement increased consumers' online privacy. To measure consumers' privacy, we focused on the amount of online tracking. We considered that consumers' privacy would increase if consumers encountered fewer tracker providers and trackers after GDPR's enforcement. Therefore, we used two measures for the amount of online tracking: (1) number of tracker providers and (2) number of trackers.

Our results show that the amount of online tracking, measured by the number of tracker providers and the number of trackers, increased over time; sites that comply with GDPR use 12 tracker providers (19 trackers) on average in the post-GDPR period. Without GDPR, they would have used 13 tracker providers (21 trackers). So, GDPR's enforcement decreased the average number of tracker providers (trackers) on each site by 1 tracker provider or -8% (2 trackers or -9%), having a minor positive impact on consumers' online privacy.

The negative effect of the GDPR's enforcement on the amount of online tracking comes in two waves. Three months after the law's enforcement (August 2018), the weaker negative impact disappears. But five months later (January 2019), the more substantial negative effect lasts until December 2019. An increase in the number of GDPR fines throughout 2019 or a €50,000,000 penalty enforced on Google in January 2019 could have caused such development of the impact.

Lastly, we find that the enforcement of the law decreased advertising, analytics, and essential trackers on news, entertainment, and business sites – corroborating a minor positive impact on consumers' privacy.

We conclude that the GDPR's enforcement increased consumers' privacy to a minor extent. But the increasing trend of the amount of online tracking poses a justifiable privacy concern for consumers. We believe our findings have important implications for consumers and regulators.

First, we show that consumers, on average, encounter more tracker providers and trackers on sites after GDPR. As an increase in consumer tracking could imply an increase in profiling and (re-)targeting, consumers face a trade-off between (1) a better personalized online experience or (2) a loss of privacy after GDPR. Second, we show that GDPR positively affected consumers' online privacy, albeit to a small extent. GDPR redefined the notion of consent in cookie banners as a method by which consumers can regulate the amount of online tracking. We suggest that consumers take advantage of that tool if they are concerned about their online privacy.

Second, European privacy regulators could benefit from our study as we show the positive yet minor impact of GDPR on consumers' privacy. We think our findings could also support the design of the upcoming ePrivacy Regulation and other upcoming privacy laws worldwide.

8. Web Appendix

8.1. Description of users of online trackers

Table 10: Description of users of trackers

	users of trackers	description of users of trackers	examples of users of trackers
1	Adult	Sites that are generally thought not to be appropriate for children.	redtube.com, pornhub.com, sex.com
2	Banking	Sites of banks.	americanexpress.com, deustche-bank.de, commerzbank.de
3	Business	Sites with a physical location providing the option to purchase items online. Official company sites and sites selling services fall within this category, too.	apple.com, paypal.com, airbnb.com
4	E-Commerce[1]	Shops whose sites allow purchasing items online without having a physical store.	amazon.com, nike.com, alibaba.com
5	Entertainment	Social networks and dating sites, online games, video-sharing and streaming, and TV channels not focusing on the news.	9gag.com, facebook.com, spotify.com
6	Government	Official sites of political parties and movements.	nasa.gov, nih.gov, oevp.at
7	News & Portals	News providers and multipurpose portals, weather forecast sites, TV channels, official sites of cities, and sports associations.	bbc.com, accuweather.com, live.com
8	Recreation	Sites where it is possible to book holidays or flights.	britishairways.com, booking.com, lufthansa.com
9	Reference	Search engines, wiki forums and communities, and online dictionaries.	google.com, wikipedia.org, dictionary.com

[1]We placed E-Commerce and Business sites in the same "Business" site category.

Notes: Table adapted from Karaj et al. (2018) and ordered alphabetically by site category.

8.2. Description of categories of online trackers

Table 11: Description of categories of trackers

trackers supporting		our tracker category	WhoTracks.me tracker category	WhoTracks.me tracker category description	examples of trackers
Advertising	1	Advertising	Advertising	Provides advertising or advertising-related services such as data collection, behavioral analysis, or retargeting.	DoubleClick, ShareThis, Experian Marketing Services
			Adult Advertising	Delivers advertisements that generally appear on sites with adult content.	Adult Webmaster Empire, ExoClick, JuicyAds
User Experience	2	Analytics	Site Analytics	Collects and analyses data related to site usage and performance.	Google Analytics, Adbrain, Piwik Pro
	3	Social Media	Social Media	Integrates features related to social media sites.	Facebook Social Plugins, Giphy, Twitter
	4	Comments	Comments	Enables comments sections for articles and product reviews.	Disqus, eKomi, Livefyre
	5	Essential	Customer Interaction	Includes chat, email messaging, customer support, and other interaction tools.	PayPal, Google Translate, LiveChat
			Essential	Includes tag managers, privacy notices, and technologies that are critical to the functionality of a website.	OneTrust, Google Tag Manager, IAB Consent
			Audio Video Player	Enables websites to publish, distribute, and optimize video and audio content.	YouTube, Twitch, Spotify
			CDN	Content delivery network (CDN) delivers resources for different site utilities and usually for many other customers.	Amazon CDN, CloudFlare, jQuery
			Hosting	Service used by the content provider or site owner.	Github Pages, FastPic, Amazon CloudFront
	6	Other	Misc.	This tracker does not fit in other categories.	Autoscout24, Oracle RightNow, Vinted
			Extensions[1]	Man-in-the-middle (MITM) trackers insert additional requests into pages by intercepting network traffic on a device or using browser extensions.	Kaspersky Labs, Adguard, Yandex Advisor
			Unknown	This tracker has either not been labeled yet or does not have enough information to mark it.	boudja.com, xen-media.com, statsy.net

