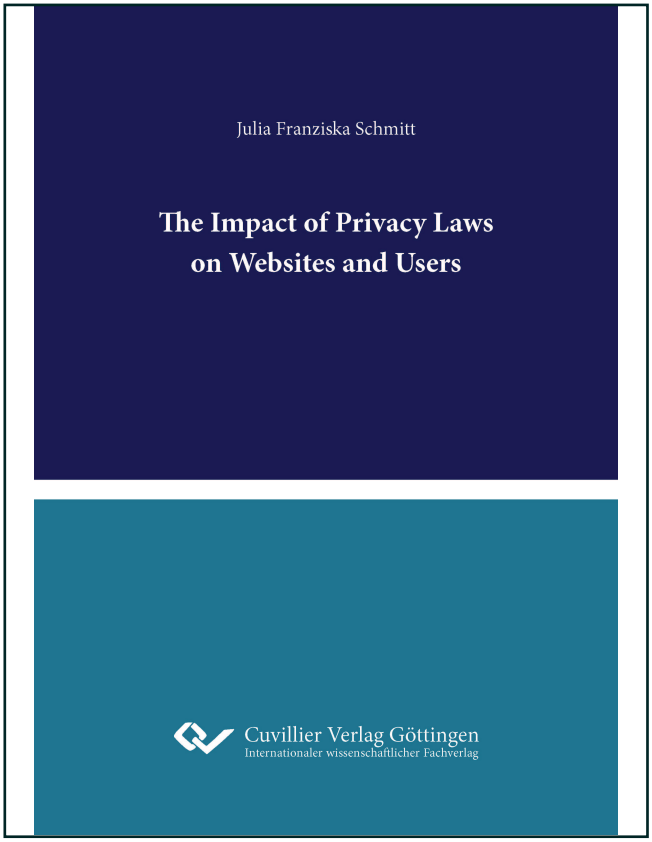
Julia Schmitt (Autor)

The Impact of Privacy Laws on Websites and Users



https://cuvillier.de/de/shop/publications/8623

1.1 Introduction

Internet users generally perceive their privacy as a cause for concern. For example, a survey in 2019 by Pew Research Center showed that 79% of American users are concerned about how companies use their data, partly because they do not know which data companies collect. In recent years, policymakers worldwide have drafted and enforced privacy laws to mitigate these types of concerns.

One of the highest-profile and most expansive laws is the European Union's (EU) General Data Protection Regulation (GDPR), which became enforceable on May 25th, 2018. Similarly, other countries such as Chile, Serbia, Brazil, India, and Thailand have also recently enforced or approved privacy laws. While the specific details of the various privacy laws differ, their basic idea is to increase the individuals' privacy, commonly defined as the individuals' control over their personal data (Holvast 1993).

In practical terms, privacy laws such as the GDPR seek to enhance data privacy by targeting the operations of companies that handle user data through two main avenues: 1) limiting companies' capacity to collect and use user data, and 2) requiring that companies be transparent about their data collection practices.

On the one hand, these requirements resulted in websites reducing the number of third-party cookies (e.g., Libert et al. 2018) and updating and providing more information in their privacy policies, likely increasing the transparency (Degeling et al. 2019; Linden et al. 2020). These findings suggest that GDPR likely increased user privacy in terms of tracker intrusiveness and transparency. On the other hand, as we will elaborate in what follows, these requirements affect companies' operations, which may lead to economic loss. Moreover, companies' attempts to recoup these losses may have negative societal effects. For example, a company might scale back its services or charge for services once provided for free, resulting in a less-informed citizenry. Moreover, some companies might cut jobs, causing financial distress to their employees; if such layoffs take place on a large scale, the societal harm could be profound.

Thus, in establishing privacy regulations, policymakers must carefully balance between ensuring citizens' right to privacy and avoiding excessive damage to the performance of companies that use user data, given the potential societal effects of such damage. Yet, it is challenging to predict how implementing data privacy laws will affect companies' performance and revenue. Part of the challenge stems from users responding in unexpected ways to efforts

to protect their privacy. Indeed, though users claim to value their privacy, it is well established that their actual behavior online does not necessarily align with these stated preferences (known as the privacy paradox; e.g., Acquisti 2004).

Accordingly, we present an empirical study to examine how the GDPR coming into effect, which we refer to as "enforcement of GDPR," affected user behavior on thousands of websites. We focus on two classes of user behavior metrics: user quantity (e.g., numbers of total visits) and usage intensity (e.g., page impressions per visit). These metrics are of interest as indicators of company performance. They often link with companies' revenues (e.g., e-commerce sites or sites with ad-based revenue; see the concluding sections of this article).

