



Information Systems Research

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Law, Economics, and Privacy: Implications of Government Policies on Website and Third-Party Information Sharing

Ram D. Gopal, Hooman Hidaji, Sule Nur Kutlu, Raymond A. Patterson, Niam Yaraghi

To cite this article:

Ram D. Gopal, Hooman Hidaji, Sule Nur Kutlu, Raymond A. Patterson, Niam Yaraghi (2023) Law, Economics, and Privacy: Implications of Government Policies on Website and Third-Party Information Sharing. Information Systems Research

Published online in Articles in Advance 16 Feb 2023

<https://doi.org/10.1287/isre.2022.1178>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2023 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Law, Economics, and Privacy: Implications of Government Policies on Website and Third-Party Information Sharing

Ram D. Gopal,^{a,*} Hooman Hidaji,^b Sule Nur Kutlu,^b Raymond A. Patterson,^b Niam Yaraghi^c

^aInformation Systems and Management, Warwick Business School, the University of Warwick, Coventry CV4 7AL, United Kingdom;

^bHaskayne School of Business, University of Calgary, Calgary, Alberta T2N 1N4, Canada; ^cBusiness Technology, Miami Herbert Business School, University of Miami, Coral Gables, Florida 33146

*Corresponding author

Contact: ram.gopal@wbs.ac.uk,  <https://orcid.org/0000-0003-4241-9355> (RDG); hooman.hidaji@haskayne.ucalgary.ca,  <https://orcid.org/0000-0002-5443-7297> (HH); sulenur.kutlu@ucalgary.ca,  <https://orcid.org/0000-0002-1193-1842> (SNK); raymond.patterson@ucalgary.ca,  <https://orcid.org/0000-0001-5408-6485> (RAP); niamyaraghi@miami.edu,  <https://orcid.org/0000-0003-3497-0251> (NY)

Received: April 4, 2021

Revised: December 1, 2021; June 24, 2022; August 10, 2022

Accepted: August 25, 2022

Published Online in *Articles in Advance*: February 16, 2023

<https://doi.org/10.1287/isre.2022.1178>

Copyright: © 2023 The Author(s)

Abstract. Widespread abuse of internet users’ privacy online has prompted user advocacy groups to implore governments to intervene and protect consumer rights. To study such interventions’ effects, we examine data-protection policies that policy makers and governments can enforce on websites, including consent-based user information sharing and subsidizing competing websites. We use a stylized analytical model to examine such policies’ impact on the decisions and outcomes of websites, users, and third-parties. Interestingly, we find that even though a consent-based policy may improve user surplus, in the absence of market entry and exit (static market), it has the unintended consequence of increasing the number of third-parties and, thus, sharing of user information. We also determine that both consent-based and website subsidization policies may reduce competition by driving websites out of the market—to the detriment of user surplus and social welfare. To validate our analytical model’s findings, we empirically investigated the impact of a consent-based policy on third-parties in a natural experiment of the California Consumer Privacy Act. These findings raise significant implications for policy making surrounding online privacy.

History: Giri Kumar Tayi, Senior Editor; Nachiketa Sahoo, Associate Editor.

 **Open Access Statement:** This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. You are free to download this work and share with others, but cannot change in any way or use commercially without permission, and you must attribute this work as “*Information Systems Research*. Copyright © 2023 The Author(s). <https://doi.org/10.1287/isre.2022.1178>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.”

Funding: Financial support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged. R. D. Gopal gratefully acknowledges support by the Gillmore Centre for Financial Technology at the Warwick Business School.

Supplemental Material: The online appendix is available at <https://doi.org/10.1287/isre.2022.1178>.

Keywords: data protection regulation • government policy • website and third-party information sharing • online privacy

1. Introduction

In today’s world, the internet is intertwined with almost every aspect of our lives. We rely on websites for everything, from entertainment and communication to banking and education. This, along with effective data-collection and -analysis technologies, has created a market in which websites sell their user data to “third” parties. As a report from the U.S. Senate (2014, p. 124) indicates, “If a user visits most ordinary websites (e.g., a newspaper website or a blog), some third-parties will likely place a cookie on that user’s computer. Almost every website examined by the sub-committee called some third-party or parties who operated cookies on that website.” Such data monetization raises serious privacy concerns and potentially harms

users through financial discrimination—including price, insurance, and mortgage discrimination (Newman 2014). Illustrating what a contentious issue this is, Facebook has recently lodged a public protest of a proposed Apple policy to require explicit user consent for tracking through apps downloaded via Apple’s App Store (Nicas and Isaac 2020, *Economist* 2021).

Yet, in the absence of any substantial regulation, websites freely collect and share users’ data with third-parties for the purposes of targeted advertising and improved functionality. Growing concerns about harmful third-party data sharing has led consumer-rights advocacy groups to call for policy interventions that protect user data and privacy from third-parties. Should

the government decide to regulate this market, it could implement one of two regulatory mechanisms. The first option is to provide users with more choices by requiring websites to disclose their third-parties and obtain users' consent before sharing their data with any particular third-party. This consent-based policy is adopted by the European Union (EU) as the General Data Protection Regulation (GDPR) and, more recently, by the state of California as the California Consumer Privacy Act (CCPA) by enforcing punitive penalties for noncompliance.¹

The second alternative policy approach for the government is to subsidize the "good actors"—those websites that follow stricter privacy guidelines and refrain from excessively sharing user data. This is a more precise non-legal intervention that can target specific websites or industries. This website subsidization policy is similar to governmental policy in many other sectors of the economy, such as post and rail (Head 1994, Geddes 2003), where the government funds public goods to compete with the private sector to enhance consumers' surplus by providing alternatives. In the context of websites, similar efforts have been implemented in select industries in the United States, such as news (NPR.org) and healthcare (www.nih.gov/health-information), and we are likely to see further efforts in this area as policy makers reinforce their involvement. Thus, here, we analyze these two alternative policies and compare them with the status quo of no regulation as benchmark. Currently, all of Canada and every state in the United States except for California have a noninterventionist policy, as does much of the world.

The complexity and convoluted relationship between the various entities involved warrants careful examination of the data-regulation policies. Although these regulatory interventions intend to protect users' data and privacy, their unintended effects on users, websites, and third-parties are subject to much debate. For example, according to Ovum (2014), strict data-protection regulations in the GDPR create financial burdens and competitive disadvantages for foreign companies, especially for those in the United States. Similarly, Johnson et al. (2019) show that the GDPR has the unintended effect of increasing the market concentration of web technology vendors, especially for ones that use data-collection techniques such as cookies. Another concern with such legislative interventions is that they behave as blunt instruments, so that once the law is implemented, it applies to all websites, regardless of market-specific characteristics such as the level of user privacy concern or competition.

With that in mind, in this paper, we focus on the implications of regulating website-oriented businesses—that is, businesses where the website content is the main source of revenue, from either user subscriptions, online advertising through third-parties, or both. These websites can exercise the option of charging users directly through subscriptions. Moreover, they can monetize their website through third-party targeting and advertising. These

third-parties gain access to user information through the website, which allows the third-parties to track users across the internet and present users with relevant advertising. Websites then receive compensation based on the accuracy of targeting and effectiveness of advertising through per-impression or per-click contracts. This kind of advertising is common across informational, news, and social media websites, such as *The New York Times* and Twitter. Such businesses are substantially impacted by third-party-limiting regulation. We exclude from our scope websites that are used to monetize something other than their website's content, such as Amazon and Target (because they are selling physical goods) and Bank of America (because they are selling financial products). In such instances, the websites' primary business model is not monetizing the website content itself.

The primary research methodology used in this paper is closed-form analytic modeling. We examine how the consent-based (such as the ones implemented by the GDPR and CCPA) and website-subsidization policies impact users, websites, and third-parties; and how these market participants alter their behavior in reaction to the policies. We show that in the absence of market entry and exit (static market), consent-based and website subsidization policies may benefit user surplus and social welfare. Interestingly, as long as there are at least some users with relatively low privacy concerns, even though the consent-based policy improves user surplus, it has the unintended consequence of increasing the number of third-parties and data sharing. This effect stems from users' willingness to accept higher data sharing in return for reduced prices. In fact, if privacy concerns are sufficiently low, then websites set prices to zero and provide their services for free. Additionally, considering market entry and exit (a dynamic market), our analysis shows that such policies may have an unintended secondary effect on players via competition (market size). Specifically, both consent-based and website subsidization policies may cause websites to exit the market, thereby reducing competition and suppressing user surplus and social welfare. To demonstrate our analytical models' validity, we empirically examine the third-party data sharing of the 100,000 most popular websites in response to implementing consent-based data-protection policies in the state of California. Our empirical investigation validates important findings from our analytic modeling with this natural experiment for the rollout of a consent-based policy.

These findings have important implications for policy makers in the domain of data-protection and privacy regulation. Most importantly, we demonstrate that a blunt legislative policy, such as the consent-based policies proposed by the GDPR and CCPA, is not always effective in addressing third-party data-sharing concerns. Such

policies may, in fact, harm user privacy by incentivizing more third-party usage. Moreover, such a policy, in some cases, suppresses competition, which has a significant indirect impact on the outcomes for user and social welfare. On the other hand, more precisely targeted mechanisms—such as subsidizing websites in particular sectors or industries—may force competing websites to improve their third-party data sharing. Thus, rather than globally implementing legislation, we advocate for a combination of policies and local subsidies that are better suited to an industry’s specific needs.

2. Literature Review

Two primary streams of research are relevant to our work. The first stream involves studies of online information leakage and user privacy concern on the internet. McDonald and Cranor (2010) find that only 20% of users prefer targeted online advertising to random advertising. Similarly, Turow et al. (2009) show that 83% of people surveyed do not want to get a personalized web page based on their browsing history and expect policy makers to protect their rights. Despite these sentiments, regulations to protect users’ privacy are lacking, and this deficiency results in greater abuse of users’ personal data. Krishnamurthy et al. (2011) find a growing disconnect between privacy leakage and protection measures. Mayer and Mitchell (2012) explain the possible harm of web tracking and the status quo reaction of authorities and policy makers to privacy issues of web users. Similarly, Kostkova et al. (2016) point out possible obstacles in the use of big data in healthcare. They highlight the need for a policy and regulatory agenda at an international level that protects user data and limits abuse by businesses.

The second stream examines third-parties employed by websites. As reported by *The Wall Street Journal* (Angwin 2010), on average, the nation’s top 50 websites place 64 pieces of tracking technology on users’ computers without their permission, track their browsing history, and collect various information continuously. Krishnamurthy and Wills (2006) show that third-parties generate negative utility for users by tracking users’ browser activities and collecting their data. They also find that third-parties’ activities are remarkably important for users. Kumar and Sethi (2009) show that hybrid revenue models, consisting of both subscription fees and advertisements, allow websites to have higher revenue compared with pure revenue models. However, they do not consider the privacy concern of users and their effect on website revenue. Similarly, Gopal et al. (2018) use a two-sided economic model to analyze how websites set user subscription fees and third-party royalties to generate revenue streams from users and third-parties, and they consider the effect of user privacy concerns on market players.

Multiple empirical investigations focus on the effect of legislative government interventions on third-parties and websites. Because the most prominent example of this type of intervention is the EU’s GDPR, many data-driven analyses have been conducted to observe this regulation’s effect. However, no clear consensus exists on whether the GDPR decreases third-party usage. Dabrowski et al. (2019) collected cookies from the Alexa Top 100,000 websites, and they compared snapshots of cookie usage by websites, not specifically third-party cookies, before and after the GDPR. Even though they observed a decrease in the use of cookies with the GDPR for EU and non-EU users, they pointed out that a need exists for further investigation with longitudinal data regarding the third-party information component. Similarly, Sørensen and Kosta (2019) showed a decrease in third-party usage with an analysis of a smaller data set over time for different countries and website categories. However, they observed a slight increase in the third-party usage of some public websites after the GDPR, but did not clearly explain the possible driving factor behind it. Hu and Sastry (2019) tracked information collected using third-party cookies by Alexa Top 500 websites for EU and non-EU users’ browsing activities over a year expanding before and after the GDPR. They found that EU users have more third-party cookies than non-EU users. Additionally, they stated that websites that offer a choice to users to opt out have more cookies than other websites. Thus, more study is needed regarding the effect of legislative government interventions.

We are not aware of any work on the combined analytical and empirical analysis of various data-protection policies’ economic implications and how they relate to third-party data sharing. Our work addresses this lacuna by analytically modeling key entities of interest: websites, users, and third-parties. How websites and third-parties react to data-protection constraints and the consequent impacts on privacy and economic outcomes are modeled, and the key analytical results are validated empirically through a large-scale natural experiment of the CCPA rollout. We examine the exchange of user data between websites and third-parties from both legal and economic perspectives.