[1] We removed such trackers from our sample, as site owners do not place them on the sites, nor are they added to sites from other trackers (i.e., "piggybacking" trackers).

Notes: Table adapted from Karaj et al. (2018).

8.3. Robustness test regarding the breakdown of effect on different treatment groups

We presented our treatment assignment framework in Section 5.3. In this section, we explore how the effect of GDPR varies between each cell of our treatment assignment framework by redefining treatment and control groups. Using this approach, we can explore how GDPR's impact differs between the "cleanest" comparison in our treatment assignment framework: EU consumers accessing sites that target EU consumers and US consumers accessing sites that target non-EU consumers.

So, we run our baseline OLS regression specification from Equation (1) per different cell in our treatment assignment framework (see Table 2), thereby redefining treatment/control group comparisons. We present the results in Table 12.

Table 12: Result of difference-in-differences analysis for the number of trackers between different combinations of treatment and control groups

Dependent Variable:	Number of trackers per site instance and month		
Model:	(1) Cell 1 vs. Cell 4	(2) Cell 3 vs. Cell 4	(3) Cell 2 vs. Cell 4
Illustration:	🟧⬜ ⬜🟦	⬜⬜ 🟧🟦	⬜🟧 ⬜🟦
Post x Treatment	-4.693*** [-6.445; -2.940]	-1.674*** [-2.663; -0.685]	-1.607 [-3.378; 0.163]
Period FE	✓	✓	✓
Site Instance FE	✓	✓	✓
N Observations	15,078	26,376	15,078
R^2	0.787	0.775	0.783
Within R^2	0.002	0.001	0.000

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

One-way standard errors are clustered at the site instance level; 95% confidence intervals are reported in brackets.

Notes: The difference-in-differences coefficients between different cells in our treatment assignment framework. The Post and the Treatment coefficients have been removed from all models due to collinearity. The total number of observations for each model is the product of the number of site instances per cell(s) in our sample and periods (T = 21).

In column (1), we compare cell 1 (i.e., EU consumers visiting sites that target EU consumers) to cell 4 (i.e., US consumers visiting sites that target non-EU consumers). In this case, we define cell 1 as our treatment and cell 4 as a control group. We refer to this comparison as the "cleanest" comparison between the treatment and the control group. We find a significant negative effect of GDPR on the number of trackers in this case (β_3 = -4.693, p < 0.001, 95% CI [-6.445; -2.940]).

Next, we compare cell 3 (i.e., EU consumers visiting sites that target non-EU consumers) to cell 4 (i.e., US consumers visiting sites that target US consumers) in column (2). Here, we define cell 3 as our treatment and cell 4 as a control group. We find significant negative effect of GDPR on the number of trackers in this case (β_3 = -1.674, p < 0.001, 95% CI [-2.663; -0.6853]).

Lastly, in column (3), we compare cell 2 (i.e., US consumers visiting sites that target EU consumers) to cell 4 (i.e., US consumers visiting sites that target US consumers). In this case, we define cell 2 as our treatment and cell 4 as a control group. In this case, we do not find a significant negative effect of GDPR on the number of trackers (β_3 = -1.607, p = 0.075, 95% CI [-3.378; 0.1634]).

8.4. Robustness test regarding the comparison of effect to different treatment assignments

Section 5.3 explained that different proxies exist to determine if the site is an "EU" or a "non-EU" firm under the GDPR. We explore two such proxies in this section: the display of a cookie banner and the site's server location.

We assume that a site that does not treat EU consumers with a cookie banner signals that it does not comply with GDPR. Likewise, if a site treats an EU (or the US) consumer with a cookie banner, it likely tries to comply with GDPR and makes our new "treatment" group.

