



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Privacy Preference Signals: Past, Present and Future

Citation for published version:

Hils, M, Woods, DW & Böhme, R 2021, 'Privacy Preference Signals: Past, Present and Future', *Proceedings on Privacy Enhancing Technologies*, vol. 2021, no. 4, pp. 249-269.
<https://doi.org/10.2478/popets-2021-0069>

Digital Object Identifier (DOI):

[10.2478/popets-2021-0069](https://doi.org/10.2478/popets-2021-0069)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Proceedings on Privacy Enhancing Technologies

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Maximilian Hils, Daniel W. Woods, and Rainer Böhme

Privacy Preference Signals: Past, Present and Future

Abstract: Privacy preference signals are digital representations of how users want their personal data to be processed. Such signals must be adopted by both the sender (users) and intended recipients (data processors). Adoption represents a coordination problem that remains unsolved despite efforts dating back to the 1990s. Browsers implemented standards like the Platform for Privacy Preferences (P3P) and Do Not Track (DNT), but vendors profiting from personal data faced few incentives to receive and respect the expressed wishes of data subjects. In the wake of recent privacy laws, a coalition of AdTech firms published the Transparency and Consent Framework (TCF), which defines an opt-in consent signal. This paper integrates post-GDPR developments into the wider history of privacy preference signals. Our main contribution is a high-frequency longitudinal study describing how TCF signal gained dominance as of February 2021. We explore which factors correlate with adoption at the website level. Both the number of third parties on a website and the presence of Google Ads are associated with higher adoption of TCF. Further, we show that vendors acted as early adopters of TCF 2.0 and provide two case-studies describing how Consent Management Providers shifted existing customers to TCF 2.0. We sketch ways forward for a pro-privacy signal.

Keywords: web standards, privacy, do not track, web measurement, advertising, TCF, GDPR, CCPA, NAI, DNT, P3P, GPC

DOI 10.2478/popets-2021-0069

Received 2021-02-28; revised 2021-06-15; accepted 2021-06-16.

1 Introduction

Privacy preference signals are digital representations of how users want their personal data to be processed. These vary from a binary “Do Not Track” signal through to more complex expressions in cookie consent dialogues.

Such signals are intended to influence how entities including websites and third parties process personal data. Web actors may collect privacy preferences in the hope of legitimizing data processing in the eyes of customers or to satisfy legal obligations.

Efforts to standardize privacy preferences go back to at least P3P, which was presented as a prototype to US regulators in 1997 and recommended as a standard by the World Wide Web Consortium (W3C) in 2002. It was adopted by around 20k websites [1], but was criticized by privacy advocates for not establishing consequences for false reporting of privacy practices [2]. Another W3C working group was formed in 2011 to specify the Do Not Track HTTP extension but it was closed before completion, citing the lack of planned support among “the ecosystem at large” [3] as exemplified by the Interactive Advertising Bureau’s withdrawal [4]. The first wave of privacy preference signals is completed by the opt-out cookies [5] created by the Network Advertising Initiative (NAI) as part of a regulatory compromise with the Federal Trade Commission [6]. The NAI never published a specification, the opt-out only concerned a narrow definition of tracking, and very few vendors participated [5].

A second wave of privacy preference signals was prompted by the passage of privacy laws like the EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). For example, the GDPR establishes that an opt-in consent signal may constitute a legal basis for processing personal data providing the consent was “freely given, specific, informed and unambiguous”. These laws prompted research that has largely focused on the interfaces through which opt-in [7–10] and opt-out [11, 12] signals are collected. An ecosystem of actors has emerged to manage the collection of opt-in consent signals on behalf of websites [13]. Often these signals are collected and shared with a pay-for-membership “Global Vendor List”, which has been termed the “commodification of consent” [14].

At this point, skeptics will rightly state that such signals exist in the world of *soft privacy* with no technical guarantees about personal data flows and that we should instead focus on the technologies associated with *hard privacy*. Such skepticism is compelling but should be qualified by the behavior of privacy advocates and AdTech firms. Both sides invested resources in P3P and DNT working groups. The latter posed a threat to AdTech business models as evidenced by the Interac-

Maximilian Hils: University of Innsbruck, Austria,
E-mail: maximilian.hils@uibk.ac.at

Daniel W. Woods: University of Innsbruck, Austria,
E-mail: daniel.woods@uibk.ac.at