3. Model Setup, Notation, and Analysis

To begin, we introduce our model of websites. We focus on website-oriented businesses—that is, businesses where the website is the main source of revenue, which can be from user subscriptions or from third-parties. For context, consider informational websites, news websites, and social media, each of which have the option to charge users and/or third-parties, and the website forms the bulk of the business. To analyze the implications of government policy, first,

we analyze a benchmark case that does not employ any policy and examine its economic outcomes. Motivated by the regulatory mechanisms used in industry, in addition to a benchmark case, we analyze the effect of two data-protection policies: first, a consent-based policy, where users decide which third-parties are allowed on a website; and second, a subsidization policy, in which the government subsidizes a website to compete in the market. We use a set of common assumptions for all models to ensure comparability. For each policy, we obtain closed-form equilibria, the details of which are provided in Online Appendix A.

The Salop circular model (Salop 1979) is used to incorporate users' heterogeneous preferences.² Users are uniformly distributed over a circle with a total circumference of one, denoting the total number of users in a market. We consider single-homing users who only visit one website with the highest utility.³ Then, according to the Salop model, we consider t to represent a misfit cost for users to visit a website. Considering symmetric and equidistant websites, without loss of generality, we assume the first website to locate at zero ($Q_1 = 0$), at the top of the circle, and the second website to locate at $1/2$ ($Q_2 = 1/2$). The distance between a website and a user represents that user's preference for the website. For example, some users may prefer CNN, whereas others prefer Fox News, and some lie in between. Figure 1 depicts the websites on the Salop circular city model. Although websites may be asymmetric in reality, we treat them as symmetric to isolate the impact from data-protection policies.⁴

Websites share their user data with third-parties. In the case of online advertising, the websites share their users' browsing and individual data with advertising third-parties by allowing use of third-party cookies and passing data through *https* requests (Englehardt et al. 2015, Libert 2015). In the case of functionality third-parties such as Google Analytics, the service is

offered in exchange for user data. Even though third-parties provide value to the websites through monetization, this negatively impacts user utility because of privacy concerns.⁵ As the number of third-parties that a website employs increases, user utility from that website decreases due to the additional privacy costs. Previous studies validate this assumption (Krishnamurthy and Wills 2006, Turow et al. 2009, Gopal et al. 2018). We take privacy concern to be the individuals' personal assessment of fairness within the scope of data privacy (Malhotra et al. 2004). In this context, if users have high privacy concerns, they are less likely to visit websites with a high number of third-parties and, instead, would be more willing to pay higher prices to avoid websites that share their data with third-parties (Gopal et al. 2018). We assume heterogeneous user privacy concerns, in that two groups of users exist: one with high privacy concern and another with low privacy concern. We assume that websites can set only one price for all of their users.

3.1. Benchmark (No Policy)

We first analyze the absence of data-protection policies in our *Benchmark (BeK)* setup, providing a baseline model to which we can compare the different policies. We use a two-sided model to study the websites, users, and third-parties. Websites act as platforms that provide services and information to their users and generate revenue from both users (through prices) and third-parties (through third-party access fees⁶). Websites almost always use at least one of these two revenue streams for monetization (Kumar et al. 2018). We see a similar setup in Gopal et al. (2018). Figure 2 shows our setup. Third-parties obtain users' personal data from the websites in return for the third-party access fee paid to the websites, usually in the form of targeted advertising.⁷ We refer to this royalty as a *third-party access fee*. Table 1 provides the notation.

A website $i \in \{1, 2\}$ decides the price, P_i , and the number of third-parties,⁸ N_{T_i} . All website costs are normalized to zero. The website revenue under Benchmark is given as:

$$\pi_{W_i}^{\text{BeK}} = N_{U_i}P_i + N_{T_i}N_{U_i}R_i \quad \forall i \in \{1, 2\}, \quad (1)$$

where N_{U_i} is the number of users of website i , and R_i is the third-party access fee of website i .

As discussed, users single-home, and the market is covered by websites. A user gains positive utility when she visits a website i that depends on the price of the website (P_i), the number of third-parties of the website (N_{T_i}), and user privacy concern. Here, a website's price can be considered as either a monetary subscription or service price, or any other cost burdened by users to consume the website's content completely, such as users' time and attention, which can be monetized by the website. We assume heterogeneous privacy concern

Figure 1. Salop Circular City Model of Two Websites

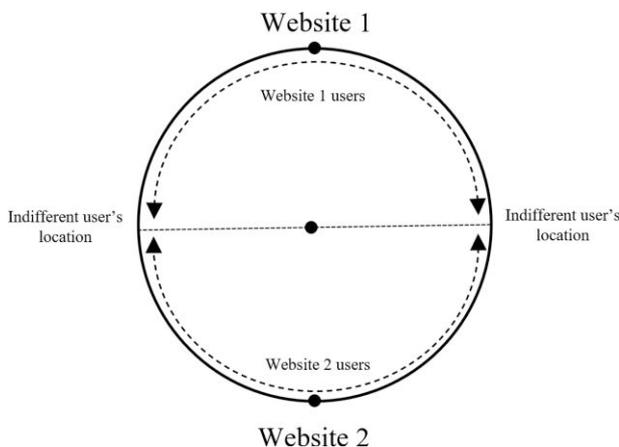
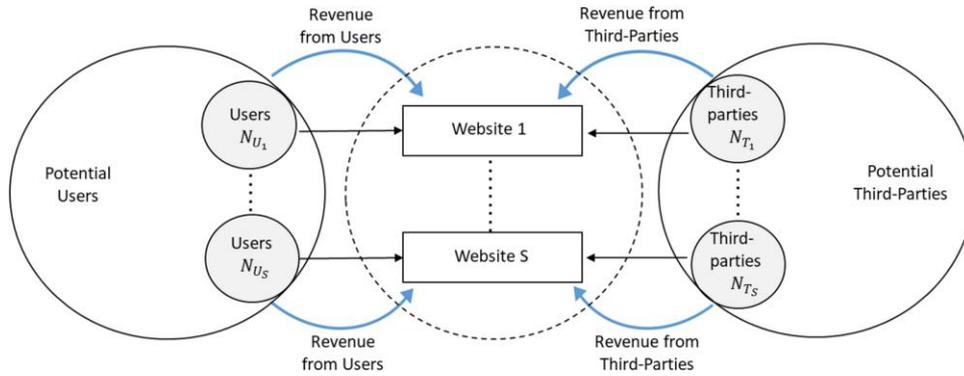


Figure 2. (Color online) Websites, Users, and Third-Parties in the General Case with S Websites



among users. Specifically, the population of users is composed of two types with respect to their privacy concern: Ratio ρ_H of the population is high-type (type H) with relatively high privacy concern ($v_H \in R^+$), and ratio ρ_L of the population is low-type (type L) with relatively low privacy concern ($v_L \in R^+$, $v_L < v_H$). The total population is normalized to one unit, where $\rho_H + \rho_L = 1$.

Based on the user's and websites' location, a user subscribes to the website that provides the highest utility. Accordingly, the utility that a user located at $0 \leq y \leq 1$ who is from type $k \in \{H, L\}$ receives from visiting website i under Benchmark is given as:

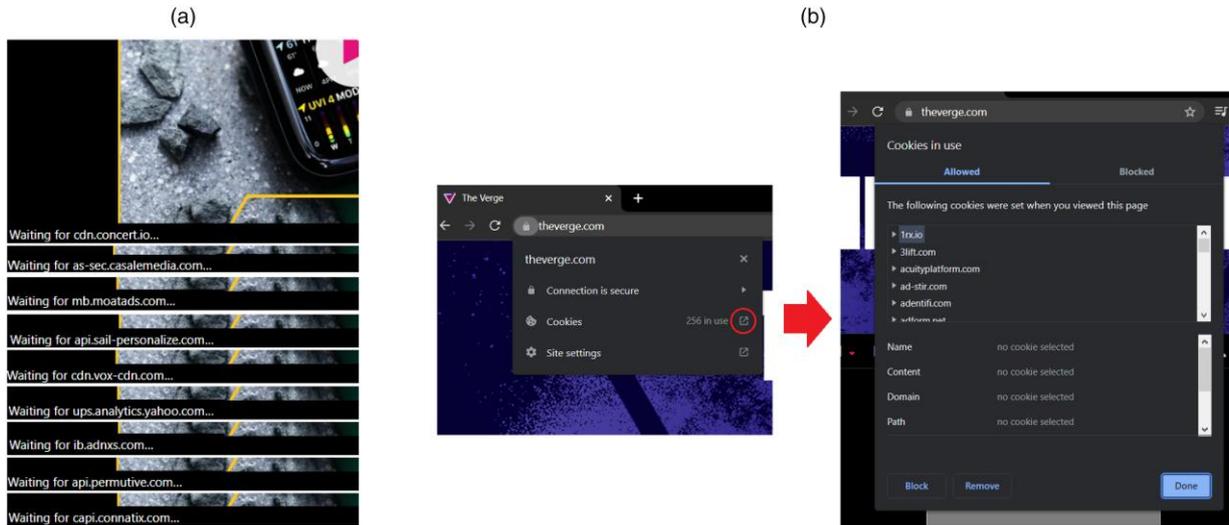
$$U_{i,k}^{\text{BeK}}(y) = X - v_k N_{T_i} - P_i - t |Q_i - y| \quad \forall i \in \{1, 2\}, k \in \{H, L\}, \quad (2)$$

where t is the misfit cost and X is the intrinsic utility gained from a website. Note that $|Q_i - y|$ gives the (circular) distance of the user to website i .⁹ We assume that all websites provide equivalent service and, therefore,

identical intrinsic utility to their users.¹⁰ However, they vary based on the number of third-parties and their price. In this specification, $v_k N_{T_i}$ captures the total third-parties' cost of website i for users of type k . Although a website may not fully reveal all the user data that are being shared with third-parties (even though there exist many policies and guidelines requiring publisher sites to reveal such information to consumers, particularly under the consent-based policy), there exist tools to do so, and privacy watchdogs regularly publish reports and audits about the third-party sharing behavior of websites that users can use to obtain such information.¹¹ Moreover, the amount of advertising that users see on a website along with the page-loading time and notifications often seen on the bottom right or bottom left of a browser window provide an indication of the number of third-parties used (refer to Figure 3(a) for an example). Users can also easily inspect the third-party cookies used on a website in modern browsers (refer to Figure 3(b) for an example). Our results continue to hold if the user information

Table 1. Notation

Notation	Definition
$U_{i,k}^j(y)$	The utility that a user located at y who is from type k receives from website i in policy case j
X	The intrinsic value provided by a website to users, $X \geq 0$
t	Misfit cost, $t \geq 0$
S	Market size or the number of websites operating in the market
Q_i	Location of website i , derived as $Q_i = (i - 1)/S$, $0 \leq Q_i \leq 1$
R_i	Third-party per-user access fee of website i , $R_i \geq 0$
N_{U_i}	Number of users for website i , $N_{U_i} \geq 0$
v_k	Privacy concern of type $k \in \{H, L\}$ user, $0 \leq v_L \leq v_H \leq 1$
B	A third-party's revenue from each user's data, $B \geq 0$
c	The fixed cost of a third-party, $c \geq 0$
C	The maximum value that the fixed cost of a third-party could take or third-parties' cost, $C > 0$
z	Nonprofit website's comparative utility provision, $\underline{z} \leq z \leq \bar{z}$
m	Nonprofit website's number of third-parties
$\pi_{T_i}^j$	Profit of a third-party from the website i in policy case j , $\pi_{T_i}^j \geq 0$
$\pi_{W_i}^j$	Revenue of the website i in policy case j , $\pi_{W_i}^j \geq 0$
Decision variables	
N_{T_i}	Number of third-parties of the website i , $N_{T_i} \geq 0$
P_i	Price of website i , $P_i \geq 0$

Figure 3. (Color online) Finding Third-Parties in a Browser for an Example Website (Chrome 95.0 for theverge.com)

Notes. (a) Loading notifications. (b) Third-party cookies used.

about third-party sharing is not perfect, but their perceived privacy concern correlates with the websites' third-party sharing behaviors.

Third-parties participate in the websites based on the profit they can make on each website. We assume third-parties make revenue from each user's shared data and denote this per-user revenue as B . Third-parties pay a per-user access fee to the website, R_i . Third-parties are assumed to be heterogeneous in their fixed cost, which we denote as c . This is because of the differences in third-party capabilities, where some third-parties are more cost-efficient than others in delivering services. The third-party capability determines their participation: Third-parties with revenues higher than the fixed cost participate in the website.¹² We assume the fixed cost of all third-parties in the market to be uniformly distributed over the range $[0, C]$. The upper bound of the uniform distribution, C , generally represents the cost of doing business for third-parties, which we simply refer to as third-parties' cost. Based on these assumptions, profit of a third-party with cost c from the website i is given as:

$$\pi_{T_i}^{\text{BeK}} = N_{U_i}(B - R_i) - c \quad \forall i \in \{1, 2\}. \quad (3)$$

Given the aforementioned setup, our first lemma characterizes the equilibrium under Benchmark. All proofs are relegated to Online Appendix A.