We visited each site in our sample in September 2021, from and outside the EU. Afterward, we took a screenshot of the site's landing page on which the site had a chance to display a cookie banner. We used a VPN service to simulate visiting a site outside the EU. We also gave the site sufficient time to load a cookie banner, especially when using the VPN service. Afterward, we manually coded whether each site displayed a cookie banner to EU or US consumers. We show the result in Table 13.

Table 13: Number of websites displaying cookie banners to consumers depending on consumer base and website target audience

		base of consumer	
website target audience	**website displays cookie banner**	**EU**	**US**
EU[1]	✓	61 (4%)	41 (3%)
	✗	29 (2%)	49 (3%)
non-EU[2]	✓	419 (29%)	136 (9%)
	✗	209 (15%)	492 (34%)
∑		718 (50%)	718 (50%)

[1]Website targets EU consumers if (1) the website uses an EU top-level domain (e.g., .de) or (2) the website received the most traffic from the EU region in at least one period. [2]Website targets non-EU consumers if (1) the website uses a non-EU top-level domain (e.g., .com) and (2) the website received the most traffic from a non-EU region in at least one period.

Notes: This table shows the number and percentage of sites in our sample that do (not) treat EU or non-EU consumers with a cookie banner, depending on if they target EU or non-EU audiences. We visited each site in our sample in September 2021 to determine if the sites serve cookie banners to consumers.

61 out of 718 sites (4%) visited by EU consumers did not serve them cookie banners, even though such sites target EU audiences. EU consumers also visited 419 sites that target non-EU consumers, but those sites still showed them a cookie banner to demonstrate GDPR compliance. 209 sites did not give EU consumers the chance to exercise their privacy by regulating the amount of online tracking in cookie banners.

On the other hand, US consumers accessed 136 sites out of 718 sites (9%) that target non-EU consumers, yet those sites still complied with GDPR – even though they did not have to obey GDPR. This example illustrates the Brussels' effect, where a European regulation has a "spillover" effect on other countries.

Next, we describe this sample's treatment framework. We show the result in Table 14.

Table 14: Distribution of site instances across cookie banner treatment assignment and consumer bases

	base of consumer	
website displays cookie banner	**EU**	**US**
✓	10,080 (33%)	3,717 (12%)
✗	4,998 (17%)	11,361 (38%)
∑	15,078 (50%)	15,078 (50%)

Notes: This table shows the number and percentage of observations in our sample with a cookie banner treatment assignment per cell. Cell belonging to the control group is colored blue. We colored the treatment group's cells orange. In total, 62% (N = 18,795) of all observations (N = 30,156) make the treatment group and 38% (N = 11,361) the control group.

62% (N = 18,795) of all observations (N = 30,156) make the treatment group and 38% (N = 11,361) the control group in this treatment assignment framework. As before, we use the baseline OLS regression model specified in Equation (1) to estimate the effect of GDPR on the number of trackers. The result can be seen in Table 15.

Table 15: Result of difference-in-differences analysis for the sample with a cookie banner treatment assignment

Dependent Variable:	Number of trackers per site instance and month
Model:	(1)
Treatment	-0.025 [-0.755; 0.704]
Post x Treatment	-1.438*** [-2.119; -0.757]
Period FE	✓
Site FE	✓
N Observations	30,156
R^2	0.729
Within R^2	0.006

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001

One-way standard errors are clustered at the site level; 95% confidence intervals are reported in brackets.

Notes: The difference-in-differences coefficients for treatment assignment using cookie banner presence or absence. Post coefficient has been removed from the model (1) due to collinearity. The total number of observations (N = 30,156) is the product of the number of site instances (N = 1,436) and periods (T = 21).

The DiD coefficient (β_3 = -1.438, p < 0.001, 95% CI [-2.119; -0.7571]) is significant and negative for the treatment group in post-GDPR period in our simple DiD model presented in column (1). The estimated effect prevails in the model that controls for period fixed effects in column (2) and our baseline model that additionally controls for site instance fixed effects in column (3). This result confirms that GDPR had a negative impact on the amount of online tracking if we used cookie banner display as treatment assignment criteria.

After using a cookie banner display to determine if a site is an "EU" or a "non-EU" firm, we use the site's server location. If the site's server location resolves within the EU region, we can consider it an "EU" firm. If not, and the site's server is outside the EU, we think it is a "non-EU" firm.

As before, we show how the distribution of our sample's site instances changes under the new treatment assignment framework. We present the result in Table 16.

Table 16: Distribution of site instances across server location treatment assignment and consumer bases

	base of consumer	
website server location	**EU**	**US**
EU	3,822 (13%)	3,822 (13%)
non-EU	11,256 (37%)	11,256 (37%)
∑	15,078 (50%)	15,078 (50%)

Notes: This table shows the number and percentage of observations in our sample with a server location treatment assignment per cell. Cell belonging to the control group is colored blue. We colored the treatment group's cells orange. In total, 63% (N = 18,900) of all observations (N = 30,156) make the treatment group and 37% (N = 11,256) the control group.