Rainer Böhme: University of Innsbruck, Austria,
E-mail: rainer.boehme@uibk.ac.at

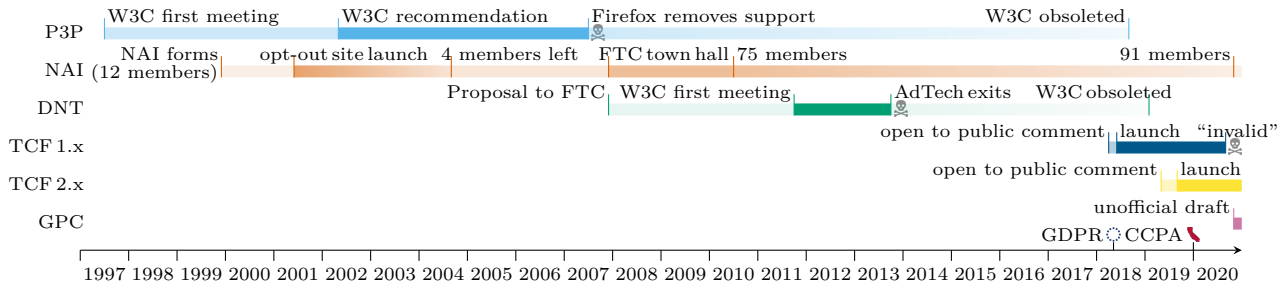


Fig. 1. Timeline of key events for privacy preference signals and relevant laws

tive Advertising Bureau withdrawing from the working group after Microsoft announced it would be turned on by default [4]. The power of these signals can also be seen in websites’ dark patterns that nudge users towards expressing certain preferences [9–11]. Given the stakes have been further increased by sanctions associated with the GDPR and the CCPA, widespread adoption of a privacy preference signal would have privacy implications.

In terms of technical design, there is disagreement over who controls the interface by which users set privacy preferences. In both P3P and DNT, the user expresses preferences to a user agent. In contrast, user preferences are collected by embedding an interface in a web page in both of the approaches developed by AdTech industry bodies, namely the Interactive Advertising Bureau (IAB) [8] and the Network Advertising Initiative (NAI) [5]. This bypasses browsers by making the signal backwards compatible with existing technology. Turning to semantics, AdTech vendors proposed opt-in signals that could represent compliance, whereas privacy advocates proposed (global) opt-out signals that empower users. In summary, these signals have a long history and also have privacy implications going forward.

This paper systematizes historical knowledge on privacy preference signals (the past), measures which signals have been adopted as of February 2021 (the present), and reflects on adoption strategies for a pro-privacy signal (the future). We show a grim state of affairs for user control over privacy: P3P is obsolete, NAI’s system still has only 75 participating AdTech firms, and the reincarnation of Do Not Track—the Global Privacy Control—has been adopted by less than 10 websites. Meanwhile, the Interactive Advertising Bureau’s TCF 1.x and TCF 2.0 have been adopted by thousands of websites. We then use high-frequency web measurements to build a longitudinal case-study of how adoption and TCF 2.0 migration varied over time, websites and AdTech vendors. Our contributions include:

- **Systematize knowledge** about first wave (P3P, DNT, and NAI opt-out) and second wave (TCF and GPC) privacy preference signals.
- **Measure present day adoption** and show that TCF adoption is roughly comparable to historical P3P adoption among websites, whereas an order of magnitude more AdTech vendors have adopted TCF than all other signals combined.
- **Test explanatory variables** for TCF adoption like website popularity, category, number of embedded third parties, and presence of Google Ads. TCF adoption is higher among websites with closer ties to AdTech.
- **Longitudinal case-study** exploring TCF 2.0 migration strategies among the two most popular Consent Management Platforms, and how the new version changed the legal basis that individual AdTech vendors claim for tracking.

Section 2 describes the five privacy preference signals and Section 3 identifies related work measuring their adoption. This motivates our empirical measurements, which are described in Section 4. Our results describing the present are contained in Section 5. Section 6 discusses the past, present and future of privacy preferences. We conclude in Section 7.

2 Background

This section compares five privacy preference signals in terms of design properties and real-world adoption, which is summarized in Table 1. We selected these signals because they were the most widely adopted among the key stakeholders, namely browsers, AdTech vendors and websites. We do not provide a background on the widespread online tracking that motivate privacy preference signals, such as cookies [15, 16] and other tracking technologies [17–19]. Similarly, we do not consider privacy preserving technologies unless they function to

express privacy preferences, such as when browsers/add-ons collect user preferences and automate sending the signal. We now turn to the five signals. Figure 1 provides an overview of the key events for each signal and Figure 2 provides a visual summary of the signal's flow.

2.1 Platform for Privacy Preferences (P3P)

P3P is one of the earliest privacy preference signals proposed for the Web. A demonstration of a P3P prototype was presented before the FTC in June 1997. The W3C recommended the P3P 1.0 specification in 2002, which describes an XML format to encode a human-readable privacy policy into a machine-readable specification stating the type, recipients and purposes of data collected. Users can define individual privacy preferences, which browsers can cross-check against a website's self-reported P3P policy. Each website's implementation could become arbitrarily complex with different policies for each web page and third-party cookie.