Lemma 1 (Benchmark). *In equilibrium under Benchmark:*

a. *The websites set a symmetric price that increases in user privacy concern (v_H and v_L), but decreases in third-party revenue from user data (B). Particularly, where user privacy concern is low and/or third-party revenue from user data is high, the website sets a price of zero.*

b. *The websites utilize a symmetric number of third-parties that decreases in user privacy concern (v_H and v_L), but increases in third-party revenue from user data (B).*

Under benchmark, where user privacy concern of the two user types is not too low or third-party revenue from user data is not too high, the websites' optimal strategy is to monetize both sides of the market—that is, to charge users for subscription to its services, as well as selling their data to third-parties. This finding is contrary to other forms of two-sided markets with positive cross-sided externality, where the platform subsidizes one side and monetizes the other. This is, nevertheless, consistent with the literature on websites' business models, where one side (third-parties) imposes negative externality on the other side (users)—for example, in Gopal et al. (2018). Intuitively, as user privacy concern increases, third-parties become less desirable, and the website's cost of using them increases. Therefore, the website shifts its monetization from third-parties to subscription price. The extent of this shift depends on the extent of change in each user type's privacy concern (v_H and v_L for types H and L , respectively) and the population ratio that these two types encompass (ρ_H and ρ_L , respectively). The increase in third-parties' cost has a similar effect to that of user privacy concern. Moreover, as the third-party benefit from user data (B) increases, websites can charge a higher access fee from third-parties, which enables them to reduce the user price.

Interestingly, if privacy concerns are sufficiently low and/or third-party revenue from user data is high, websites may not directly charge users; that is, they set a corner solution of $P_i = 0$ and monetize users only indirectly through third-parties. This is the case

for many websites, particularly informational websites such as news websites (e.g., CNN or Fox News). However, the free-to-use model is not the only business model used by such websites, and other websites charge subscription fees. For example, *The New York Times*, *The Washington Post*, *The Wall Street Journal*, *The Financial Times*, and *The Globe and Mail* have a subscription-based model. It has been shown that publisher websites that employ a subscription model generally use fewer third-parties than those focusing on monetization through third-parties (Gopal et al. 2018). Even though this decision may be based partly on a website’s content and its value to users, it also is impacted by the website’s strategy of monetizing users via third-parties.

3.2. Consent-Based Policy

Inspired by data-protection regulations such as the GDPR and CCPA, this policy requires user consent for sharing data with third-parties. Under this policy, users’ explicit consent is taken before their data are shared with any third-party, and users have the option to deny access to any particular third-party. Such policies enforce compliance by administering punitive penalties on noncompliant websites. Given the substantial penalties set by the GDPR and CCPA, we assume the policy to be implemented perfectly; that is, once enforced, all websites comply. We use the general term *Consent-Based (CoB)* to refer to such a policy. The difference from Benchmark is that, in this case, users can decide which third-parties to allow and which ones to reject, resulting in a different decision process. Whereas under Benchmark, websites decide both their prices and the number of third-parties, under CoB policy, users decide the number of third-parties from a list provided by the website, which, in turn, determines the price that websites charge. We formalize this as a two-stage game. In the first stage, the website provides users

with a menu of third-parties. At one extreme, this menu allows for a price of zero and the number of third-parties that enables the free service, and at the other extreme, it allows for a high price that enables the website to use no third-parties at all. This offered menu also includes the corresponding price for each chosen number of third-parties. Then, in the second stage, given this menu, users decide the number of third-parties that they allow at a website with its associated website price. This timeline lets us consider the equilibrium results at the website, insofar as the website sets prices to account for the users’ choice of the number of third-parties. Note that although users may decide the number of third-parties that they allow, unlike Benchmark, under CoB, different types of users may choose a different number of third-parties. Websites within jurisdictions that enforce CoB policies (such as the EU or California) are forced to ask users for consent for each third-party with which they share data. In most websites, this is operationalized as a list of third-parties that users can enable or disable by selecting or deselecting as they desire, as Figure 4 shows. The website, however, can only set one price for all user types.

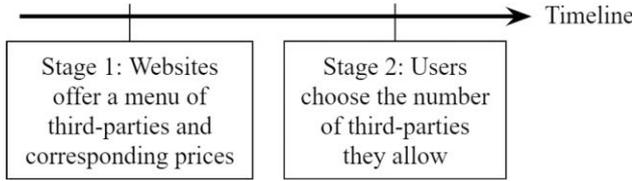
Figure 5 provides the timeline of decisions for the Consent-Based setting. We use backward induction to solve this two-stage game. Note that even though we model heterogeneity of privacy concern among users through high- and low-type users, the privacy concern of users in either or both types can still be relatively low or high. In other words, the high- and low-type users are comparative to each other, and it is possible that even the high-type users have relatively low privacy concern or low-type users have relatively high privacy concern.

Lemma 2 (Consent-Based). *Under Consent-Based policy, dependent on users’ privacy preferences, four different equilibria are possible:*

a. *If the privacy concern for both user types is relatively low ($v_H < \tau_H$ and $v_L < \tau_L$), then both types of users allow all*

Figure 4. (Color online) Selecting Third-Parties on a Website Under Consent-Based Policy (<https://cookie-script.com/knowledge-base/consent-for-cookies>)



Figure 5. The Decision Timeline Under Consent-Based Policy

third-parties from the menu. The websites set the price to zero.

b. If the privacy concern for type H is relatively high ($v_H > \tau_H$), but for type L is relatively low ($v_L < \tau_L$), then type H users allow no third-parties and type L users allow all third-parties from the menu. The websites set the price to zero.

c. If the privacy concern for type H is relatively low ($v_H < \tau_H$), but for type L is relatively high ($v_L > \tau_L$), which may occur only where type H is dominant ($\rho_H > \rho_L$), then type H users allow all third-parties and type L users allow no third-parties from the menu. The websites set the price to zero.

d. If the privacy concern for both types is relatively high ($v_H > \tau_H$ and $v_L > \tau_L$), then neither user types allow any third-parties from the menu. The websites set a nonzero price.

Under the Consent-Based policy, unless both types of users have high privacy concern, websites set their price to zero and monetize users only indirectly through third-parties. This is because of the preference of one (type L in scenario (b) and type H in scenario (c)) or both types of users (scenario (a)) to not pay for the service, as they are relatively relaxed about their privacy. As the websites set the price of zero, in scenarios (b) and (c), one type subsidizes free use of the website for the other type. Interestingly, where type H dominates the population of users, it is possible for type L users to allow fewer third-parties than type H . This is because the dominant group of users are more important to the websites, and websites decide their pricing strategy based on the dominant user type. Therefore, the nondominant type, type L in this case, can allow no third-parties without causing the website to increase its price.

On the other hand, if both types of users have relatively high privacy concern, then neither type allows any third-parties, as the privacy cost of using third-parties is prohibitively high. In this scenario, the website sets the price above zero and monetizes users only directly, forgoing the revenue from the third-party side. In the current market conditions, ample evidence exists that, if not most, then at least some users do not care about their online privacy enough to be willing to pay for the websites' services. For

example, a study found that 87% of users either ignored cookie banners or accepted all cookies when prompted after the GDPR went into effect (Mueller 2018). Moreover, third-parties are present on all websites (U.S. Senate 2014, Gopal et al. 2018), implying that there are at least some users who prefer the use of third-parties over subscription fees. Therefore, we believe that the scenario where all users have high privacy concern (scenario (d)) does not reflect the conditions under which most website-oriented businesses operate. Moreover, we note that the default number of third-parties set by the website is the number of third-parties that allow the website to set a price of zero for users with relatively low privacy concern (this can be either the high- or low-type users).

3.3. Website Subsidization Policy

Under the Website Subsidization policy, a government or nonprofit entity subsidizes a website that competes with commercial websites. This subsidized website is provided for free to users and uses a minimal number of third-parties (m) to cover its expenses and operate. Examples of such subsidized websites exist in select industries such as news (e.g., www.NPR.org, www.digital.gov, and www.canada.ca/en/news.html) and healthcare (e.g., www.nih.gov/health-information or www.canada.ca/en/health-canada). The subsidized website is not necessarily a government website, but a nonprofit website that can be supported by the policy maker. We refer to this policy as *Website Subsidization (WeS)*. Without loss of generality, we assume the nonprofit website to locate at zero, replacing the commercial website at this location. The government or nonprofit policy maker decides the *comparative utility provision* (z) of its website, which determines the subsidized website's intrinsic utility relative to commercial websites. The nonprofit website can arrange its comparative utility provision based on its objective, which may be to maximize user surplus or social welfare. We assume that if the market is not covered by commercial websites, then the government sets the nonprofit website's comparative utility provision ($z \in [\underline{z}, \bar{z}]$) high enough at least to cover the market. The minimum threshold for the nonprofit website's comparative utility provision, \underline{z} , assures that the market is covered. The maximum threshold, \bar{z} , is there to assure that the commercial website is viable.

The utility that a user located at y receives from the nonprofit website and commercial websites under the Website Subsidization policy are given as:

$$U_1^{\text{WeS}}(y) = Xz - v_k m - t|Q_1 - y| \quad \forall k \in \{H, L\}, \quad (4)$$

$$U_{2,k}^{\text{WeS}}(y) = X - v_k N_{T_2} - P_2 - t|Q_2 - y| \quad \forall k \in \{H, L\}. \quad (5)$$

The nonprofit website utilizes some monetization third-parties so that it can recoup its operational costs.

In reality, nonprofit and government websites are limited in the types of third-parties they can use and profit that they can make,¹³ so this assumption is realistic. Users and third-parties contribute to the revenue of the commercial website:

$$\pi_{W_2}^{WeS} = N_{U_2}P_2 + N_{T_2}N_{U_2}R_2. \quad (6)$$

Given this setup, our next lemma characterizes the equilibrium under Website Subsidization.

Lemma 3 (Website Subsidization). *In equilibrium under Website Subsidization:*

a. *The commercial website sets a price that increases in user privacy concern (v_H and v_L), but decreases in third-party benefit from user data (B) and comparative utility provision (z).*

b. *The commercial website utilizes a positive number of third-parties that decreases in user privacy concern (v_H and v_L) and comparative utility provision (z), but increases in third-party revenue from user data (B).*

The subsidized website does not utilize any third-parties and does not charge a price. For commercial websites, the impact of privacy concern and third-party market characteristics on prices and the number of third-parties under Website Subsidization are similar to those of Benchmark. Under this policy, an additional parameter impacts the price and number of third-parties for the commercial website, and that is the subsidized website’s comparative utility provision. Comparative utility provision intensifies the commercial website’s competition for users, and, therefore, as it improves, the commercial website is forced to reduce both its price and number of third-parties to retain appeal to users.

4. Data-Protection Policies’ Implications in Static Markets

We use our results from the previous section and compare the Consent-Based and Website Subsidization policies to Benchmark to study the impact of these policies on users, websites, and third-parties. In this section, we restrict our analysis to the case where the market size (number of websites in the market, S) is fixed to two websites—that is, where policy does not impact the number of players in the market. Section 5 extends our analysis to the dynamic market case, where websites can enter and exit the market.

Proposition 1 (Implications of Consent-Based Policy).

a. *If there are at least some users with relatively low privacy concern, then the Consent-Based policy increases the default number of third-parties set by the website; decreases website prices and website revenue; and increases user surplus and third-party surplus. The Consent-Based policy decreases social welfare if the third-parties’ cost and misfit cost are both moderate ($\underline{C} < C < \hat{C}$ and $\underline{t} < t < \hat{t}$).*

b. *If there are no users with relatively low privacy concern, then the Consent-Based policy decreases the default number of third-parties set by the website; increases website prices and website revenue; and decreases user surplus and third-party surplus. The Consent-Based policy decreases social welfare.*

In the scenario where both types of users have relatively low privacy concern, they allow all third-parties, and this drives the website price down to zero. In this case, the websites monetize users only indirectly through third-parties. This is indeed what we observe in many industries, where websites do not charge a price. This results in lower revenue for the websites and shifts the surplus to users and third-parties. As for the overall social welfare, this impact depends on third-parties’ cost and misfit cost. If both costs are moderate, then Consent-Based policy decreases social welfare; otherwise, it improves social welfare. Therefore, the policy maker’s decision on whether the Consent-Based policy is effective depends on the characteristics of the market (the intensity of competition among websites) and the third-parties (their efficiency in benefiting from user data). These results continue to hold when only one type of user has relatively low privacy concern, with the caveat that in these scenarios, some users allow no third-parties, whereas others allow all third-parties. However, if there are no users with relatively low privacy concern, then the website cannot monetize users through third-parties, as users would not allow this. Instead, the website sets a higher subscription fee and only monetizes users directly. This removes third-parties from the market, causing social welfare to decrease. We note that there is evidence that the current conditions for website-oriented firms include at least some users who are not concerned about their privacy, and part (b) of Proposition 1 does not currently occur in practice.

Focusing on the case where there are at least some users not concerned about privacy, consistent with the original intent of legislation such as the GDPR and CCPA, our results reveal beneficial aspects of the Consent-Based policy for consumers in certain scenarios—lower prices and higher user surplus. Interestingly, however, this policy also benefits third-parties. As users are in the driver’s seat in terms of third-party usage, websites entice them to consent to third-party usage through lower prices. For websites, this policy shifts the balance of revenues away from users and toward third-parties. This has the counterintuitive effect of increasing the default number of third-parties set by the website and data sharing relative to Benchmark. In Section 7, we discuss empirical evidence for this important insight.