63% (N = 18,900) of all observations (N = 30,156) make the treatment group and 37% (N = 11,256) the control group in this new treatment assignment framework. We use the baseline OLS regression model specified in Equation (1) to estimate the effect of GDPR on the number of trackers. The result can be seen in Table 17.

Table 17: Result of difference-in-differences analysis for the sample with a server location treatment assignment

Dependent Variable:	Number of trackers per site instance and month
Model:	(1)
Treatment	0.674 [-0.089; 1.436]
Post x Treatment	-1.629*** [-2.333; -0.926]
Period FE	✓
Site FE	✓
N Observations	30,156
R^2	0.728
Within R^2	0.003

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001

One-way standard errors are clustered at the site level; 95% confidence intervals are reported in brackets.

Notes: The difference-in-differences coefficients for treatment assignment using cookie banner presence or absence. Post coefficient has been removed from the model (1) due to collinearity. The total number of observations (N = 30,156) is the product of the number of site instances (N = 1,436) and periods (T = 21).

The DiD coefficient (β_3 = -1.629, p < 0.001, 95% CI [-2.333; -0.9258]) is significant and negative for the treatment group in post-GDPR period in our simple DiD model presented in column (1). The estimated effect prevails in the model that controls for period fixed effects in column (2) and our baseline model that additionally controls for site instance fixed effects in column (3). This result confirms that GDPR had a negative impact on the amount of online tracking if we used server location as treatment assignment criteria.

8.5. Robustness test regarding the parallel trends' assumption

As explained in Section 5.1, we use a sample of the WhoTracks.me data to observe how the amount of online tracking differs between the EU and US consumer groups accessing 718 sites over 21 months. However, WhoTracks.me also provides a sample of data for "global" consumers, whose location we cannot determine. But that sample of data is available from May 2017 onward (T = 32 months). So, we can use it to test if the parallel trends assumption holds between the group means in the pre-GDPR period.

First, we find the same sites present in our primary sample as we want to compare how tracking changes for those sites in our preliminary analysis. Using that approach, we find 359 out of 718 sites present in our primary (EU/US) sample, and which we track over 32 months.

Second, as we cannot determine the consumer location for this sample, we split the observations into treatment/control groups based on the site's target audience (see Section 5.3). So, a site targeting EU consumers should comply with GDPR, whereas a non-EU targeting site does not have to obey GDPR. We present the descriptives of this sample in Table 18.

Table 18: Distribution of site instances across audiences and consumer bases for the sample with a longer pre-GDPR period

website target audience	number and percentage of observations
EU[1]	2,176 (19%)
non-EU[2]	9,312 (81%)
∑	11,488 (100%)

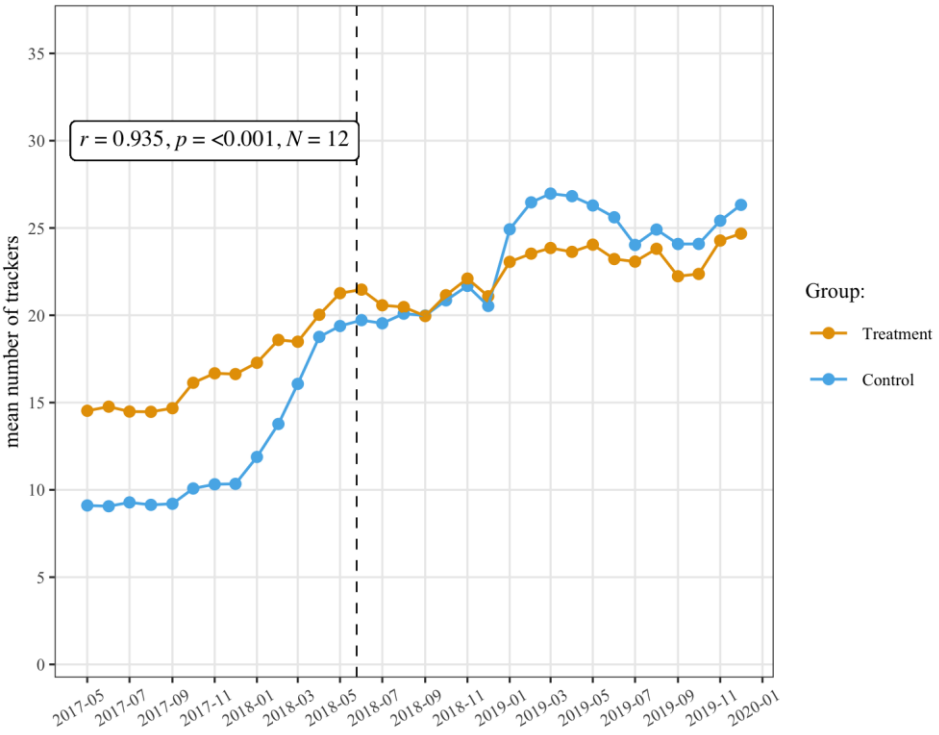
[1]Website targets EU consumers if (1) the website uses an EU top-level domain (e.g., .de) or (2) the website received the most traffic from the EU region in at least one period. [2]Website targets non-EU consumers if (1) the website uses a non-EU top-level domain (e.g., .com) and (2) the website received the most traffic from a non-EU region in at least one period.