Our analysis builds upon the premise that enforcing a privacy law can positively and negatively affect user quantity and usage intensity. Regarding user quantity, limitations on data collection and usage restrict companies' marketing activities, such as targeting new customers through personalized ads. As a result, users might be less aware of certain companies than they would have been otherwise and face increased search costs to find them. Consequently, traffic to those companies' websites might decrease. At the same time, traffic to certain websites might increase among users who find themselves with fewer alternatives – indeed, shortly after the enforcement date of GDPR, some websites operating outside the EU blocked access to EU-users to avoid having to comply with the law (Lecher 2018).

Regarding usage intensity, the requirements for transparency and consent to collect data may require websites to adjust their appearances – thereby affecting the user experience. For example, users might face a pop-up with information regarding the website's cookie usage or other data collection activities and then have to click to accept or decline cookies and the respective data collection. This interaction might increase users' awareness of their data disclosure and influence their usage intensity (Dinev and Hart 2006). In particular, they might spend less time on the website to reduce the amount of data it can collect, or they might abandon the website to avoid having to authorize it to collect data. Alternatively, once users have consented to have their data collected, they might use the website more than they would otherwise – to avoid having to visit other websites and authorize them to collect data. Lastly, there might be users who do not change their behavior at all.

These arguments suggest that, overall, the enforcement of a privacy law such as the GDPR, i.e., the GDPR becoming effective, may have positive or negative effects, or no effect at all, on

the number of users who visit a particular website and on their usage intensity. Moreover, different websites might be affected differently, as users' expectations regarding their privacy and their consequent responses to privacy-driven changes in website operations may vary across regions (cultures) or websites in different industries (e.g., Dinev et al. 2006). It is also essential to understand how these effects develop over time, as it might take users several months to adjust their usage habits.

Thus, our study aims to achieve the following specific objectives:

1) Quantifying the effects of the enforcement of the GDPR on five metrics of user quantity and four metrics of usage intensity on websites over time (from 3 months up to 18 months after the enforcement of the GDPR);
2) Identifying how these effects vary as a function of website characteristics (i.e., website industry and popularity) and user characteristics (i.e., a user's country of origin).

Our analysis relies on a dataset capturing user behavior on 6,286 unique websites spanning 24 industries; these websites represent the most popular websites in 13 countries (11 EU countries, Switzerland, and the United States). The data cover the period from July 2017 to December 2019 – i.e., 10 months before and 18 months after the enforcement of the GDPR (hereafter referred to as "GDPR") on May 25th, 2018 – enabling us to construct a before-and-after analysis.

Within our dataset, some website-user interactions are subject to the GDPR (i.e., interactions involving EU-websites or EU-users). In contrast, others are not (i.e., interactions involving Non-EU-websites and Non-EU-users), effectively creating a "control group." Thus, we can use a panel differences estimator similar in spirit to a difference-in-differences (DiD) estimation (e.g., Janakiraman et al. 2018, Kumar et al. 2016, Goldstein et al. 2014). We combine the panel differences estimator with a synthetic control group (SCG) approach (Abadie et al. 2015) to isolate the effect of the GDPR on our metrics of interest.

We obtain the following results:

1) Among websites to which the GDPR is applicable, the average number of visits per website decreases by almost 5% in the short-term and about 10% in the long-term; about two-thirds of websites continue to be negatively affected by the GDPR in the long-term. We similarly observe short-term decreases of 0.8%-3% in the average number of unique visitors, page impressions, and amount of time on the website, and long-term decreases of 6.6%-9.7%.

2) Among websites that suffer from a reduction in user quantity, the remaining users exhibit an increase in usage intensity – for example, the number of visits per user increases, on average, by about 4.8% at 18 months post-GDPR. Conversely, among websites that gain users after the GDPR, usage intensity decreases; e.g., the number of visits per user decreases, on average, by about 9.1% at 18 months post-GDPR.
3) The effects of the GDPR vary across websites; for example, less-popular websites lose more total visits (10%-21% drop) than more-popular websites (2%-9% drop), suggesting that the GDPR increases market concentration. The effects also vary across industries, with Entertainment and Leisure websites being most negatively affected (-12.5 to -13.8% after 18 months). In contrast, Business and Consumer Service websites even experience a positive effect (+4.7% after 18 months).
4) User characteristics (i.e., a user's country of origin) have only a small effect on how the GDPR affects user behavior.