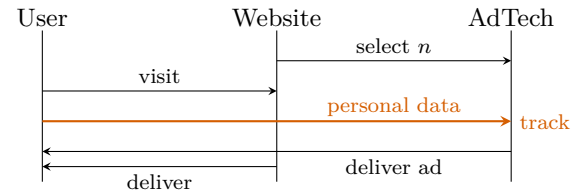
P3P was adopted by, respectively, 588 (10%), 463 (8.34%), 2.3k (2.3%), and 33.1k (60%) of the sites in samples from 2003 [20], 2007 [21], 2007 [22], and 2010 [1]. The final sample [1] is not representative of the wider web because the majority of sites were discovered by the Privacy Finder search engine, which specifically aimed to identify web sites that respect a user's privacy. However, the finding of 19 820 websites [1] implementing P3P in 2010 serves as a reasonable lower bound in Table 1. The same study [1] found that 11 (15%) of a sample of AdTech vendors had a P3P privacy policy.

Microsoft was the only browser developer to fully adopt P3P and stopped support in 2016. Mozilla supported only some P3P features, but removed them by 2007. Other browsers shunned P3P and instead allowed users to set defaults like blocking all third party cookies [23]. P3P-specific browser extensions provide a more meaningful perspective on conscious user adoption than usage statistics for each browser. For example, Privacy Bird, an add-on for Internet Explorer 5 and 6 that displays a website's P3P policy in an easy to understand language, was downloaded 20k times in the first 6 months [24].

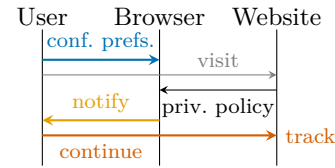
2.2 Network Advertising Initiative (NAI) Opt-Out

AdTech vendors founded a self-regulatory body, the NAI, as a compromise following the Federal Trade Commission's (FTC) report on web privacy submitted to Congress in 1998 [6]. The NAI established a system of

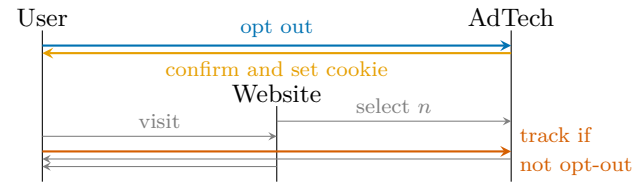
Baseline: Personal data flow in web advertising



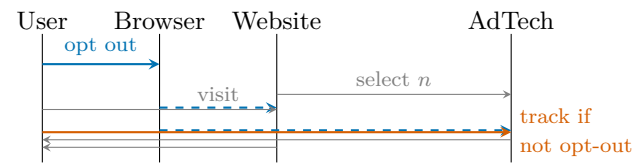
P3P privacy preference signal



NAI opt-out privacy preference signal



DNT/GPC privacy preference signal



TCF privacy preference signal

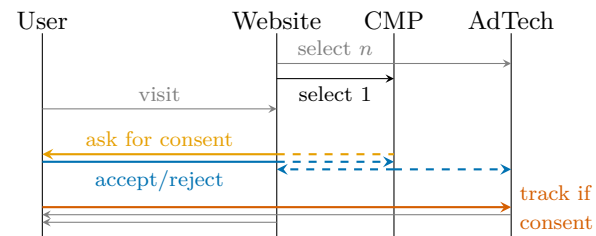


Fig. 2. User prompt, privacy preference signals, and personal data flows when using each approach.

opt-out cookies. Users can visit the NAI's website¹ and set an opt-out cookie for each participating vendor to signal that the user does not want to be tracked by that firm. Critics [5] note that the NAI's narrow definition of tracking would not cover many techniques observed in the wild [17–19].

¹ <https://optout.networkadvertising.org/>

The NAI provide a list of all participating vendors, which was just 4 in 2004, 75 in 2010 [1] and stands at 75 participating vendors as of January 2021. Websites and browsers do not need to adopt the NAI’s system because it piggy-backs on existing browser cookie functionality. The NAI reported one million visits to the the opt-out page in 2006 [5] but we cannot differentiate unique visitors. Returning to browser extensions, there were at least 44.9k users of the Targeted Advertising Cookie Opt-Out (TACO) add-on², which maintained an up to date list of opt-out cookies.

2.3 Do Not Track (DNT)

Acknowledging the failure of P3P, the W3C created a working group in 2011 to standardize the Do Not Track (DNT) mechanism [25]. DNT was less expressive than P3P. Implementation involved browsers sending a DNT: 1 header with each HTTP request to signal that their user did not wish to be tracked. Stakeholders disagreed on whether DNT should default to on or off [4, 26]. This opposition was part of the reason why the W3C working group was closed without success in 2019 [3].

DNT was implemented in browsers by Microsoft, Apple, Mozilla and eventually Google [27]. Websites and third-party vendors could signal in an HTTP response header if they respected the user’s DNT signal. This signal was not exposed in any browser’s user interface³ (outside of add-ons), which meant users were largely unaware of website adoption. Only 9 companies issued public statements regarding support of DNT [28]. In 2011, Mozilla reported DNT adoption by Firefox users to be at 17% in the US and 11% outside [29], although this oversamples privacy aware users.