Notes: This table shows the number and percentage of observations in our sample with a longer pre-GDPR period per cell. The consumer base is omitted from this table as the sample does not provide information on which consumer base accessed sites. Cell belonging to the control group is colored blue. We colored the treatment group cell orange. In total, 19% (N = 2,176) of all observations (N = 11,488) make the treatment group and 81% (N = 9,312) the control group.

As seen from Table 18, 81% of all observations in this sample make the control group, and 19% the treatment group.

Next, we plot the development of the amount of online tracking, measured by the number of trackers, for those two groups over time. We also measure the correlation between the two group means in the pre-GDPR period to support the assumption of the parallel trends. This result is visible in Figure 9.

Figure 9: Development of the number of trackers in treatment and control groups for the sample with a longer pre-GDPR period



Group means move similarly over time in the pre-GDPR period. This relationship is also quantified by the strong positive correlation coefficient, $r = 0.935$, $p < 0.001$, $N = 12$.

Additionally, we calculate the effect of GDPR on the number of trackers using the model specification from Equation (1) for this sample with a longer pre-GDPR period. We make a single adjustment; we use the site instead of a site instance as our unit of observation in this case. As seen in Table 19, we find significant negative effect of GDPR on the number of trackers (β_3 = -5.871, $p < 0.001$, 95%

CI [-8.866; -2.875]). However, as we cannot control for the consumer base, in this case, the model likely overestimates the level of the effect.

Table 19: Result of difference-in-differences analysis for the number of trackers in the sample with a longer pre-GDPR period

Dependent Variable:	Number of trackers per site instance and month
Model:	(1)
Post x Treatment	-5.871*** [-8.866; -2.875]
Period FE	✓
Site FE	✓
N Observations	11,488
R^2	0.745
Within R^2	0.020

Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

One-way standard errors are clustered at the site level; 95% confidence intervals are reported in brackets.

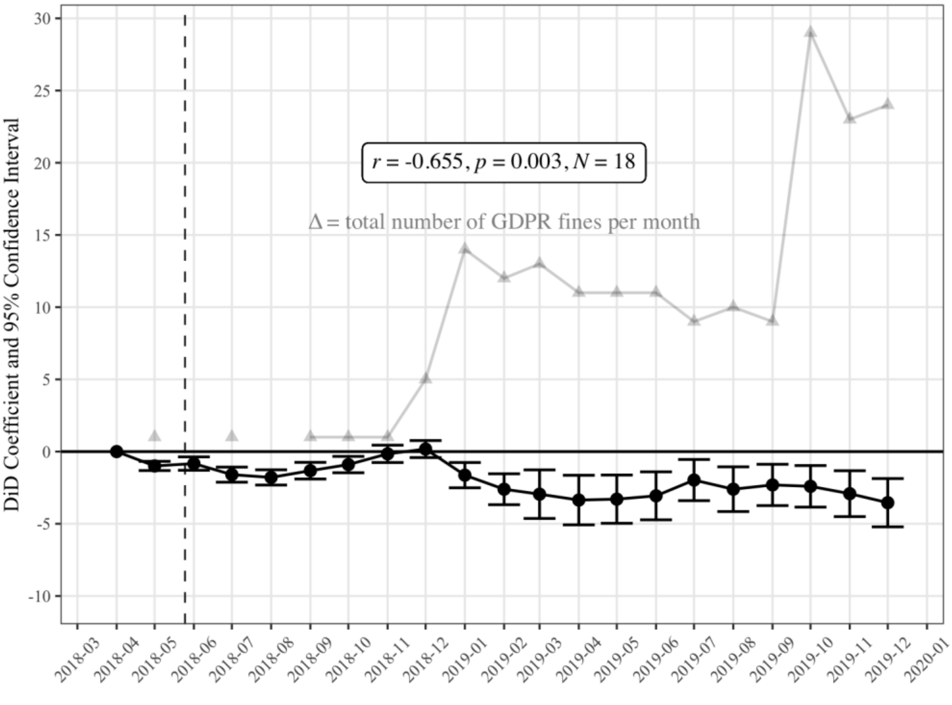
Notes: The Post and the Treatment coefficients have been removed from model (1) due to collinearity. The total number of observations (N = 11,488) is the product of the number of sites (N = 359) and periods (T = 32).