1.2 Knowledge on Effects of Privacy Changes on Online User Behavior

We draw from and contribute to two main streams of literature. Through surveys and lab experiments, the first stream attempts to illuminate users' attitudes towards data privacy and their responses to different levels of privacy or control over their data. The second stream uses field studies to examine the effects of privacy laws on various outcomes of interest.

1.2.1 User Attitudes and Behavior with Regard to Privacy

Lab experiments and survey-based studies have examined how users' attitudes and website usage behavior are affected by websites' handling of user privacy. The results of these studies point to a nuanced relationship between privacy and user behavior. For example, several studies based on consumer surveys suggest that when users perceive themselves as having more control over their privacy – specifically, more options to regulate their privacy – they experience lower privacy concerns (Martin 2015), a higher level of trust in a website, an increase in purchase intentions (Martin et al. 2017) and a higher willingness to disclose data to websites (Brandimarte et al. 2013; Acquisti et al. 2013; Malhotra et al. 2004; Culnan and Armstrong 1999). They can even react more positively to personalized ads (Tucker 2013).

Other studies, in contrast, find that different privacy levels do not affect user behavior: For example, Belanger and Crossler (2011) show that users share data with companies despite privacy concerns. This result may have been induced by users' feelings of powerlessness regarding their privacy (Few 2018). Acquisti et al. (2012) further show that users' privacy concerns and preferences for the same level of privacy are not stable. The willingness to disclose data can depend on other factors like the amount and order of such data requests. These findings align with the privacy paradox, indicating that users' stated privacy preferences often differ from their actual behavior (e.g., Acquisti 2004).

Still, other studies suggest that including more privacy control options for users might negatively affect website usage. In particular, privacy features, such as requesting users' explicit consent for data collection and more transparency (as required by GDPR), can make users aware of data disclosure that they were not previously aware of (Dinev and Hart 2006), increase privacy concerns and thus reduce ad effectiveness (Kim et al. 2018). This awareness may lead users to feel warier about using the site and thus diminish their usage.

Dinev and Hart (2006) proposed the privacy calculus theory, which provides a framework encompassing all these different responses to privacy controls. Specifically, the theory suggests that the extent to which a user values privacy on a particular website depends on the user's privacy concerns, the user's trust in the website, and the value that the user derives from the website's offerings. Users with higher privacy concerns or lower trust towards a website may be more likely than others to respond favorably to more stringent privacy measures. In turn, when users attribute a high value to the website's offerings, they may be willing to sacrifice privacy in exchange for convenient access to those offerings and thus may be indifferent to privacy levels – or even respond unfavorably if privacy hurts the website's accessibility.

This theory suggests that users' responses to changes in a website's handling of privacy may vary across users and websites. Indeed, several studies show that differences in privacy perceptions and expectations depend on a user's country and cultural background (e.g., Dinev et al. 2006; Steenkamp and Geyskens 2006; Miltgen and Peyrat-Guillard 2014) and on the device used by a user to access a website (Melumad and Meyer 2020). The current study extends these findings by comparing how users in different countries vary in their responses to privacy laws and by considering variations across websites with different characteristics.

1.2.2 Field Studies: Effects of Privacy Laws on Various Outcomes

The findings outlined above suggest that it is likely to be challenging to predict how large populations of users will respond to the enforcement of new privacy laws. Accordingly, several studies use field data to construct event studies of users' revealed behavior after enforcing such laws. Examples that predate the GDPR are the work of Goldfarb and Tucker (2011a), who show that implementing the EU Privacy and Electronic Communications Directive reduces ad effectiveness on websites, making it more challenging for ad-financed websites to generate revenues, and of Campbell et al. (2015) who show that privacy laws especially hurt smaller online companies. At the same time, Goldfarb and Tucker (2011b) further show that irrespective of privacy laws, ad effectiveness can diminish for strongly obtrusive and targeted ads, suggesting a positive effect of privacy laws on user welfare.

Several recent studies have specifically sought to characterize various effects of the GDPR. Some of these works focus on websites' actions in response to the law, showing that many update their privacy policies (Degeling et al. 2019) and increase their privacy policy length (Linden et al. 2020). Furthermore, an apparent reduction in third-party cookies occurs (Libert et al. 2018; Hu and Sastry 2019). Partly due to the anticipated reduction in third-party cookies, Mirreh (2018) predicts that websites could lose almost half of their traffic because of an inevitable shift of retargeting strategies, making it more challenging for companies to get users to their websites.