2.4 Global Privacy Control (GPC)

The unofficial GPC draft specification [30], which was released in October 2020, continues the work of DNT in extending HTTP requests with a single bit value. Perhaps the most important change is re-framing *Do Not Track* as a “Do Not Sell” and “Object To Processing” signal, which is closer to the language of the GDPR

and the CCPA, which became effective in May 2018 and January 2020. This means GPC references (enforceable) laws, which DNT lacked.

As of February 2021, Mozilla and the Brave browser are listed as publicly supporting GPC, but only Brave have implemented it. We do not provide any estimates for user size given it was released so recently.

2.5 Transparency and Consent Framework (TCF)

After the enactment of the GDPR, an advertising industry body (IAB Europe) formed a working group to develop the Transparency and Consent Framework (TCF), “the only GDPR consent solution built by the industry for the industry” [31]. Participants predominantly representing private firms from the advertising and publishing industries co-developed the TCF, which defines the legal terms and data processing purposes that users consent to and the format by which consent signals are stored and exchanged between third parties. A new version (TCF 2.0) was introduced in 2020.

TCF is implemented by websites in the form of a consent dialog that does not require browser buy-in, much like NAI. It creates the role of *Consent Management Providers (CMPs)*, who implement the framework on individual websites. CMPs are central to the TCF in providing an interface between website, user, and ad vendors. They provide websites with a (customizable) cookie prompt to embed, store users’ choices as browser cookies, and provide an API for advertisers to access this information. We refer to Hils et al. [13, Fig. 2] for a visual depiction of the ecosystem.

The IAB maintains a public list of CMPs, which lists 119 participating providers as of February 2021.⁴ A website wishing to implement the TCF independently must become a CMP, otherwise they can out-source this to an existing CMP. In reality, a handful of CMPs dominate the market [8]. The largest CMPs are OneTrust and Quantcast, which account for 37.4% of all CMP implementations in the Tranco 100k (see Section 5).

To receive TCF consent signals from CMPs, AdTech vendors must register with the IAB and pay a yearly maintenance fee to join the *Global Vendor List (GVL)*⁵. As of Feb. 2021, 684 companies are registered on this list. Most CMPs collect consent for the entire GVL by







² <https://web.archive.org/web/20110920055245/https://addons.mozilla.org/en-us/firefox/addon/targeted-advertising-cookie-op/>






³ <https://www.w3.org/TR/tracking-dnt/#responding>

⁴ <https://iab europe.eu/cmp-list/>

⁵ <https://iab europe.eu/vendor-list/>

Table 1. Comparison of privacy preference signals

	P3P	NAI Opt-Out	DNT	GPC	Transparency & Consent Framework	
					TCF 1.x	TCF 2.x
General						
Convened by	W3C	AdTech & FTC	Privacy Advocates		AdTech	
Legal Basis	none	self-regulat.	self-regulat.	CCPA	GDPR	
Standardized by	W3C	NAI	W3C	GPC	IAB	
Design Properties						
Implementation	Privacy Policy XML	Opt-Out Cookie	HTTP Header		Consent Cookie from CMP	
User Interface	UA indicator	central website	UA setting or ext.		dialog on website	
User Decision	configure prefs.	opt out	turn on		select allowed purposes	
Decision Scope	all browsing	cookie lifetime	all browsing		website until re-request	
Vendor Decision	define policy	join NAI	adopt standard		declare processing purposes	
Website Decision	define policy	none	adopt standard		pick vendors	vendors + purposes
Adoption						
Websites	> 20k [1, 20–22]	–	> 9 [28]	?	> 1,539*	> 6,726*
AdTech Vendors	> 11 [1]	> 75 [1]	≈ 0 [28]	?	602	684
Browsers		–	   		–	–

? = unknown, – = compat. with exist. tech., * = in Tranco 100k, see Sec. 5,  Safari  Brave  Chrome  Internet Explorer  Firefox

default, which means privacy preferences apply to the whole list [14].

The specifications of TCF 1.x and TCF 2.0 both define a more complex signal than DNT/GPC. Under TCF 1.x, users may affirmatively consent to any combination of five data processing purposes. They may also state individual preferences for each vendor on the GVL. TCF 2.0 expands this model to ten purposes and two special features, increasing complexity even further.

In both TCF versions, users are prevented from expressing certain preferences. Vendors can claim that they have a legitimate interest in a specific purpose, which serves as their legal basis to process data even if the user clicks “Reject all”. Starting with TCF 2.0, some CMPs provide users with the additional option to object to this processing (GDPR asks for such functionality), but this needs to be done separately in a subdialog. As such, the “Reject all” button commonly does not actually express *all* possible preferences. With TCF 2.x, vendors can declare that their legal basis is flexible. This means they would like to process data with the user’s consent, but they can also perform (limited) processing based on a legitimate interest. As the only exception, TCF 2.x removes the option for vendors to claim a legitimate interest in Purpose 1—“Store and/or access information on a device”—, possibly preempting an intervention by regulators. The policy changes between TCF 1.x and TCF 2.0 motivate measuring the transition.