8.6. Correlation between the monthly difference-in-differences coefficients and GDPR fines

We combine data from privacyaffairs.com, which tracks the number (and amount) of GDPR fines over time, with the WhoTracks.me data. We do so to find the correlation between the GDPR's effect on the amount of online tracking, measured by the number of online trackers, and the number of GDPR fines. As Privacy Affairs does not report dates for some GDPR penalties – as they are generally unknown – we removed such observations. €50,000,000 penalty enforced on Google in January 2019 was the highest GDPR fine in our observation period.

We present the calculation of the Pearson's product-movement correlation coefficient between the monthly GDPR difference-in-difference coefficients and the total number of GDPR fines in Figure 10. As seen from Figure 10, there exists a strong negative correlation between the number of GDPR fines and monthly difference-in-differences coefficients, r = -0.655, p = 0.003, N = 18.

Figure 10: Correlation between monthly difference-in-difference coefficients and the total number of GDPR fines per month



Notes: The model includes period and site instance fixed effects. One-way standard errors are clustered at the site instance level. The first period (April 2018) before GDPR enforcement was excluded from the estimation as a reference period.

References

Aggarwal, A., Viswanath, B., Zhang, L., Kumar, S., Shah, A., & Kumaraguru, P. (2018). I Spy with My Little Eye: Analysis and Detection of Spying Browser Extensions. *2018 IEEE European Symposium on Security and Privacy (EuroS&p)*, 47–61. https://doi.org/10.1109/EuroSP.2018.00012

Andrew, J., & Baker, M. (2021). The General Data Protection Regulation in the Age of Surveillance Capitalism. *Journal of Business Ethics*, *168*(3), 565–578. https://doi.org/10.1007/s10551-019-04239-z

Bergemann, D., & Bonatti, A. (2015). Selling Cookies. *American Economic Journal: Microeconomics*, *7*(3), 259–294. https://doi.org/10.1257/mic.20140155

Binns, R., Lyngs, U., Van Kleek, M., Zhao, J., Libert, T., & Shadbolt, N. (2018). Third-Party Tracking in the Mobile Ecosystem. *Proceedings of the 10th ACM Conference on Web Science*, 23–31. https://doi.org/10.1145/3201064.3201089

Bleier, A., Goldfarb, A., & Tucker, C. (2020). Consumer Privacy and the Future of Data-Based Innovation and Marketing. *International Journal of Research in Marketing*, *37*(3), 466–480. https://doi.org/10.1016/j.ijresmar.2020.03.006

Bradford, A. (2020). *The Brussels effect: How the European Union Rules the World* (pp. 1–425). Oxford University Press.

CMS International Law Firm. (2020). *Statistics: Fines Imposed Over Time*. CMS Hasche Sigle Partnerschaft von Rechtsanwälten und Steuerberatern mbB; enforcementtracker.com. https://www.enforcementtracker.com/?insights

Data & Marketing Association (DMA). (2019). *GDPR: A Consumer Perspective 2018* (pp. 1–20). https://dma.org.uk/article/gdpr-a-consumer-perspective

Davies, J. (2019). *’2019 Is the Year of Enforcement': GDPR Fines Have Begun*. Digiday. https://digiday.com/media/2019-is-the-year-of-enforcement-gdpr-fines-have-begun/

Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., & Holz, T. (2019). We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. *Proceedings 2019 Network and Distributed System Security Symposium*, 1–20. https://doi.org/10.14722/ndss.2019.23378

Deighton, J., & Kornfeld, L. (2020). *The Socioeconomic Impact of Internet Tracking* (February; pp. 1–40). Interactive Advertising Bureau. https://www.iab.com/insights/the-socioeconomic-impact-of-internet-tracking/

Eijk, R. van, Asghari, H., Winter, P., & Narayanan, A. (2019). The Impact of User Location on Cookie Notices (Inside and Outside of the European Union). *Workshop on Technology and Consumer Protection (ConPro'19)*, 1–6. https://ssrn.com/abstract=3361360

Englehardt, S., & Narayanan, A. (2016). Online Tracking: A 1-million-site Measurement and Analysis. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 1388–1401. https://doi.org/10.1145/2976749.2978313

European Commission. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council*. https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04

European Commission. (2021). *Proposal for an ePrivacy Regulation*. https://digital-strategy.ec.europa.eu/en/policies/eprivacy-regulation

Evidon. (2021). *Company Database*. https://www.evidon.com/resources/company-database/

Exactag. (2021). *This is How Advertisers Bridge the Gap Between Online and Offline Data*. https://blog.exactag.com/en/this-is-how-advertisers-bridge-the-gap-between-online-and-offline-data