Proposition 2 (Implications of Website Subsidization Policy). *The Website Subsidization policy decreases (increases) the commercial website price, number of third-parties, and website revenue, as well as third-party surplus if comparative utility provision is high (low).*

The Website Subsidization policy has an additional lever that can be controlled—a comparative utility provision (z), which impacts the implications of this policy. As the comparative utility provision increases, a commercial website faces stiffer competition from the subsidized website. This drives the commercial website to reduce both its prices as well as the number of third-parties it utilizes. This also results in lower revenue for the commercial website. Even though it is not possible to analytically study the implications for user surplus and social welfare, later in the analysis, we numerically demonstrate that it is possible to achieve higher user surplus than the Benchmark by appropriately determining the comparative utility provision.

Proposition 3 (Consent-Based versus Website Subsidization).

a. *If there are at least some users with relatively low privacy concern, then the Consent-Based policy relative to the Website Subsidization policy decreases the commercial websites' price. Moreover, the Consent-Based policy decreases (increases) the default number of third-parties set by the commercial website and commercial website's revenue if the comparative utility provision is low (high).*

b. *If there are no users with relatively low privacy concern, then the Consent-Based policy relative to the Website Subsidization policy increases the commercial websites' price, decreases the default number of third-parties set by the commercial website, and decreases the third-party surplus. Moreover, the Consent-Based policy decreases (increases) a commercial website's revenue if comparative utility provision is low (high).*

The commercial website's preference between the two policies depends on the comparative utility provision offered by the nonprofit website. If this utility provision is low, then the commercial website prefers the Website Subsidization policy. However, if the competitive pressure increases through a higher-quality subsidized website, then the commercial website's preference for the Consent-Based policy increases. If there are at least some users with relatively low privacy concern, then the third-party surplus may either decrease or increase when the Consent-Based policy is used compared with Website Subsidization. On the other hand, if there are no users with relatively low privacy concern, then no third-party is used under Consent-Based, which eliminates the third-party surplus. As discussed, there is evidence that currently, there are at least some users with relatively low privacy concerns in the market, implying that part (b) of Proposition 3 does not reflect the current reality.

Even though it is not possible to compare the Website Subsidization to Benchmark and Consent-Based policies analytically in terms of user surplus and social welfare, we provide a numerical analysis for this. In particular, we are interested in examining whether either the Benchmark or Website Subsidization policy

can yield superior outcomes to the Consent-Based policy. Essentially, we assess whether nonlegal interventions can yield superior outcomes in terms of user surplus and social welfare over legal interventions. To perform this analysis, we conduct an extensive numerical analysis of the comparison between user surplus and the policies' social welfare. Note that even though we analytically cannot confirm the results for the comparison, our model provides the closed-form equations for each of these constructs, and, therefore, the numerical results we present are of high confidence.

Result 1 (User Surplus).

a. Under the Website Subsidization policy, user surplus increases with the nonprofit website's comparative utility provision.

b. There exists a minimum threshold value of the nonprofit website's comparative utility provision, beyond which the user surplus of the Website Subsidization policy is higher than that of the Benchmark and Consent-Based policy (where there are at least some users with relatively low privacy concern).

c. The minimum threshold value is decreasing in the users' privacy concern.

Our results show that it is always possible to achieve higher user surplus with Website Subsidization compared with the Consent-Based policy. Thus, from the perspective of user surplus, a nonregulatory option is a viable policy instrument. We find that privacy concern is negatively associated with the minimum threshold value. So, if privacy concern is high, even a lower-quality nonprofit website that competes with the commercial website can be beneficial to users.

Although higher user surplus always can be achieved through the Website Subsidization policy, this is not the case for social welfare. In our numerical analysis, we observe that it is not always possible for the Website Subsidization policy to provide higher social welfare than the Consent-Based policy. We explore conditions that make it disadvantageous for the Website Subsidization policy and the Benchmark case to provide higher social welfare than the Consent-Based policy through a regression analysis. Results are provided in Result 2.

Result 2 (Social Welfare).

a. Under the Website Subsidization policy, social welfare first increases, then decreases with the nonprofit website's comparative utility provision.

b. As user privacy concern (v_H and v_L) increases, where there are at least some users with relatively low privacy concern, the Consent-Based policy produces higher social welfare than the Benchmark and the Website Subsidization policy.

Results 1 and 2 shed interesting insights on the comparative utility provision's opposing effects. Lower values of the nonprofit website's utility provision lower user surplus, but increase social welfare. The results

switch at high values of utility provision. So, thwarting competition with a high-quality nonprofit website is harmful for social welfare. In this case, it is better to give control to users who can leverage their control to determine the number of third-parties they allow. Furthermore, when competition is intense, the nonprofit website’s marginal effect diminishes, and the Website Subsidization policy loses its appeal. Section 7 provides empirical evidence for these findings.

5. Data-Protection Policies’ Implications in Dynamic Markets

In our analyses thus far, we assumed that the market size (i.e., the number of websites in the market) is fixed under different policies. However, given that these policies have revenue implications, policy enforcement may alter the incentive of websites to enter or exit the market. In the following analysis, we consider the effects of a fixed cost of entry. For parsimony, we do not consider sequential entry, but model firms as entering simultaneously and examine equilibrium outcomes.

Considering the increased competition (increased market size of up to $S = 5$), intuitively, as competition increases, user surplus increases, but prices and the number of third-parties used by websites decrease. Competition also lowers each website’s revenue and third-party surplus. Therefore, if a policy impacts the number of websites operating in an industry, then that policy will have an indirect effect on websites and users. To analyze this indirect effect, we consider a fixed entry cost, F , which, along with website revenue, determines the number of firms in the market, as websites need to make positive payoffs. In other words, the market size is decided such that each website earns revenues higher than the fixed entry cost F , but adding one more website makes the revenues lower than F . This setup follows a two-stage Cournot game, where websites make entry decisions in the first stage and then decide their prices and third-parties in the second stage.

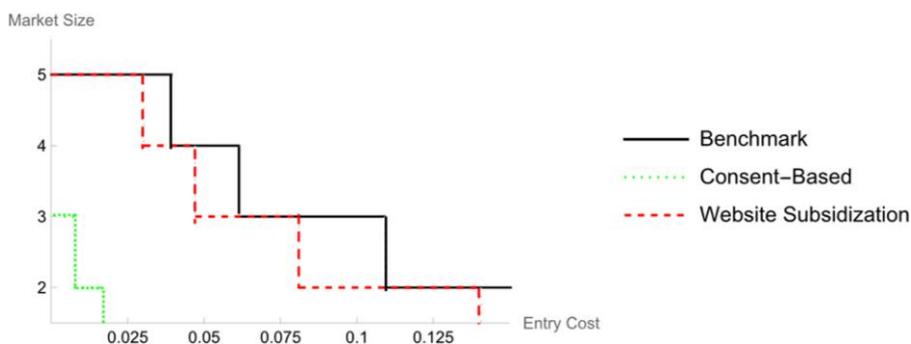
We know from Proposition 1 that the Consent-Based policy yields lower website revenue than the Benchmark. Therefore, the market size under the Consent-Based policy is lower than or equal to the Benchmark ($S^{CoB} \leq S^{BeK}$). Comparing website revenue and market size for the Website Subsidization policy to other policies, however, depends on the nonprofit website’s comparative utility provision (z). If the comparative utility provision is high, then the commercial website revenue is lower, which reduces the market size in presence of barriers to entry. However, if the comparative utility provision is low, then the commercial websites’ revenue is high, implying that there can be more commercial websites in the market. Thus, endogenizing the market size can alter the policy implications significantly for the website and users. Figure 6 depicts how market size under different policies changes with entry cost in an example scenario.

Proposition 4 (Dynamic Markets). *Considering barriers to entry and competition (market size ranging from two to five websites):*

- a. *If the entry cost is small, then the number of websites does not change with policies, and the findings are identical to that of static markets.*
- b. *If there are at least some users with relatively low privacy concern, then as the entry cost increases, the Consent-Based policy may cause some websites to exit, thereby hindering competition and diminishing user surplus and social welfare.*
- c. *If there are no users with relatively low privacy concern, then as the entry cost increases, the Consent-Based policy allows for additional websites to enter, thereby improving competition, user surplus, and social welfare.*
- d. *If the nonprofit website’s comparative utility provision is high, then the Website Subsidization policy hinders competition and diminishes user surplus and social welfare. If the comparative utility provision is low, however, then the Website Subsidization policy improves competition, which improves user surplus and social welfare.*

When entry cost is small, the results remain unchanged. In this case, Benchmark yields the highest revenue for

Figure 6. (Color online) Entry Cost and Market Size Under Different Policies ($C = 5.8, t = 1.48, B = 3.96, v_H = 1, v_L = 0.41, X = 1, z = 0.99, m = 0$)



Downloaded from informs.org by [137.205.8.152] on 16 February 2023, at 04:53. For personal use only, all rights reserved.

websites, followed by the Consent-Based policy. Depending on the nonprofit website's comparative utility provision, the Website Subsidization policy may be anywhere on this spectrum. User surplus is highest under the Consent-Based policy, followed by Benchmark, with the Website Subsidization policy depending on the nonprofit website's comparative utility provision. The social welfare ordering of policies depends on the parameters and nonprofit website's comparative utility provision.

On the other hand, when barriers to market entry are substantial, if there are at least some users with relatively low privacy concern, then some websites may drop out of the market under both the Consent-Based and Website Subsidization policies compared with Benchmark. The Consent-Based policy dampens market size as the cost of entry increases. Market size for the Website Subsidization policy depends on the nonprofit website's comparative utility provision, where for low comparative utility provision, the market size increases, and for high comparative utility provision, the market size decreases. Therefore, incorporating the market size effect significantly impacts our earlier results. However, if all users have high privacy concern, then the Consent-Based policy allows for more websites and, thereby, improves competition. As we discuss, this does not seem to be the case in the current environment, as evidence exists that at least some users have relatively low privacy concern.

Depending on cost of entry, the Consent-Based policy may hinder competition, which is detrimental for user surplus, and the Consent-Based policy may no longer be the best policy when considering user surplus or social welfare. Depending on the nonprofit website's comparative utility provision, the Website Subsidization policy may improve (for low comparative utility provision), hinder (for high comparative utility provision), or not change (for moderate comparative utility provision) competition. If the nonprofit website's comparative utility provision is high, then the Website Subsidization policy hinders competition, which harms users. In this case, there are two opposing effects from the Website Subsidization policy on user surplus: Adding a nonprofit website improves their surplus, whereas reducing competition lowers their surplus. If the nonprofit website's comparative utility provision is low, however, the Website Subsidization policy improves competition, which benefits users. These findings have important implications for data-protection policy making, in that the secondary effect of policies on changing the competition may alter the outcomes.

6. Extensions

Next, we extend our model and analysis to handle scenarios where websites are asymmetric and where users may multihome.

6.1. Asymmetric Commercial Websites

In our main model, we consider commercial websites to provide symmetric intrinsic utility to users. Here, we extend our model to consider commercial websites that are asymmetric in their intrinsic value provided to users, where website 1 offers X_1 and website 2 offers X_2 . Without loss of generality, we assume website 1 to be superior to website 2—that is, $X_1 > X_2$. With this setup, the utility that a user located at $0 \leq y \leq 1$ who is from type $k \in \{H, L\}$ receives from visiting website i under the Benchmark and Consent-Based policy is given as:

$$\begin{aligned} U_{i,k}^{BeK}(y) &= X_i - v_k N_{T_i} - P_i - t |Q_i - y|, \\ U_{i,k}^{CoB}(y) &= X_i - v_k N_{T_{i,k}} - P_i - t |Q_i - y| \\ &\quad \forall i \in \{1, 2\}, k \in \{H, L\}. \end{aligned}$$

Using this setup and a similar analysis to that of symmetric websites in the previous sections, we study data-protection policies' impact on websites, users, and third-parties where websites are asymmetric. Given that the intrinsic utility from the nonprofit website under Website Subsidization is different from the commercial website, this policy is fundamentally asymmetric. Therefore, here, we only focus on the Benchmark and Consent-Based policy, noting that our prior results on Website Subsidization continue to hold. Our next proposition summarizes our findings for asymmetric websites.