3 Related Work

Section 3.1 briefly describes the privacy practices employed by websites in order to motivate why privacy preferences matter. Section 3.2 surveys research into privacy preferences including the previous five signals. Section 3.3 links the paper to the general question of *why are technical standards adopted?*

3.1 Privacy Practices

Researchers consistently demonstrate privacy eroding techniques deployed in the wild [15–19] motivated by online advertising business models [32]. Personal data is leaked via social networks [33], third-party web scripts [34], apps [35], software development kits [36], and organizational breaches [37]. The scale of tracking motivate re-designing systems to provide privacy guarantees. For example, multihoming can be used to defend against fingerprinting [38] and trusted hardware can ensure compliance to stated privacy policies [39].

Turning to so-called soft privacy, data processors are constrained by law and social norms. These constraints are far from absolute. For example, half of websites in a 2017 sample violated laws implementing the EU Privacy Directive by installing cookies before collecting user consent [40]. This is likely because organizations do not incur significant costs following data breaches and privacy violations in terms of either regulatory fines or lost shareholder value [41]. Nevertheless, firms’ privacy prac-

tices are *somewhat* impacted by data processors’ self-declared privacy policies [42–45] and even the privacy preferences expressed by users, to which we now turn.

3.2 Privacy Preferences

Interviews [46] and surveys [47, 48] can use natural language to understand users’ actual privacy preferences, which tend to contradict observed behavior [49–51]. Privacy languages aim to express preferences more precisely than natural language. For example, APPEL encodes user preferences to be compared against P3P policies [52]. It could not express acceptable practices nor capture the realities of secondary sharing, which motivated XPref [53] and P2U [54], respectively. Alternative languages focus on the usability for developers [55], enabling audits [56], and providing explanations [57]. Privacy languages have been regularly surveyed by academics [58–61] but unfortunately there has been little adoption in practice [59]. This motivates our focus on signals deployed in the Web ecosystem.

In terms of the first wave of signals, measurements of DNT and NAI opt-out adoption relied on organizations disclosing private data sources like Firefox configurations [29], opt-out web page visits [5], or the NAI’s membership [5]. P3P differed in that website adoption could be quantified via web scraping [1, 20–22, 62, 63] often sampling via commercial website rankings.

Turning to the second wave, there are no GPC adoption studies because only a draft specification has been released so far. The TCF ecosystem has been probed from a range of academic disciplines. Legal methods are relevant to the semantic content of the signal. For example, the purposes for collecting personal data standardized in the TCF may not be specific enough [64].

User interface research is important because the TCF does not standardize how the consent decision is presented to users, which is known to be influential [10, 65–67]. At least two studies have found that consent dialogues used to collect consent under the TCF contain design choices that nudge users towards providing consent [7, 9].

Web scraping studies have focused on implementation problems with TCF [64] or the ecosystem of consent management providers (CMP) [13]. These studies provide measurements of TCF in passing. For example, both studies measure TCF vendor registrations and their claimed purposes for processing data for TCF 1.x [13, p. 9] and both TCF 1.x and TCF 2.0 [64]. The latter study measures aggregate TCF 2.0 adoption,

whereas we measure and visualise at the vendor level. Matte et al. [8] show how TCF 1.x adoption varies by top-level domain (TLD) and identify the most popular CMPs across the top 1k sites in five EU country code TLDs. Hils et al. [13] use longitudinal measurements to show the market growth of six CMPs, highlighting how fast the ecosystem changes.

3.3 Standards Adoption

We build on a body of work emphasizing the role of institutions in technical standards adoption. For example, many vendors initially saw the TCP/IP protocols as a nuisance [68]. Leiner et al. [68] describe how a series of “conferences, tutorials, design meetings and workshops” were organized to educate a generation of vendors and engineers. The rest is history.

The community was slow to turn to adoption questions like “What Makes for a Successful Protocol?”, which was posed by RFC 5218 in 2008. Noting the qualitative nature of the resulting research, Nikkah et al. [69] provide an illuminating statistical analysis of the association between technical features of 250 RFCs and adoption success. Analysing unchanging technical features cannot explain why it took two decades before IPv6 was widely adopted [70, 71]. Economic considerations like the scarcity of IPv4 addresses and the supply of compatible hardware can help explain *when* standards are adopted [72].

Thus, standards should be considered in the context of wider ecosystems governed by economic incentives. For example, HTTPS adoption relies on X.509 certificate infrastructure that was “in a sorry state” in 2011 with many websites relying on shared or invalid certificates [73]. The situation was worse in the long tail likely because certificates are costly [74]. Felt et al. [75] report on significant improvements in 2017 and attribute improvements in the long tail to institutions like Let’s Encrypt and publishing platforms—we show how similar economic considerations explain why TCF was adopted.