Falahrastegar, M., Haddadi, H., Uhlig, S., & Mortier, R. (2014). Anatomy of the Third-Party Web Tracking Ecosystem. *ArXiv*, 1–12. https://arxiv.org/abs/1409.1066

Falahrastegar, M., Haddadi, H., Uhlig, S., & Mortier, R. (2016). Tracking Personal Identifiers Across the Web. In *Lecture Notes in Computer Science* (Vol. 9631, pp. 30–41). https://doi.org/10.1007/978-3-319-30505-9_3

Fruchter, N., Miao, H., Stevenson, S., & Balebako, R. (2015). Variations in Tracking in Relation to Geographic Location. *ArXiv*, 1–8. https://arxiv.org/abs/1506.04103

Goldfarb, A., & Tucker, C. E. (2011). Privacy Regulation and Online Advertising. *Management Science*, *57*(1), 57–71. https://doi.org/10.1287/mnsc.1100.1246

Haddon, H. (2018). *Exec Recruitment Firms Fear Exposure To GDPR and Fines*. https://www.thehrdirector.com/business-news/data-analytics/recruitment-fear-gdpr-fines/

Hanson, M., Lawler, P., & Macbeth, S. (2018). *The Tracker Tax: The impact of Third-Party Trackers on Website Speed in the United States* (pp. 1–11). Cliqz/Ghostery. https://cdn.ghostery.com/website/wp-content/uploads/2020/10/29102940/Ghostery_Study_-_The_Tracker_Tax.pdf

Johnson, G., Shriver, S., & Goldberg, S. (2021). Privacy and Market Concentration: Intended and Unintended Consequences of the GDPR. *SSRN Electronic Journal*. https://ssrn.com/abstract=3477686

Karaj, A., Macbeth, S., Berson, R., & Pujol, J. M. (2018b). WhoTracks.me: Shedding Light on the Opaque World of Online Tracking. *ArXiv*, 1–15. https://arxiv.org/abs/1804.08959v1

Karaj, A., Macbeth, S., Berson, R., & Pujol, J. M. (2018a). WhoTracks.me: Shedding Light on the Opaque World of Online Tracking. *ArXiv*, 1–15. http://arxiv.org/abs/1804.08959v2

Kraft, L., M. Miller, K., & Skiera, B. (2021). Privacy and the Prevalence of Inconsistencies in Third-Party Consumer Profiling on the Internet. *Working Paper*.

Kummer, M., & Schulte, P. (2019). When Private Information Settles the Bill: Money and Privacy in Google's Market for Smartphone Applications. *Management Science*, *65*(8), 3470–3494. https://doi.org/10.1287/mnsc.2018.3132

Lambrecht, A., Goldfarb, A., Bonatti, A., Ghose, A., Goldstein, D. G., Lewis, R., Rao, A., Sahni, N., & Yao, S. (2014). How do Firms Make Money Selling Digital Goods Online? *Marketing Letters*, *25*(3), 331–341. https://doi.org/10.1007/s11002-014-9310-5

Lambrecht, A., & Tucker, C. (2013). When Does Retargeting Work? Information Specificity in Online Advertising. *Journal of Marketing Research*, *50*(5), 561–576. https://doi.org/10.1509/jmr.11.0503

Laub, R., Miller, K. M., & Skiera, B. (2021). The Economic Value of User Tracking and Behavioral Targeting for Publishers. *Working Paper*.

Lerner, A. (University. of W., Simpson, A. K. (University. of W., Kohno, T. (University. of W., & Roesner, F. (University. of W. (2016). Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. *Proceedings of the 25th USENIX Security Symposium*, 997–1013.

Libert, T. (2015). Exposing the Hidden Web: An Analysis of Third-Party HTTP Requests on 1 Million Websites. *International Journal of Communication*, *9*(1), 3544–3561. https://ssrn.com/abstract=2685330

Libert, T., Graves, L., & Nielsen, R. K. (2018). *Changes in Third-Party Content on European News Websites after GDPR* (Reuters Institute for the Study of Journalism Reports: Factsheet, pp. 1–7). Reuters Institute for the Study of Journalism. https://ora.ox.ac.uk/objects/uuid:5a5d4eea-6e74-49b4-8c77-71ec6760f127

Libert, T., & Nielsen, R. K. (2018). *Third-Party Web Content on EU News Sites: Potential Challenges and Paths to Privacy Improvement* (Reuters Institute for the Study of Journalism Factsheets, pp. 1–11). Reuters Institute for the Study of Journalism. https://ora.ox.ac.uk/objects/uuid:c57241a8-f520-46e5-9e1a-8e4a7f66c23b