A study that is particularly relevant to our research is that of Goldberg et al. (2021), who measure how the GDPR affected recorded web traffic and e-commerce sales four months after the enforcement of the regulation. The authors show an average 11.70% drop in recorded page views from EU-users (Goldberg et al. 2021). Our empirical study delivers insights that greatly extend Goldberg et al.'s research. Primarily, our study adopts a long-term orientation for a substantially larger website sample, providing a more comprehensive analysis of GDPR. Given that the GDPR was the first major new privacy law in the EU since the e-Privacy Directive in 2002, users may have needed some time to adjust their behavior to the GDPR. Therefore, the full effect of the privacy law might only become observable after some time. Furthermore, our study examines differences in the effects across websites and users. Finally, the data sample of our study enables an empirical estimation of metrics covering actual traffic, whereas Goldberg et al.'s available data only allow an examination of recorded traffic. As the authors mention in their study, a change in recorded traffic after GDPR is, in fact, a combination of two changes:

A change in the number of consenting users and a change in the actual traffic that these consenting users generate.

1.3 Description of Empirical Study

Our empirical study aims to analyze the effects of the enforcement of the GDPR on online user behavior, as reflected in measures of user quantity and usage intensity; to understand how these effects evolve over time (distinguishing between short-term effects – 3 months after enforcement –, up to long-term effects – 18 months after enforcement); and to reveal how these effects vary as a function of website and user characteristics.

1.3.1 Background on the GDPR

The GDPR, which came into effect on May 25th, 2018, is the first major privacy law in Europe since the e-Privacy Directive in 2002. The GDPR regulates any activity performed on personal data from users located in the EU. As a regulation, the law is further binding for all websites based in EU countries; according to Article 3 of the GDPR, a website's "base" (and thus the applicability of the GDPR) is determined according to the geographical location where the website's data processing takes place. Websites within the scope of GDPR that do not comply with the privacy law face significant fines of up to 4% of the website's global annual turnover or €20 million, depending on the severity of the infringement.

The GDPR handles various privacy aspects that can affect how a user engages with a website. Similar to other approved or enforced privacy laws such as Brazil's Lei Geral de Protecao de Dados Pessoais (LGPD), India's Personal Data Protection Bill (PDPB), or Thailand's Personal Data Protection Act (PDPA), the GDPR has stringent privacy protection requirements (Lucente and Clark 2020). For example, the mentioned privacy laws all require websites to obtain a user's explicit consent for data processing like the GDPR, i.e., they all follow an opt-in approach for consent. Given the similar nature of GDPR compared with other privacy laws, the findings of this study likely mirror the effects of other privacy laws on user quantity and usage intensity on websites. At the same time, for privacy laws that are less strict than GDPR, such as the California Consumer Privacy Act (Lucente and Clark 2020), the findings of this study might serve as an upper bound of the effects.

1.3.2 Description of Set-Up of Empirical Study

Before the GDPR becoming effective, users could not anticipate how websites would react to the diverse set of requirements imposed by GDPR. In our empirical study, we examine the effect of the enforcement of GDPR, i.e., the GDPR coming into effect, on the user behavior on websites. Most likely, websites complied with GDPR to different degrees. So, we do not measure the effect of all websites behaving entirely according to GDPR. Instead, we observe the effect of the websites' interpretation of the privacy law. Thus, we measure what happened after GDPR came into effect – the intention-to-treat effect of GDPR. Therefore, our treatment "enforcement of GDPR" refers to "GDPR coming into effect" (on May 25, 2018) and not to a situation in which GDPR was enforced such that all websites behaved entirely with GDPR.

The GDPR provides a useful setting for quantifying the effect that the enforcement of privacy laws has on user behavior because it implicitly divides website-user interactions (here referred to as "website-instances") into a treatment group (i.e., GDPR is applicable) and a control group (i.e., GDPR does not apply), as depicted in Figure 1.1.

Figure 1.1: Scope of GDPR and Resulting Assignment to Treatment and Control Group

		Base of User	
		EU	Non-EU
Base of Website	EU	GDPR applies = Treatment Group	GDPR applies = Treatment Group
	Non-EU	GDPR applies = Treatment Group	GDPR does not apply = Control Group

□ Treatment Group □ Control Group

As noted above, the GDPR's scope includes all websites based in the EU and further encompasses the processing of personal data from all users located in the EU. Thus, the treatment group comprises website-instances corresponding to EU-users visiting any website or to Non-EU-users visiting EU-websites. The control group consists of website-instances corresponding to Non-EU-users visiting Non-EU-websites. In line with Article 3 of the GDPR, we use the website's server location (retrieved from https://check-host.net) to determine the respective website's data processing location and the GDPR's applicability. We use the enforcement date of GDPR (May 25th, 2018) to construct a before-and-after analysis, comparing the treatment

group to the control group to quantify the intention-to-treat effect of GDPR. This approach allows us to construct a panel differences estimator that is similar in spirit to a DiD estimator and rests upon two critical assumptions: the stable unit treatment value assumption (SUTVA) and the parallel pre-treatment trends of the control and the treatment group.