3.4 Contribution

Our main empirical contribution involves measuring the adoption of privacy preference signals among websites as of February 2021. Following the demise of P3P and DNT, the TCF has become dominant and the Global Privacy Control is still in its infancy. We explore variables explaining which websites adopt TCF,

and also longitudinally measure migration to a new version (TCF 2.0).

This work differs from existing work by focusing exclusively on the adoption of privacy preference signals. We largely ignore the actors [13, 64] and interfaces [7, 9, 10] harvesting such signals and instead focus on which factors (e.g. website type, popularity, and partners) are associated with TCF adoption. Further, we are the first to systematize strands of research ranging from works in the late 1990s to post-GDPR studies. Finally, we provide the first results about migrating between versions of such signals using our the longitudinal methodology introduced in [13]. Our previous work focuses on detecting specific CMPs, some of whom collect non-TCF signals exclusively or only collect TCF signals for a subset of customers.

4 Methods

We adopt a mixed approach⁶ conducting both longitudinal high-frequency measurements to determine historic adoption of TCF and migration between versions, as well as a large-scale snapshot measurement to examine site-specific factors that may influence adoption. Section 4.1 describes our snapshot measurement of the Tranco 100k toplist. Section 4.2 explains how we use the Netograph platform to conduct longitudinal high-frequency measurements.

4.1 Snapshot Measurements

To measure the prevalence of TCF and its different versions on the web, we crawled the top 100k entries from the Tranco toplist, which aggregates the ranks from the lists provided by Alexa, Cisco Umbrella, Majestic, and Quantcast [76]. Our automated browser crawls were performed in February 2021 using a Tranco toplist from January 2020⁷. We used this older toplist dated shortly before publishers transitioned to TCF 2.x in order to avoid survivorship bias in our observations. Picking a later toplist would over-sample websites created post-2020 who are certain to adopt TCF 2.0 and de facto avoid a migration decision. Our toplist and a current

Table 2. Data sources for figures.

Figure	Approach	Data Source	<i>N</i>	CMP
3, 5	Snapshot (Feb. '21)	Tranco Toplist	100k	all
4	Snapshot (Feb. '21)	Tranco Toplist	10k	all
6, 9	Longitudinal	Netograph	7.2M	QC/OT
7	Longitudinal	Netograph	5.7M	QC
8	Longitudinal	Netograph	1.4M	OT
10–11	Diff. of vendor list	IAB	293	–

Tranco toplist (Tranco id KGNW from Feb. 19th 2021) overlap by 76.5%.

We first converted the Tranco list of domains to a list of URLs that can be crawled. For each *domain*, we attempted to establish a TLS and a TCP connection with `www.domain` and `domain` on port 443 and 80, respectively. This was repeated three times over a week to catch temporary service disruptions. We then picked a configuration that was reachable at least once, preferring TLS over TCP and secondly `www.domain` over `domain` to construct our crawl URL. An error in the TLS certificate verification was treated as unreachable. We used `http://domain` as a fallback if no connections were successful.

Our crawling infrastructure was set up in a European university network. Websites were opened using Google Chrome on Linux with its current default user agent,⁶ a desktop resolution of 1024×800, and en-US as the preferred browser language. All other settings were set to their defaults: third party cookies are allowed, the DNT and GPC HTTP headers are not set. The low desktop resolution and all other settings were chosen to match that of our longitudinal measurements described below. Crawls are automated using custom browser instrumentation based on the Chrome DevTools Protocol. Unsuccessful crawls were retried twice within a week.

For every capture, we collected the following data points using custom browser instrumentation. First, HTTP headers are stored for all requests and responses. Second, connection-related metadata such as IP addresses and TLS certificate chains are logged. Third, for every domain in a capture, its relation to the main page, all cookies, IndexedDB, LocalStorage, SessionStorage and WebSQL records are saved. Fourth, we store the browser’s DOM tree and record a full-page screenshot (including scrolling).

⁶ Supplementary Material:

<https://github.com/mhils/pets2021-privacy-preference-signals>

⁷ Available at <https://tranco-list.eu/list/K8JW>

4.1.1 TCF Adoption

We automatically detect whether crawled websites implement the TCF. To do this, we wait for the website’s DOMContentLoaded event to fire, then wait another ten seconds, and then inject JavaScript code into the execution context of the root document. This approach for CMP detection was already validated by Matte et al. [8] with more aggressive timeouts. As each CMP must implement a `__cmp()` function for TCF 1.x and `__tcfapi()` function for TCF 2.x, we check for the presence of these functions to determine if TCF is being used. We additionally checked for other signs of TCF (such as the presence of `__tcfapiLocator` or `__cmpLocator`), but this search did not turn up any new results. For every TCF API we find, we issue a ping command to learn more about the implementation. In the case of TCF 2.x, the `PingReturn` object (as specified by the TCF) is expected to contain the CMP’s identifier (as assigned by the IAB) as well as the CMP/GVL/TCF versions in use. We also considered that a CMP may masquerade as a different CMP here. We correlated the reported CMP ids with contacted domains and did not find any evidence of misrepresentation.