Proposition 5 (Asymmetric Websites). *Where websites are asymmetric:*

- a. *Under the Benchmark, the high-quality website sets a higher price, utilizes more third-parties, and generates higher revenue compared with the low-quality website.*
- b. *Under Consent-Based, where there are at least some users with relatively low privacy concern, a high-quality website utilizes more third-parties and generates higher revenue, but sets the same price of zero compared with a low-quality website.*
- c. *Under Consent-Based, where there are no users with relatively low privacy concern, both websites utilize zero third-parties, but the high-quality website charges a higher price and generates higher revenue compared with the low-quality website.*
- d. *Comparing the websites under Benchmark and Consent-Based policies, where there are at least some users with relatively low privacy concern, under Consent-Based, each website (whether high or low quality) sets a higher number of third-parties and charges a lower price.*
- e. *Comparing the websites under Benchmark and Consent-Based policies, where there are no users with relatively low privacy concern, under Consent-Based, each website (whether high or low quality) sets a lower number of third-parties (zero) and charges a higher price.*

It is evident that under the Consent-Based policy, where at least some users have relatively low privacy

concern, the website with higher intrinsic value sets a higher number of third-parties. Thus, those users who accept all third-parties are exposed to more third-parties. The high-privacy user rejects the use of third-parties, and the price is set to zero for both websites. On the other hand, if there are no users with relatively high privacy concern, then both websites do not use any third-parties and only monetize users directly. In this scenario, the high-quality website sets a higher price. Our findings on the Consent-Based policy's impact on the number of third-parties and prices are consistent with the symmetric model, as provided in Proposition 1. Given the asymmetric model's complexity, we cannot derive all of the results analytically that we had from the symmetric model. This is expected, as others have discussed (Gopal et al. 2018). Nevertheless, we are able to provide numerical results based on a broad numerical analysis of the parameters.

Result 3 (Asymmetric Websites). Comparing websites under Benchmark and Consent-Based policy, we find the following:

- a. Both websites (high and low quality) gain lower revenue under a Consent-Based policy where there are at least some users with relatively low privacy concern, but both websites gain higher revenue when there are no users with relatively low privacy concern.
- b. User surplus is higher under Consent-Based policy where there are at least some users with relatively low privacy concern, but user surplus is lower where there are no users with relatively low privacy concern.
- c. The total third-party surplus is higher under Consent-Based policy where there are at least some users with relatively low privacy concern, but third-party surplus is lower (zero) where there are no users with relatively low privacy concern.

Thus, our main results continue to hold where websites are asymmetric in terms of their intrinsic value for users.

6.2. Multihoming Users

In our main model, we consider users to single-home. That is, they use only one website: the one that provides the highest utility to them. Here, as an extension to our main model, we study the scenario where users can multihome. We use a model of multihoming consistent with Choi (2010). We focus our analysis on two scenarios: First, we consider the scenario where third-parties enjoy the same benefit from the data from both single-homing and multihoming users. Then, we consider the scenario where multihoming users' data are more valuable than the single-homing users.

We make a few additional assumptions in the multihoming setting. First, we consider the decision of homing to be endogenous—that is, users access more

than one website only if both websites provide positive utility. Second, we assume that the total time that users can spend on all websites is fixed. This implies that if a user single-homes, then all of her attention is spent on one website. On the other hand, if a user multihomes, then she divides her attention among the websites. Further, this division is proportional to the utility that the user receives from each website. In other words, users spend more time at the website that provides more utility to them. For example, if a user gets twice as much utility from website 1 than website 2 (because website 1 better suits their needs), then they spend twice as much attention (or time) on website 1 than website 2, and this is reflected in the utility they gain from each website. In this case, when deriving values for the payoff to websites and third-parties in equilibrium, we consider the amount of attention received from users, rather than number of users. Moreover, we assume that the users' price and cost of privacy are according to the amount of time or attention they spend on each website, as they can choose from different subscription tiers, and the data they generate for third-parties correlate with the amount of time or attention spent. These assumptions are consistent with the way users visit websites and form part of our contribution.

Using this setup, similar to the single-homing case, a user's utility for each website is given as in (2). We assume that if a user accesses one website exclusively, then they get the whole utility from that website—that is, $U_{i,k}^j(y)$. However, if they use both websites, then the amount of time and attention they give to each website is according to the utility they get from each website. The user utility in this case is given as:

$$U_k^j(y) = \begin{cases} U_{1,k}^j(y), & \text{if single-home website 1} \\ U_{2,k}^j(y), & \text{if single-home website 2} \\ \frac{U_{1,k}^j(y)}{U_{1,k}^j(y) + U_{2,k}^j(y)} U_{1,k}^j(y) \\ + \frac{U_{2,k}^j(y)}{U_{1,k}^j(y) + U_{2,k}^j(y)} U_{2,k}^j(y), & \text{if multihome.} \end{cases} \quad (7)$$

Using these assumptions, as provided in the proof, we find that all equations are the same as the single-homing scenario, and, therefore, the results from the multihoming scenario are identical to those of single-homing. Assuming that the revenue of websites and third-parties are derived using the same assumptions as the main model, we find that this setup for multihoming does not impact our results, as provided in our next proposition.

Proposition 6 (Multihoming Users). *Where users can multihome according to their preference for websites, and third-parties do not differentiate between single-homing*

and multihoming users, all of our results from the main model continue to hold.

In other words, the comparison of Benchmark, Consent-Based, and Website Subsidization policies does not change if users multihome.

Next, we consider a variation of our model to better capture the nature of user data in the multihoming case. In the online advertising industry, data gathered from multiple sources allows third-parties to form a better picture of users and to target them with higher accuracy. This implies that users that multihome may be more valuable to such third-parties. To capture this in our model, we consider the data that multihoming users generate to be more beneficial to third-parties. That is, rather than considering the benefit from all users to be B , as is the case for single-homing, we assume that single-homing users allow for a benefit of B for third-parties, whereas multihoming users generate a benefit of $B_{multi} > B$. As multihoming does not impact the revenue for third-parties in Website Subsidization (the subsidized website does not share user data with third-parties), we focus on the Benchmark and Consent-Based policy for this analysis. The addition of multihoming users makes our analysis quite complicated, and some results cannot be determined analytically, but we can confirm some of our main findings.

Proposition 7 (Multihoming Users with Higher Value to Third-Parties). *Where users can multihome according to their preference for websites and third-parties benefit more from multihoming users than single-homing users, we find that*

- a. *If there are at least some users with relatively low privacy concern, then the Consent-Based policy increases the number of third-parties set by websites and reduces the price compared with Benchmark.*
- b. *If there are no users with relatively low privacy concern, then the Consent-Based policy decreases the number of third-parties set by websites (down to zero) and increases the price compared with Benchmark.*

7. Empirical Investigation

In this section, we intend to corroborate our analytical findings by empirically examining real-world data on websites' behavior. Note that we have modeled three different policies, of which only the Benchmark and Consent-Based policies exist in the real world. As we will show, although we cannot empirically examine all the insights from our analytical models, those insights that are empirically testable are in line with our analytical findings.

We begin by developing a set of hypotheses and propositions based on our analytical findings. To test these, we first use Difference in Differences (D-i-D) and Interrupted Time Series (ITS) analyses to draw causal inferences about the impact of the Consent-

Based policy on the number of third-parties. We then conduct an exploratory analysis to investigate how the prevalence of nonprofit websites in a particular market is associated with its level of privacy concern.

7.1. Hypotheses Development

In Proposition 1, we find that under the current settings where there are at least some users with low privacy concern, the Consent-Based policy increases the default number of third-parties set by the websites. Given this insight, we hypothesize that once a consent-based policy is implemented, websites tend to increase the number of third-parties with which they intend to share their users' data. Put formally, our first hypothesis is as follows:

Hypothesis 1. *Implementation of the consent-based policy increases the default number of third-parties set by the websites.*

Result 1 presented in Section 4 indicates that the minimum threshold value of the nonprofit website's comparative utility provision, where user surplus under Website Subsidization is higher than the Benchmark and the Consent-Based policy, decreases with privacy concern. In an empirical setting, we can shed light on this result by examining the prevalence and utility provision of nonprofit websites in categories with different levels of privacy concern. In this context, we define privacy-sensitive websites as those that provide privacy-sensitive information and services, where users may have higher concerns about their privacy. With that definition, we can argue that, for example, in the healthcare context, websites focusing on substance abuse are more privacy-sensitive than websites providing information about animal health. Specifically, we expect to observe the following outcomes:

- The proportion of nonprofit websites is higher in markets with higher privacy concern.
- The comparative utility provision of nonprofit websites is lower in markets with higher privacy concern.

7.2. Empirical Analysis of the Consent-Based Policy's Impact on Third-Parties

California's CCPA creates rights for California residents to access, correct, delete, and opt out of the sale of personal information. More specifically, the CCPA prohibits websites from selling their users' data to third-parties, unless they have obtained explicit consent from them. Because the CCPA highly resembles the Consent-Based policy in the analytical models, we use its implementation as a natural experiment to test Hypothesis 1 about its impact on the number of third-parties.

To collect data for our analyses, we first used Alexa.com's Application Programming Interface (API) to create a list of the top 100,000 websites according to their rankings. We then created an automated crawler

that visited each website in the list on a daily basis and collected data on the number of the third-parties used. Websites optimize their revenue from data sharing by selectively choosing the third-parties with whom they share data. To ensure that websites' behavior is unaffected by our crawling history, our automated algorithm cleared the cookies and browser history after every website visit. Therefore, the third-parties that our crawler identifies for each website are the default third-parties that a website sets for a new user. The crawler was installed on two virtual servers: one with a California Internet Protocol (IP) address and one with a New York IP address. Starting December 17, 2019, the crawler visited every website from both IP addresses on a daily basis for a 32-day period until January 17, 2020. This lets us compare the behavior of websites that were visited from IP addresses both inside and outside of California over a period of time before and after CCPA went into effect. Because CCPA pertains only to California residents, we can consider websites' behavior when they are visited from a California IP address as our treatment group and their behavior when they are visited from a New York IP address as our control group.

To identify the types of websites, we use their domain extensions. Websites with extensions “.com,” “.net,” “.io,” “.co,” “.us,” “.tv,” “.club,” “.biz,” and “.ai” are categorized as “commercial.” Websites with extensions “.gov,” “.org,” “.edu,” “.info,” and “.me” are categorized as “nonprofit.” Websites with any other extension are categorized as “other.”

It is important to note that in our empirical analysis, we only observe the default number of third-parties that the website sets in its menu, and not the choice of users, as our crawler does not reject any third-parties in the process of visiting these websites. This is consistent with our analytical model, including those in Proposition 1 on the default number of third-parties set by the websites. Table 2 presents the number of third-parties before and after CCPA implementation in California and New York across various subcategories of websites.

We begin our empirical examination with a D-i-D analysis at the website-day level. We specify the following equation that includes both website and day dummies, mimicking the approach by Agrawal and Goldfarb (2008):

$$Y_{it} = \alpha + \beta CCPA_{it} + \theta_i + \delta_t + \epsilon_{it}. \quad (8)$$

In this equation, Y_{it} represents the number of third-parties that website i has on day t . $CCPA_{it}$ is our key explanatory variable and is equal to one if the law applies to website i on day t . That is, $CCPA_{it}$ is equal to one if it satisfies the following conditions: First, the website is visited from a server with a California IP address; and second, the website is visited after January 1, 2020 (the day on which the CCPA went into effect). The website and day fixed effects are

Table 2. Summary Statistics on the Number of Third-Parties

	Observations	Mean	Standard deviation
California			
Before CCPA			
All	1,398,310	85.33	131.52
Commercial	1,149,165	89.71	134.93
Nonprofit	138,407	54.04	95.24
Other	110,738	78.96	129.17
After CCPA			
All	1,584,179	83.44	123.67
Commercial	1,301,964	87.56	126.25
Nonprofit	156,828	53.5	93.43
Other	125,387	78.13	124.40
New York			
Before CCPA			
All	1,398,310	83.33	122.65
Commercial	1,149,165	87.37	125.45
Nonprofit	138,407	53.19	88.95
Other	110,738	78.96	129.17
After CCPA			
All	1,584,179	80.81	115.49
Commercial	1,301,964	84.63	117.40
Nonprofit	156,828	52.21	90.22
Other	125,387	76.92	118.36

represented by θ_i and δ_t , respectively. We first conduct this analysis using the whole population of websites and then repeat it on subsets of “commercial,” “nonprofit,” and “other” categories.

As Table 3 shows, the CCPA variable's coefficient is positive for all website categories, indicating a causal increase in the average number of third-parties of the treatment group after introducing the CCPA. Such an effect is highly significant ($p < 0.001$) for all categories except for nonprofit websites, in which the effect is almost significant ($p = 0.1060$). It is important to ensure that the trends before the implementation of the shock are parallel across the two groups. Although we can visually inspect the trends and examine the parallel trends assumption, it is also possible to statistically test this assumption following the method proposed by Autor (2003). Specifically, we create indicator variables for the seven days prior to the CCPA implementation date ($CCPA_{t-7}, \dots, CCPA_{t-1}$). Note that $CCPA_{t-7}$ is equal to one for seven periods or more before the CCPA implementation date. We also created 11 indicator variables ($CCPA_{t+1}, \dots, CCPA_{t+11}$) for the 11 days postimplementation date of CCPA. Note that $CCPA_{t+11}$ is equal to one for 11 days or more after the CCPA implementation date. We then have these dummy variables interact with our treatment indicator, while eliminating the time period right before the shock ($CCPA_{t-1}$), so that we can use it as our baseline. As Autor (2003) mentions, this technique lets us examine the possible anticipatory responses leading to the implementation date, in effect allowing us to statistically test the assumption of parallel trends prior to the shock. We present the

Table 3. Estimates of the CCPA Implementation’s Effects on the Number of Third-Parties

Number of third-parties	All websites	Commercial	Nonprofit	Other
CCPA	0.6594*** (0.0658)	0.6152*** (0.0701)	0.4402 (0.2724)	1.3934*** (0.2125)
Intercept	36.3616*** (7.0827)	29.1254*** (6.8399)	37.7019*** (9.2396)	-1.2186 (6.4412)
R ²	0.8977	0.9086	0.6880	0.9165

****p* < 0.01.

estimation results of the leads and lags model for all of the website subcategories in Table 4. We also graphically present the results in Figure 7.