Lubowicka, K. (2019). *6 New Privacy Laws Around The Globe You Should Pay Attention To*. Piwik PRO blog. https://piwik.pro/blog/privacy-laws-around-globe/

Macbeth, S. (2017). *Tracking the Trackers: Analysing the Global Tracking Landscape with GhostRank* (pp. 1–13). Technical report, Ghostery. https://medienkraft-1067b.kxcdn.com/cms/wp-content/uploads/2018/10/user-tracking-studie-ghostery.pdf

Macbeth, S. (2018). *Government Websites Leak Data to Google & Co.* https://cliqz.com/en/magazine/government-websites-leak-data-to-google-co

Makeuseof.com. (2021). *Advertise With Us*. https://www.makeuseof.com/advertise/

Mayer, J. R., & Mitchell, J. C. (2012). Third-Party Web Tracking: Policy and Technology. *2012 IEEE Symposium on Security and Privacy*, 413–427. https://doi.org/10.1109/SP.2012.47

McCoy, M. S., Libert, T., Buckler, D., Grande, D. T., & Friedman, A. B. (2020). Prevalence of Third-Party Tracking on COVID-19–Related Web Pages. *JAMA*, *324*(14), 1462–1464. https://doi.org/10.1001/jama.2020.16178

Neumann, N., Tucker, C. E., & Whitfield, T. (2019). Frontiers: How effective is Third-Party Consumer Profiling? Evidence from Field Studies. *Marketing Science*, *38*(6), 918–926. https://doi.org/10.1287/mksc.2019.1188

Peukert, C., Bechtold, S., Batikas, M., & Kretschmer, T. (2021). Regulatory spillovers and data governance: Evidence from the GDPR. *Forthcoming in Marketing Science, SSRN Electronic Journal*. https://ssrn.com/abstract=3560392

Privacy International. (2019). *Your Mental Health for Sale* (pp. 1–37). Privacy International. https://www.privacyinternational.org/campaigns/your-mental-health-sale

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Ryan, J. (2020). *New Data on GDPR Enforcement Agencies Reveal Why the GDPR is Failing*. Brave. https://brave.com/dpa-report-2020/

Sakamoto, T., & Matsunaga, M. (2019). After GDPR, Still Tracking or Not? Understanding Opt-Out States for Online Behavioral Advertising. *2019 IEEE Security and Privacy Workshops (SPW)*, 92–99. https://doi.org/10.1109/SPW.2019.00027

Samarasinghe, N., & Mannan, M. (2019). Towards a Global Perspective on Web Tracking. *Computers & Security*, *87*, 1–19. https://doi.org/10.1016/j.cose.2019.101569

Sanchez-Rola, I., Dell'Amico, M., Kotzias, P., Balzarotti, D., Bilge, L., Vervier, P.-A., & Santos, I. (2019). Can I Opt Out Yet? *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*, 340–351. https://doi.org/10.1145/3321705.3329806

Soltani, A. (2011). *Identifiers and Online Tracking* (pp. 1–5). https://www.w3.org/2011/track-privacy/papers/soltani.pdf

Sørrensen, J., & Kosta, S. (2019). Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites. *The World Wide Web Conference*, 1590–1600. https://doi.org/10.1145/3308558.3313524

Tucker, C. E. (2012). The Economics of Advertising and Privacy. *International Journal of Industrial Organization*, *30*(3), 326–329. https://doi.org/10.1016/j.ijindorg.2011.11.004

Urban, T., Tatang, D., Degeling, M., Holz, T., & 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, ASIA CCS, 20*. https://doi.org/10.1145/3320269.3372194

Utz, C., Degeling, M., Fahl, S., Schaub, F., & Holz, T. (2019). (Un)informed Consent. *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 973–990. https://doi.org/10.1145/3319535.3354212

Viscomi, R. (2020). *Web Almanac* (R. Viscomi, Ed.; 2020th ed., pp. 1–580). HTTP Archive. https://almanac.httparchive.org/en/2020/table-of-contents

WhoTracks.me Privacy Team. (2017a). *What is a Tracker?* WhoTracks.me Blog. https://whotracks.me/blog/what_is_a_tracker.html

WhoTracks.me Privacy Team. (2017b). *Where Does the Data Come From?* WhoTracks.me Blog. https://whotracks.me/blog/where_is_the_data_from.html

WhoTracks.me Privacy Team. (2018). *April Update - Preparing for Internationalisation*. WhoTracks.me Blog. https://whotracks.me/blog/update_apr_2018.html

Zeber, D., Bird, S., Oliveira, C., Rudametkin, W., Segall, I., Wollsén, F., & Lopatka, M. (2020). The Representativeness of Automated Web Crawls as a Surrogate for Human Browsing. *Proceedings of the Web Conference 2020*, 167–178. https://doi.org/10.1145/3366423.3380104