The adoption of TCF is naturally higher on some types of websites, such as those who typically display paid advertisements. To quantify this, we divided the Tranco 10k toplist into categories with the help of Symantec Rulespace [77], a categorization database already used in related work by Sanchez-Rola et al. [78]. We limit our analysis to the Tranco 10k as a non-negligible share of websites (11.7%) in the top 100k is not categorized, compared to only 2.4% for the top 10k websites. We note that recent work has shown that most categorization services are not fit for detecting specialized content or content-blocking [79], but this does not significantly affect our coarse classification of popular domains.

To determine the number of third parties present on each website, we normalized all requested URLs to their effective second-level domain using Mozilla’s Public Suffix List [80]. This list contains all suffixes under which internet users can directly register names, including non-standard “TLDs” such as `blogspot.com`. We note that this approach does not account for recent obfuscation techniques such as CNAME cloaking [81].

We also examined the fraction of websites that appear to be collecting data versus those showing a cookie prompt. To determine a lower bound, we took all third-party domains that were included on at least 1,000 websites in the Tranco 100k (158 domains) and manually

removed shared resources such as content delivery networks which may not constitute tracking (12 domains). We then determined for each website if any of the remaining 146 third parties were embedded. For example, we exclude `s3.amazonaws.com` as this domain is commonly used to serve static assets and not for tracking. In contrast, almost all remaining domains clearly belong to ad companies. We include both lists in the supplementary material.⁶

Finally, we estimated the prevalence of non-TCF cookie notices or consent prompts in our snapshot measurements using a simple back-of-the-envelope heuristic. For every capture, we scan the stored copy of the browser’s final DOM tree for the occurrence of the phrase “cookie”. The resulting estimates only indicate orders of magnitude, which is acceptable given they are not core to any of our results. Rather they are intended to provide context, such as showing government websites are significantly less likely to present a cookie notice than our other categorizations (see Figure 4). In a manual inspection of 50 randomly picked domains with and 50 domains without “cookie” in their DOM tree, we found five domains that had a “Cookie Notice” link in their footer (but no dialog) and no false negatives (which yields a 5% error rate overall). Again, this part of our analysis is not as rigorous as our other measurements and is only intended to provide context in Figure 4.

4.2 Longitudinal Measurements

To measure the adoption and transition between TCF versions longitudinally, we analyze automated browser crawls recorded by the Netograph web measurement platform.⁸ Netograph continuously ingests a live feed of social media posts, extracts all URLs, and visits them from crawlers located in EU and US data centers. For brevity, we refer to [13] for a discussion of the validity and reliability of this measurement method. Most importantly, HTTP message contents are not retained due to storage constraints, but a large amount of metadata is stored, such as the HTTP headers of every request.

Relying on metadata in our longitudinal data means we have to measure TCF adoption using CMP-specific indicators. Instead of building quick and dirty heuristics for over 90 CMPs, we focus our efforts on creating a set of reliable indicators for two of the leading providers in the consent management market, Quant-

⁸ <https://netograph.io/>

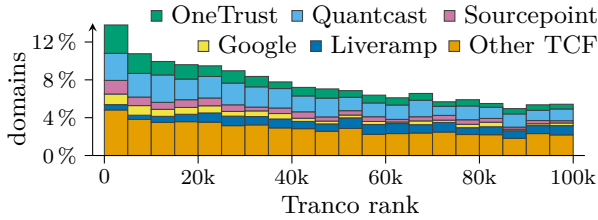


Fig. 3. Share of websites in the Tranco 100k that use a CMP. OneTrust and Quantcast are the most popular providers, followed by Sourcepoint, Google, and Liveramp.

cast and OneTrust, which are embedded on 9.7% of websites in the Tranco 10k (Feb. 2021). We manually analyzed their respective dialog implementations and identified distinct HTTP requests that indicate the use of specific TCF versions⁶. For Quantcast, we detected the use of TCF for all implementations dating back to May 2018. For OneTrust, we identified the use of TCF 1.x or TCF 2.0 in their Cookie Consent SDK launched at the end of 2019 (`otSDKStub.js`).

From Netograph’s 177 million captures in the social media dataset, we obtained all 5.7 million captures that include a Quantcast consent dialog and all 1.4 million captures that include a OneTrust consent dialog. We grouped captures by their effective second-level domain to not overcount repeated measurements with varying subdomains. Due to Netograph’s sampling strategy, less popular domains may not be observed for a several days. We account for this by explicitly marking the period between the last TCF 1.x and the first TCF 2.0 measurement as an (unobserved) transition phase.