The coefficients on the shock leads are close to zero and statistically insignificant, showing no evidence of an anticipatory response from websites about the law’s looming implementation date. This observation

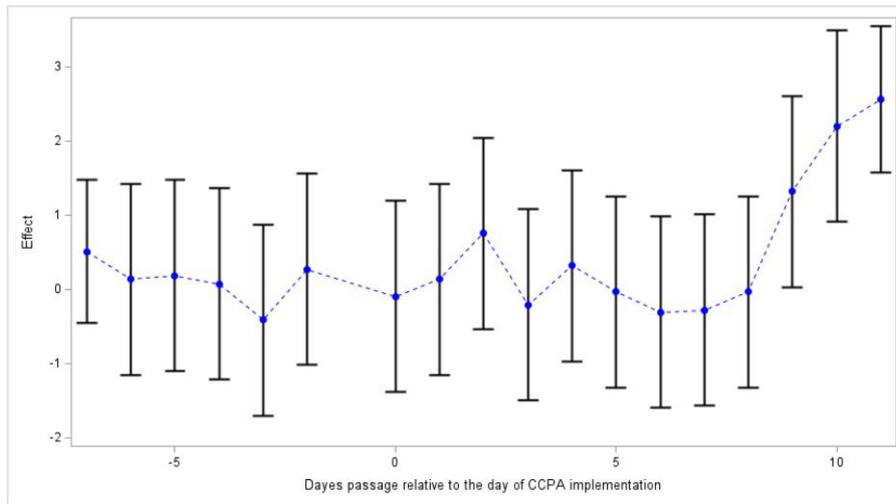
confirms the visual evidence for the validity of the preshock parallel trends assumption. As shown in Figure 8, the trends in daily average number of third-parties in California and New York are parallel to each other, not only in the period leading up to the CCPA implementation on January 1st, but also over the next eight days until January 9th. This is because the number

Table 4. CCPA’s Estimated Impact on the Number of Third-Parties Before and After January 1, 2020

Number of third-parties	All websites	Commercial	Nonprofit	Other
CCPA _{t-7}	0.512 (0.490)	1.089 (0.996)	0.078 (0.415)	0.364 (0.999)
CCPA _{t-6}	0.135 (0.657)	0.256 (1.336)	0.226 (0.557)	-0.093 (1.341)
CCPA _{t-5}	0.186 (0.657)	0.779 (1.336)	0.109 (0.557)	-0.340 (1.341)
CCPA _{t-4}	0.076 (0.657)	-0.004 (1.336)	0.058 (0.557)	0.171 (1.341)
CCPA _{t-3}	-0.413 (0.657)	-0.245 (1.336)	-0.132 (0.557)	-0.883 (1.341)
CCPA _{t-2}	0.271 (0.657)	1.094 (1.336)	0.078 (0.557)	-0.384 (1.341)
CCPA _{t0}	-0.096 (0.657)	-0.205 (1.336)	0.352 (0.557)	-0.430 (1.341)
CCPA _{t+1}	0.138 (0.657)	-0.942 (1.336)	0.110 (0.557)	1.241 (1.341)
CCPA _{t+2}	0.759 (0.657)	0.826 (1.336)	0.160 (0.557)	1.298 (1.341)
CCPA _{t+3}	-0.205 (0.657)	-0.280 (1.336)	0.207 (0.557)	-0.536 (1.341)
CCPA _{t+4}	0.318 (0.657)	0.120 (1.336)	-0.097 (0.557)	0.911 (1.342)
CCPA _{t+5}	-0.030 (0.657)	-0.044 (1.336)	0.057 (0.557)	-0.100 (1.341)
CCPA _{t+6}	-0.303 (0.657)	-1.468 (1.336)	-0.221 (0.557)	0.801 (1.341)
CCPA _{t+7}	-0.278 (0.657)	-0.143 (1.336)	-0.256 (0.557)	-0.453 (1.341)
CCPA _{t+8}	-0.029 (0.657)	-0.014 (1.336)	-0.642 (0.557)	0.575 (1.341)
CCPA _{t+9}	1.321* (0.657)	0.551 (1.336)	0.783 (0.557)	2.631** (1.341)
CCPA _{t+10}	2.200*** (0.657)	1.719 (1.336)	1.095** (0.557)	3.782*** (1.341)
CCPA _{t+11}	2.563*** (0.502)	2.577** (1.021)	1.290*** (0.425)	3.834*** (1.025)
Intercept	-0.835 (5.041)	66.964*** (5.986)	22.713*** (2.497)	-2.642 (5.991)
R ²	0.9363	0.9306	0.9622	0.9314

p* < 0.1; *p* < 0.05; ****p* < 0.01.

Figure 7. (Color online) Day Passage Relative to the CCPA Implementation Date

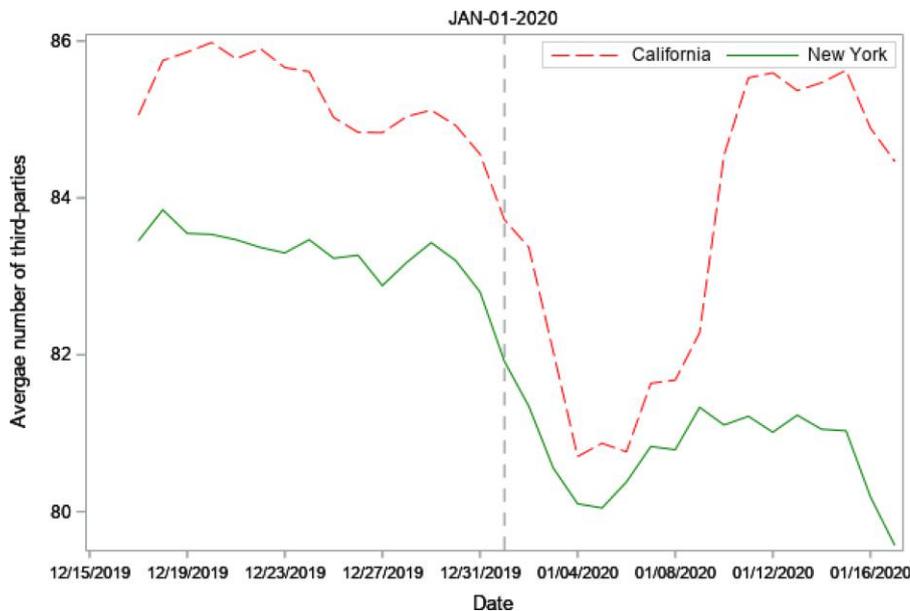


of third-parties in New York dropped with those in California since the implementation of the law. However, although the websites resumed their normal data-sharing practices for users from California within eight days, they maintained their reduced number of third-parties for users in New York. The empirical results in Table 4 support this observation. As reported in Table 4, for the eight days since January 1st, we do not observe any statistically significant difference in the number of third-parties for any of the website categories. However, for days 9, 10, and 11 and afterward, we observe a significant difference in the number of third-parties. This indicates that the D-i-D results reported earlier in Table 3 are driven by

the differences that are materialized in the second week in the post-CCPA adoption period.

Although the D-i-D method compares the changes in the average value of the dependent variable before and after a shock between the control and independent groups, it does not shed light on how such changes occur. This is especially important when the before and after periods are relatively long, as they are in our case. We have time-series data on the websites' third-parties over a 32-day period, equally divided before and after the introduction of CCPA. Although our D-i-D results confirm that the number of third-parties is larger in the period after the CCPA

Figure 8. (Color online) Daily Average Number of Third-Parties



implementation, they cannot specify whether such an increase happened immediately after the introduction of the shock or resulted from a gradual increase over the 16-day period after the shock.

The ITS methodology bridges this gap, as it can parse out the changes by estimating both an intercept and a slope after implementing the shock. The intercept would signify the immediate changes, whereas the slope would show the gradual changes in the postshock period. Such insights would have valuable practical implications for policy makers, as they can understand the mechanism through which a policy shift would exhibit its effects.

It is important to note that although the set of websites are identical, the users who visit them are different. The CCPA only regulates websites' behavior toward users in California, and, therefore, for a particular website, the number of third-parties with whom it shares its California users' data would be considered as observations from the treatment group, and the number of third-parties with whom the same website shares its New York users' data would be considered as observations from the control group. ITS analysis examines whether the difference between the average number of third-parties changes after implementing CCPA. As discussed earlier, the advantage of ITS to the more common D-i-D approach is that it lets us study gradual changes and see whether policies' effects diminish over time. It also provides better estimates, as the method considers and adjusts for correlations between observations from the same websites over multiple consecutive time periods.

ITS is essentially a form of time-series analysis. When individual-level data are available in a form of a panel data set, a typical approach in ITS analysis is to average the data at each time point to transform the panel data structure into a time-series data structure and then model the time series over these time points. In other words, all outcome variable measurements available from individuals are averaged at each time point, and then these averages are used as population-level data for performing the ITS analysis (Hudson et al. 2019, Bazo-Alvarez et al. 2020).

In following two ITS analyses, the dependent variables (Y_t) are the daily average number of third-parties in treatment (California IP address) and the *difference* in the daily average number of third-parties in treatment and control (New York IP address) groups, respectively. We fit the dependent variable in each period as a function of three main explanatory variables. The first is a continuous variable (t) that counts the periods since the time series' start. This variable's coefficient captures time trends. The second is a binary variable ($CCPA_t$) that indicates the shift in policy. This variable is equal to one if the period is after the CCPA's implementation (January 1, 2020) and zero (zero) if it

is before. The binary variable's coefficient indicates whether there is a change in the outcome variable immediately after the CCPA went into effect. The third is a continuous variable ($TimeAfterCCPA$) that counts the number of periods (days) after the CCPA went into effect. This variable's value in periods before the rules' implementation is equal to zero. This variable's coefficient indicates whether a change occurs in the slope of outcomes after implementing the rules compared with the trend in the preimplementation period. To account for observations' correlation in consecutive periods, we follow the recommendations of Penfold and Zhang (2013) to apply the Durbin Watson test and, when necessary, account for autoregressive terms (Y'_{t-i}) in the model.

$$Y_t = \beta_0 + \beta_1 CCPA_t + \beta_2 t + \beta_3 (TimeAfterCCPA) + \alpha Y'_{t-i} + \varepsilon_t \quad (9)$$

We first estimate (9) using the daily average of the number of third-parties when visited from a California IP address as our dependent variable. The results are presented in the first panel of Table 5. We find no statistically significant decrease in the number of third-parties on January 1, 2020, when the CCPA went into effect. More interestingly, we observe a small increase in the slope of the number of third-parties (0.2196, $p = 0.0599$) after the act's implementation. The next three panels in Table 5 present the results of the analysis when we separately estimate the model for different types of websites. Except for the nonprofit websites, no statistically significant change occurs in the number of third-parties immediately after implementing the CCPA.

In Table 5, the dependent variable is the daily average of the number of third-parties with whom websites share the data of their California users. Although this specification utilizes temporal variations in the average number of third-parties before and after the exogenous shock of CCPA implementation, it cannot rule out the effects of unobserved factors on the number of third-parties. To address this issue, in Table 6, the dependent variable is, instead, the difference between the daily averages of the number of third-

Table 5. ITS Estimates for Websites Visited from California IP Addresses

California	All websites	Commercial	Nonprofit	Other
CCPA	-0.6607 (0.5676)	-0.5078 (0.5936)	-2.0146*** (0.5124)	-0.7503 (0.6284)
t	-0.1720** (0.0732)	-0.0663** (0.0292)	-0.0348** (0.0143)	0.001104 (0.0270)
Time after CCPA	0.2196* (0.1116)	0.0847* (0.0461)	0.1180*** (0.0169)	0.0175 (0.0411)
Intercept	86.2711*** (0.7110)	90.6869*** (0.8133)	54.9126*** (0.3693)	78.4846*** (0.7932)

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

parties with which websites share their California and New York users’ data. Under this specification, we assume that the unobservable factors that could have affected the number of third-parties for California users would have had the same effect on the third-parties that receive New York users’ data. We provided formal statistical evidence for examining the parallel trends assumption in Table 4 and Figure 7. We can examine this assumption further by visually inspecting the trends presented in Figure 8. This figure presents a visual confirmation of this assumption by showing that the two groups have quite similar trends, as they remain parallel to each other up until the exogenous shock date. Although we cannot observe such factors, because we take the difference between these two values as the dependent variable, such unobserved effects will be canceled out, and we expect the subsequent results to be unbiased. Immediately after implementing CCPA, we observe no statistically significant change in the difference between the average number of third-parties, whether it is for all websites or for any website subcategories.