Several factors might bias the treatment effect that we observe using the described methodology. For example, there might be concerns regarding the possible late or early compliance of websites with GDPR or the potential existence of confounding factors. The major concern, however, might be regarding the validity of our control group. This concern stems from the possible situation that websites in our control group might voluntarily comply with GDPR. Furthermore, the mere knowledge of Non-EU-users about the GDPR already represents a "treatment" that affects our control group as well. Both situations would represent a violation of the SUTVA that is integral to our analysis.

We thoroughly examine the robustness of our results to all of those factors, i.e., late or early compliance, confounding factors, and the possibility that GDPR also treats our control group. All robustness checks indicate that the mentioned factors do not bias our results (see Sections 1.9.4, 1.9.5, and 1.9.7 in the Appendix). Thus, even if a potential bias existed within our results, its impact is likely relatively small. Furthermore, such an effect only yields to underestimating GDPR's actual effect because the treatment might also impact the control group.

1.3.3 Overview of Data

1.3.3.1 Description of Data Sample.

This study utilizes data from SimilarWeb for the Top 1,000 websites – as listed in Alexa Top Sites in April 2018 – of two Non-EU countries (Switzerland and USA) and 11 EU countries (Austria, Denmark, France, Germany, Hungary, Italy, Netherlands, Poland, Spain, Sweden, and the UK[3]). The authors choose the USA and Switzerland as Non-EU-countries as both countries are culturally similar to the EU. SimilarWeb draws on a diversified and rich global user panel to measure online user behavior. Companies (e.g., Google, Alibaba, eBay, P&G) primarily use data from SimilarWeb, but also researchers in top-tier academic journals (e.g., Calzada and Gill 2020, Lu et al. 2020). The websites in our sample span diverse industries (see

[3] During the time of our study, the United Kingdom (UK) was still a member of the EU. Its membership ended on January 31th, 2020.

Figure A1 - 1 in Section 1.9.1 in the Appendix), audiences, and popularity levels (here measured by SimilarWeb ranks). For each website in our sample, the dataset includes information about the website industry as well as the global, country, and industry rank, based on the website's popularity worldwide, in the analyzed country, and within the website's industry.

For each website in the sample, the dataset further includes information on the user quantity metrics of users accessing that website from one EU country: the one in which the website is most popular. Additionally, for each website, user quantity data are available for users accessing the website from the US. Thus, if a website does not appear in the Top 1,000 of any EU country, data are available only for US users. These data span the period between July 1st, 2017 and December 31st, 2019 – i.e., almost a year before GDPR's enforcement (May 25th, 2018) and 1.5 years after the enforcement – and can therefore be used for a before-and-after analysis as outlined above.

We start with 13 countries with 1,000 websites each. Our initial sample includes 7,332 unique websites after we removed duplicate websites. For example, "google.com" is a duplicate website as it is among the Top 1,000 websites in all 13 countries. Instead of occurring 13 times, google.com just occurs once in our sample. For each of these 7,332 websites, we have user behavior data corresponding to Non-EU-users. For 6,460 websites of those 7,332 websites, the dataset additionally includes user behavior data of EU-users.

Thus, for 6,460 websites, we have two sets of observations, corresponding, respectively, to the Non-EU-user base and to the EU-user base of that website. For the remaining 872 websites, we only observe the Non-EU-user base. In what follows, we consider each website's Non-EU and EU-user bases separately and refer to each combination of a website with one of the two user bases, for convenience, as a "website-instance." For example, for a website such as "zeit.de" that is based in an EU country (here, Germany), we observe two website-instances: One website-instance corresponds to the set of observations for the EU-user base of "zeit.de." The second website-instance corresponds to the set of observations for the Non-EU-user base of "zeit.de." As "zeit.de" is EU-based, GDPR applies to both its website-instances, and both website-instances belong to the treatment group (Figure 1.1).

Accordingly, for a website such as "nzz.ch" that is based in a Non-EU country (here: Switzerland), we observe two website-instances: one website-instance corresponding to the set of observations for the Non-EU-user base of "nzz.ch" and the second website-instance corresponding to the set of observations for the EU-user base. As the website of this second example