Because the trend post-CCPA is the sum of coefficients of time (t) and *Time after CCPA* variables, we apply the Wald test to examine whether the sum of their coefficients is equal to zero ($\beta_2 + \beta_3 = 0$). The results are highly significant for all four models in Table 6 and confirm that the trend after the CCPA is indeed positive, and, thus, the number of third-parties increase over time.

To further check our results’ robustness, we redid our ITS analysis presuming different intervention dates. As Table 7 shows, presuming that the intervention occurred prior to January 1 does not capture the CCPA’s effect, indicating that no alterations in website behavior occurred in anticipation of the law. On the flip side, presuming intervention after January 1 consistently indicates a significant increase in the number of third-parties throughout the postintervention period. This is consistent with findings from the leads and lags analysis presented in Table 4.

Figure 9 shows the magnitude of the CCPA’s effect on increasing third-parties. As discussed earlier, to

Table 6. ITS Estimates for Websites Visited from California and New York IP Addresses

California–New York	All	Commercial Nonprofit	Other	
CCPA	−0.8059 (0.5489)	−0.6762 (0.5468)	−0.8456 (0.5192)	−0.3583 (0.7620)
t	−0.0813 (0.0556)	−0.0312* (0.0181)	0.004348 (0.0146)	−0.009323 (0.0311)
Time after CCPA	0.3191*** (0.0794)	0.1145*** (0.0249)	0.0413** (0.0190)	0.1011** (0.0478)
Intercept	2.5342*** (0.5191)	2.9361*** (0.4760)	0.7553* (0.3855)	−0.3314 (0.9423)

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 7. ITS Analysis Results Presuming Different Shock Dates

California–New York	$t - 6$	$t - 3$	$t + 3$	$t + 6$
CCPA	−0.5049 (0.5997)	−0.6237 (0.5631)	−1.6798*** (0.4045)	−0.3283 (0.4264)
t	−0.0237 (0.1568)	−0.0872 (0.0985)	−0.0448* (0.0253)	−0.0663*** (0.0172)
Time after CCPA	0.1697 (0.1903)	0.2563* (0.125)	0.4488*** (0.0456)	0.5028*** (0.0604)
Intercept	1.7316 (1.2656)	2.4838** (0.882)	2.3849*** (0.2957)	2.5747*** (0.2301)

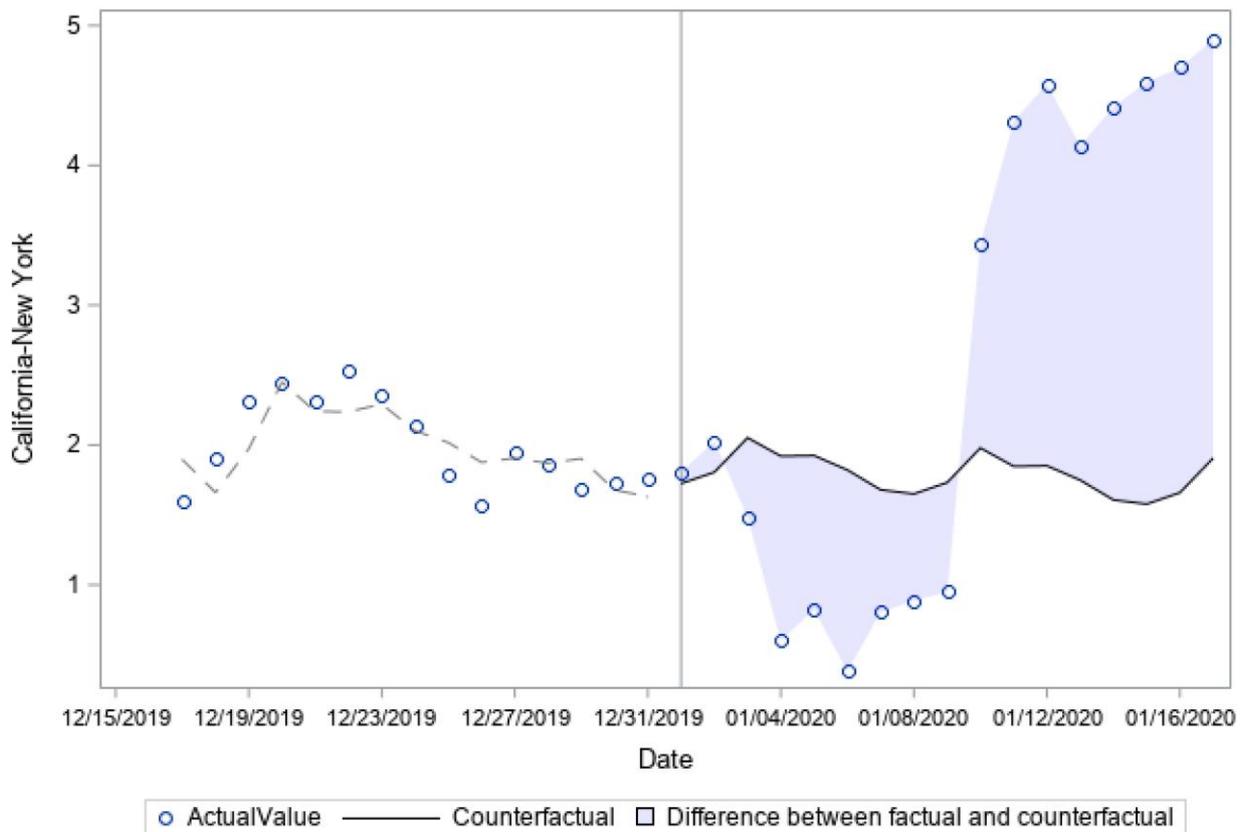
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

remove the bias of unobserved variables, we examine the difference in the number of third-parties of each website when visited from a California IP address from when it is visited from a New York IP address. As presented in the figure’s left half, prior to the law’s implementation, the difference between the two is stable, hovering around an average of two. However, after January 1, 2020, when the CCPA went into effect, this stable trend is disrupted. Despite initial reduction, we observe a significant upward slope, resulting in a net increase in the number of third-parties. The solid line in the graph shows the counterfactual difference in the number of third-parties had the CCPA not gone into effect based on forecasts of the trends prior to the CCPA implementation date. The shaded band shows the difference between the actual and counterfactual values.

These empirical findings confirm Hypothesis 1 and result in Proposition 1, in that a Consent-Based policy results in an increase in the default number of third-parties set by a website.

7.3. Impact of Privacy Concerns

As discussed in Section 7.1, based on the insights presented in Result 1 from our analytical analysis, we expect the nonprofit websites to be more ubiquitous, but have a lower comparative utility provision in markets with higher privacy concerns. To explore these observations, we use the website categories and subcategories provided in “top sites by category” from Alexa.com. We use these categories as an indication of different markets with varying privacy concern. We collected third-party data on all websites that are listed within the “Health” category, which are grouped into 34 subcategories. For each website, we collected data on daily page views per visitor, percentage of traffic from search engines, and number of sites linking into the focal website. We argue that these three variables all measure the comparative utility provision of a website’s content from various aspects. As the comparative utility provision of the content in a website increases, users tend to spend more time on the website and view more pages. The website’s

Figure 9. (Color online) Difference in the Daily Average Number of Third-Parties Before and After CCPA Implementation

relevance and usefulness highly impact search-engine results, making it easier for users to access websites. If a website's content and service are more relevant, then the traffic from search engines increases accordingly. Finally, the more reliable and comprehensive a website's content is, the more likely it is that other sites link into that website. We create a single measure of comparative utility provision for each website based on these three dimensions using Principal Component Analysis and calculate the average comparative utility provision of websites within each of the 34 health subcategories. We also calculate the average comparative utility provision for each of the three website types (commercial, nonprofit, and other) within each of the subcategories. The ratio of the average comparative utility provision of nonprofit websites to the average comparative utility provision of each subcategory indicates the extent to which the comparative utility provision of nonprofit websites is better or worse than the average within each subcategory.

Figure 10 presents the comparative utility provision of nonprofit websites relative to the average comparative utility provision of all websites (shown as the horizontal line above the x-axis) within each subcategory. We observe that the nonprofit websites in privacy-sensitive categories, such as *reproductive health*, *addiction*,

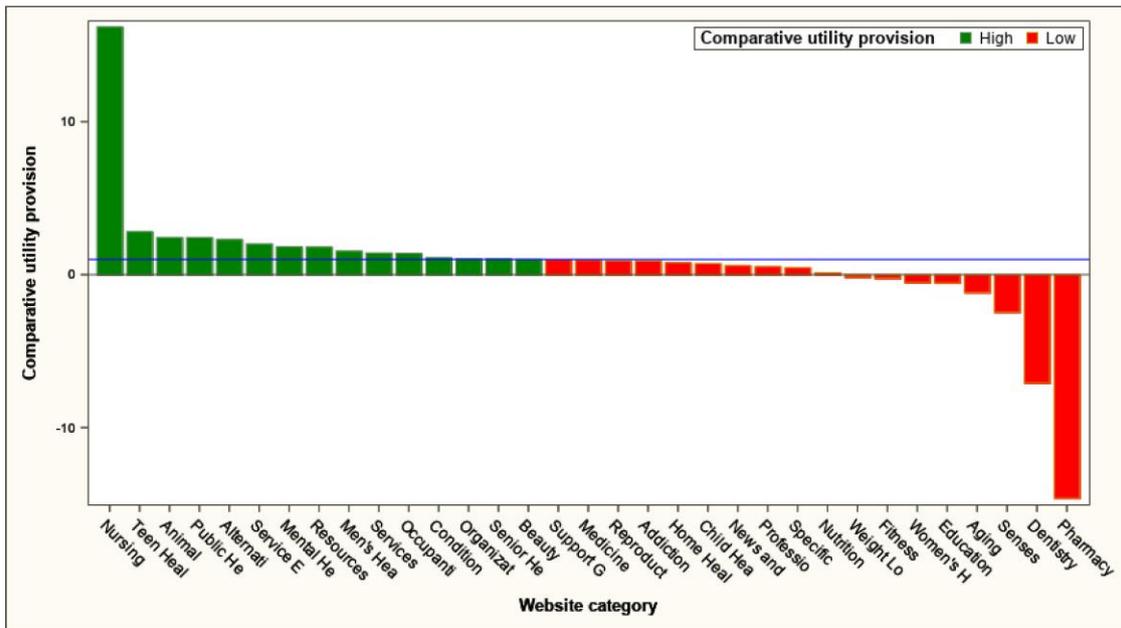
specific substances, *support groups*, and *aging*, have a lower comparative utility provision than the average of respective categories. On the other hand, the comparative utility provision of nonprofit websites in less privacy-sensitive categories, such as *nursing*, *public health*, *animals*, *resources*, and *beauty*, is higher than the average of the respective category. These observations provide support for the supposition that nonprofit websites' comparative utility provision is lower in markets where privacy concern is high.

To study the impact of privacy concern on nonprofit websites' prevalence, we calculate the percentage of nonprofit websites within each subcategory. As Figure 11 shows, we observe that the percentage of nonprofit websites in privacy-sensitive categories, such as *addiction*, *specific substances*, *support groups*, and *aging*, is higher than average, whereas the percentage of these websites in less privacy-sensitive groups, such as *animals*, *resources*, and *beauty*, is lower. This supports our observation that nonprofit websites' prevalence is lower in markets with high privacy concern.

8. Discussion and Conclusion

Third-parties are integral to the internet ecosystem, as they are used by almost all websites. Although third-

Figure 10. (Color online) Comparative Utility Provision of Nonprofit Websites Relative to Their Respective Categories

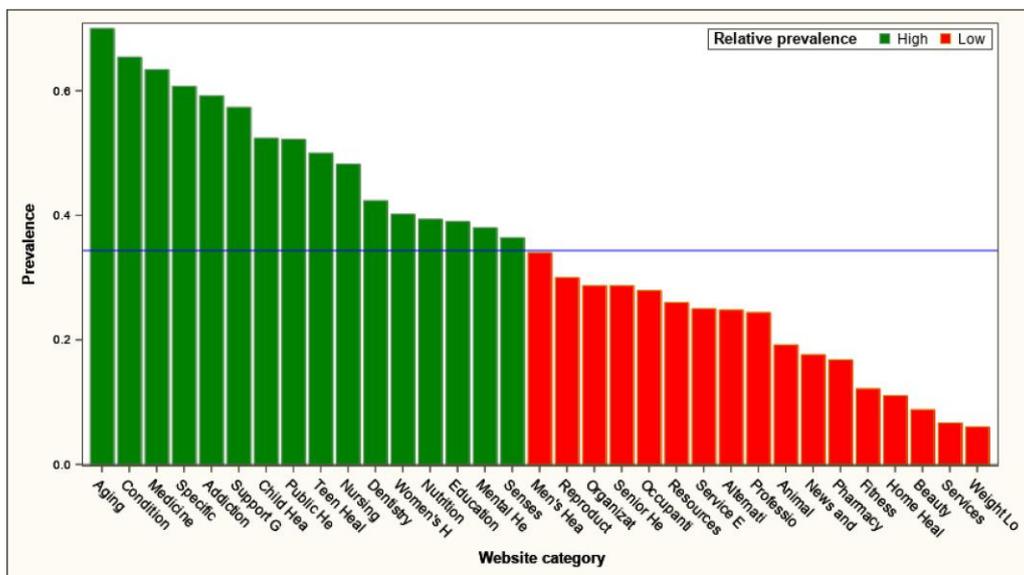


parties can provide value for users, they also pose serious privacy threats to consumers. In recent years, user privacy concerns have heightened, and, in response, governments and policy makers are scrambling to enforce data-protection policies to improve user privacy.

Regulatory mechanisms available to policy makers include consent-based user information sharing and subsidizing a competing website. These policies are intended to improve user surplus and social welfare, yet laws and public policy often have unforeseen

outcomes. Using a stylized analytical model, we examine such policies' impact on the decisions and outcomes of various entities involved, including websites, users, and third-parties. We find that in the absence of market entry and exit—and where at least some users have relatively low privacy concern—even though a consent-based policy may improve user surplus, it has the unintended consequence of increasing the number of third-parties and, thus, sharing user information. Considering the impact of policy making on websites' entry and exit, we find that both Consent-Based and

Figure 11. (Color online) Prevalence of Nonprofit Websites Relative to Their Respective Categories



Downloaded from informs.org by [137.205.8.152] on 16 February 2023, at 04:53. For personal use only, all rights reserved.

Website Subsidization policies may reduce competition by driving websites out of the market to the detriment of users and society.

Based on our findings, we discuss several important implications on data-protection policies for websites, third-parties, and policy makers. Either Consent-Based or Website Subsidization policies can be effective, in the absence of market entry and exit considerations. Under more realistic conditions, in which websites can enter and exit the market, we show that both Consent-Based and Website Subsidization policies may drive websites out of the market, which may affect users negatively.

We also demonstrate that Consent-Based policies where the number of third-parties are not under websites' control are not beneficial to websites, as they decrease their revenue. However, Consent-Based policies are beneficial for third-parties, because users prefer to share their data with third-parties to avoid paying for content. Similarly, Website Subsidization policies decrease website revenues, and their effect on third-parties depends on nonprofit websites' comparative utility provision. We validate a number of the findings from our analytical model through an empirical investigation of the impact of Consent-Based policy on third-parties in a natural CCPA experiment. Specifically, we verify the impact of competition and Consent-Based policies on third-parties. Interestingly, we find that Consent-Based policies have the opposite and unintended effect of increasing the number of third-parties engaged by a website. We also provide an exploratory analysis of the impact of privacy concern on nonprofit websites, where we find that an increase in user privacy concern translates into a higher number of nonprofit websites, but with lower comparative utility provision. These empirical findings also have important implications for governmental data-protection policies.

Going forward, policy makers can use our study to consider the different policy mechanisms at their disposal and choose which one works best for their specific context. Website Subsidization is similar to a scalpel, enabling policy makers to sculpt around and impact specific target markets. However, certain political and business forces may resist introducing publicly funded competitors, as they prevent a level playing field. Consent-Based policies are more comparable to a sledgehammer that uniformly affects all market segments. For example, policy makers can employ a nonprofit website to focus on improving social welfare or user surplus in a highly specific market segment with Website Subsidization or to roll out Consent-Based policies that apply to a much broader set of target markets. For circumstances where it is difficult for the government to enact a law for the entire market, Website Subsidization policies are appealing alternatives, as they may provide even better user surplus than Consent-Based policies.

Finally, there are some limitations to this research. We assume that third-parties only create negative utility for users, whereas in some instances, third-parties provide services that could increase the utility for users. We also do not consider the interaction of third-parties such as ad networks with the websites and their implications. These questions are outside the scope of this paper, and future research can examine privacy-protection policies in markets where third-parties can create positive user utility while simultaneously extracting consumer data for future exploitation or can create platforms for ad networks, which themselves include different externalities. Moreover, our model provides additional insights on policy's impact on website prices, which we were unable to investigate empirically; this would make for interesting avenues of future research.

Acknowledgments

The authors are thankful to anonymous reviewers at Information Systems Research and participants at the International Conference on Information Systems, the INFORMS Annual Meeting, and the Annual INFORMS Society for Marketing Science Conference. This paper draws on research supported by the Social Sciences and Humanities Research Council of Canada.

Endnotes

¹ Even though the EU and the United States are often at opposing ends of the regulatory spectrum, the CCPA, known as California's GDPR-equivalent policy, enforces similar policies to the GDPR on the State of California. Particularly, as it relates to our study, both enforce strict consent and opt-in requirements on companies dealing with user data.

² In the case with two websites, a Hotelling model yields identical results. The reason for using the Salop circular model here is for consistency, as we provide results for markets with more than two websites in Section 5.

³ We extend our model to incorporate multihoming users in Section 6.2 and find that the model with multihoming setting provides similar results to our main model.

⁴ We extend our analysis to asymmetric websites in Section 6.1.

⁵ Our results continue to hold qualitatively if some users receive positive utility from third-parties.

⁶ Third-party access fees may consist of cash payments from the third-party to the website or may be the value received for unpaid services such as Google Analytics and user authentication in exchange for user data.

⁷ The sharing of user information may include the sharing of a user device and browser information and the use of cookies or other tracking technology such as device fingerprinting. The number of third-parties construct that we use in this paper can more generally be considered as the amount of information shared with third-parties, for which the number of third-parties is a proxy.

⁸ Third-party access fee (R_i) and the number of third-parties (N_{T_i}) have a one-to-one relation; therefore, choosing a third-party access fee is equivalent to choosing the number of third-parties and vice versa. In this paper, we use the two interchangeably.

⁹ In this notation, the location of website 1 can be either zero or one, depending on the user's location. If the user is located at $0 \leq y \leq 1/2$, then $Q_1 = 0$, and if the user is located at $1/2 < y \leq 1$, then $Q_1 = 1$.

¹⁰ We relax this assumption in Section 6.1.

¹¹ Example tools for monitoring third-parties include Lightbeam (<https://github.com/mozilla/lightbeam-we>) and Privacy Badger (<https://privacybadger.org/>), and examples for websites publishing information about websites' third-party sharing behavior include WhoIs (whoisxmlapi.com) and WebXRy (webxray.org).

¹² Assuming third-parties to be differentiated based on their per-user revenue (B) yields qualitatively similar results.

¹³ <https://observer.com/2012/09/should-government-web-sites-sell-commercial-advertisements/>.

References

- Agrawal A, Goldfarb A (2008) Restructuring research: Communication costs and the democratization of university innovation. *Amer. Econom. Rev.* 98(4):1578–1590.
- Angwin J (2010) The web's new gold mine: Your secrets. *Wall Street J.* (July 30), <https://www.wsj.com/articles/SB10001424052748703940904575395073512989404>.
- Autor DH (2003) Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *J. Labor Econ.* 21(1):1–42.
- Bazo-Alvarez JC, Morris TP, Pham TM, Carpenter JR, Petersen I (2020) Handling missing values in interrupted time series analysis of longitudinal individual-level data. *Clin. Epidemiol.* 12:1045–1057.
- Choi JP (2010) Tying in two-sided markets with multi-homing. *J. Indust. Econom.* 58(3):607–626.
- Dabrowski A, Merzdovnik G, Ullrich J, Sendera G, Weippl E (2019) Measuring cookies and web privacy in a post-GDPR world. Choffnes D, Barcellos N, eds. 20th Internat. Conf. Passive Active Measurement, *Lecture Notes in Computer Science*, vol. 11419 (Springer, Cham Switzerland), 258–270.
- Economist* (2021) Cook v Zuck: Apple's privacy policy kicks Facebook where it hurts. *The Economist*. Accessed February 10, 2021, <https://www.economist.com/business/2021/02/06/apples-privacy-policy-kicks-facebook-where-it-hurts>.
- Englehardt S, Reisman D, Eubank C, Zimmerman P, Mayer J, Narayanan A, Felten EW (2015) Cookies that give you away: The surveillance implications of web tracking. *WWW'15 Proc. 24th Internat. Conf. World Wide Web* (International World Wide Web Conferences Steering Committee, Geneva), 289–299.
- Geddes R (2003) Technological change and the case for government intervention in postal services. Foldvary FE, Klein DB, eds. *The Half-life of Policy Rationales: How New Technology Affects Old Policy Issues* (New York University Press, New York), 208–224.
- Gopal RD, Hidaji H, Patterson RA, Rolland E, Zhdanov D (2018) How much to share with third parties? User privacy concerns and website dilemmas. *MIS Quart.* 42(1):143–164.
- Head K (1994) Infant industry protection in the steel rail industry. *J. Internat. Econom.* 37(3–4):141–165.
- Hu X, Sastry N (2019) Characterising third party cookie usage in the EU after GDPR. *WebSci'19 Proc. 10th ACM Conf. Web Sci.* (Association for Computing Machinery, New York), 137–141.
- Hudson J, Fielding S, Ramsay CR (2019) Methodology and reporting characteristics of studies using interrupted time series design in healthcare. *BMC Med. Res. Methodol.* 19(1):1–7.
- Johnson G, Shriver S, Goldberg S (2019) Privacy & market concentration: Intended & unintended consequences of the GDPR. Preprint, submitted November 15, <https://dx.doi.org/10.2139/ssrn.3477686>.
- Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, Koczan P, et al (2016) Who owns the data? Open data for healthcare. *Front. Public Health* 4:7.
- Krishnamurthy B, Wills CE (2006) Generating a privacy footprint on the Internet. *IMC'06 Proc. 6th ACM SIGCOMM Conf. Internet Measurement* (Association for Computing Machinery, New York), 65–70.
- Krishnamurthy B, Naryshkin K, Wills C (2011) Privacy leakage vs. protection measures: The growing disconnect. *Proc. Web.* 2:1–10.
- Kumar S, Sethi SP (2009) Dynamic pricing and advertising for web content providers. *Eur. J. Oper. Res.* 197(3):924–944.
- Kumar V, Khattar D, Gairola S, Kumar Lal Y, Varma V (2018) Identifying clickbait: A multi-strategy approach using neural networks. *41st Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval* (Association for Computing Machinery, New York), 1225–1228.
- Libert T (2015) Exposing the hidden web: An analysis of third-party HTTP requests on 1 million websites. Preprint, submitted November 2, <https://arxiv.org/abs/1511.00619>.
- Malhotra NK, Kim SS, Agarwal J (2004) Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Inform. Systems Res.* 15(4):336–355.
- Mayer JR, Mitchell JC (2012) Third-party web tracking: Policy and technology. *2012 IEEE Sympos. Security Privacy* (IEEE, Piscataway, NJ), 413–427.
- McDonald A, Cranor LF (2010) Beliefs and behaviors: Internet users' understanding of behavioral advertising. Preprint, submitted November 23, <https://dx.doi.org/10.2139/ssrn.1989092>.
- Mueller R (2018) 76% ignore cookie banners—the user behavior after 30 days of GDPR. *Amazee Metrics*. Accessed November 23, 2022, <https://www.amazeeetrics.com/en/blog/76-ignore-cookie-banners-the-user-behavior-after-30-days-of-gdpr/>.
- Newman N (2014) How big data enables economic harm to consumers, especially to low-income and other vulnerable sectors of the population. *J. Internet Law* 18(6):11–23.
- Nicas J, Isaac M (2020) Facebook takes the gloves off in feud with Apple. *New York Times* (December 16), <https://www.nytimes.com/2020/12/16/technology/facebook-takes-the-gloves-off-in-feud-with-apple.html>.
- Ovum (2014) Data privacy laws: Cutting the red tape. *Intralinks*. Accessed November 23, 2022, <https://www.intralinks.com/resources/analyst-reports/ovum-report-data-privacy-laws-cutting-red-tape>.
- Penfold RB, Zhang F (2013) Use of interrupted time series analysis in evaluating healthcare quality improvements. *Acad. Pediatr.* 13(6):S38–S44.
- Salop SC (1979) Monopolistic competition with outside goods. *Bell J. Econom.* 10(1):141–156.
- Sørensen J, Kosta S (2019) Before and after GDPR: The changes in third party presence at public and private European websites. Liu L, White R, eds. *WWW'19 World Wide Web Conf.* (Association for Computing Machinery, New York), 1590–1600.
- U.S. Senate (2014) Online advertising and hidden hazards to consumer security and data privacy (Report, Majority and Minority Staff Report Permanent Subcommittee on Investigations, Committee on Homeland Security and Governmental Affairs, Washington, DC).
- Turov J, King J, Hoofnagle CJ, Bleakley A, Hennessy M (2009) Americans reject tailored advertising and three activities that enable it. Preprint, submitted September 29, <https://dx.doi.org/10.2139/ssrn.1478214>.