4.2.1 Measuring Vendor Adoption

To track the adoption of TCF 2.0 by AdTech vendors, we downloaded all previously published lists of vendors registered as participating in the TCF from the IAB and verified their accuracy using the Internet Wayback Machine. These lists include each vendor’s declared purposes for processing personal data. As of Feb. 2021, there are 215 revisions of this list for TCF 1.x and 78 revisions for TCF 2.0. We then inspected these previous versions for longitudinal changes and measured every instance when an AdTech vendor joins, leaves, or switches to TCF 2.0. While TCF 2.0 is not backwards compatible from a publisher’s point of view, a vendor that has declared support for TCF 2.0 may still accept TCF 1.x consent strings from publishers.

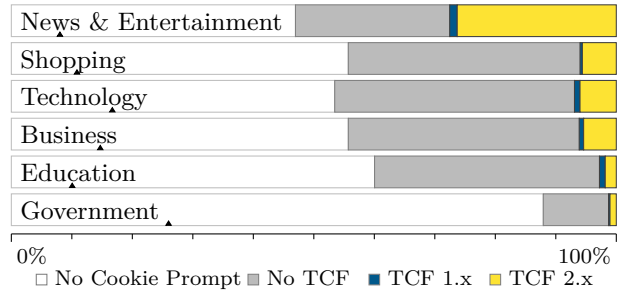


Fig. 4. Share of websites in the Tranco 10k with a (TCF) cookie prompt. For reference, \blacktriangle marks the share of websites which do not embed popular third parties.

5 Results

Section 5.1 focuses on the relationship between website characteristics and TCF adoption mainly using snapshot measurements. Section 5.2 explores how vendors and websites migrated to TCF 2.0 using our longitudinal approach. Table 2 maps each figure to the approach, data source, and covered CMPs. We provide the underlying data in the supplementary material.⁶

5.1 TCF Adoption

We first explore how TCF adoption varies by the popularity and category of website. Figure 3 shows that TCF is more prevalent among popular websites (e.g the Tranco 5k) and that adoption is relatively consistent through the Tranco 100k. Websites embedding OneTrust comprise a greater fraction of TCF implementations for more popular sites (Tranco 20k), whereas Quantcast embeds are more evenly distributed. Quantcast’s free self-service solution may be better suited to less popular sites than OneTrust’s, which requires an interaction with a sales associate. By offering a free and usable solution, Quantcast is playing a similar role to Let’s Encrypt with HTTPS adoption [75].

Figure 4 shows that TCF adoption in the Tranco top 10k is highest among websites classified as News & Entertainment and is lowest among Government websites. The grey bars provide a relatively coarse indication (see the previous section) of what percentage of each category displays a cookie prompts. Few Government websites display prompts, which helps to explain the low TCF adoption. Almost half of all cookie prompts on News & Entertainment sites implement TCF, whereas this fraction is less than 15% for each of the other five even though the first five categories have a similar frac-

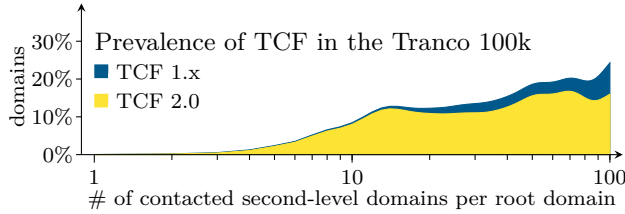


Fig. 5. Adoption of the TCF increases significantly for websites that embed a large number of third parties.

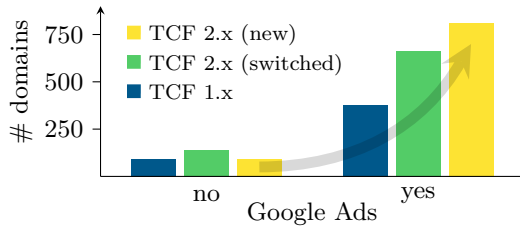


Fig. 6. Google did not participate in TCF 1.x and only joined TCF 2.0. Their partners’ websites were far more likely to adopt TCF 2.x but not TCF 1.x.

tion of websites showing cookie prompts. This motivates exploring alternative explanations.

We explored whether web relationships can help explain varying adoption rates. Figure 5 shows that TCF adoption increases with the number of embedded third parties. This result could be caused by third parties influencing partner websites to adopt TCF, but it could also be mere correlation. Websites with business models based on personal data may be *both* more likely to embed many third parties and also more likely to adopt the TCF.

Causality could be probed via a natural experiment in which websites were randomly assigned a partner that exerts influence. It can be argued the decision of Google to join TCF 2.0 but not TCF 1.x provides such an opportunity. By comparing the relative adoption of TCF 1.x and TCF 2.0 among websites which embed Google with those who do not, we can isolate the effect on TCF adoption of partnering with Google. If partnering with Google influences websites’ decisions, we would expect a higher fraction of such websites to adopt TCF 2.0 but not TCF 1.x as compared to the same fraction among non-partners. Indeed, Figure 6 shows that for websites supporting TCF 2.x and not using Google Ads, 60% had already joined TCF 1.x, whereas this applies to only 45% of the websites using Google Ads. We cannot tell whether the influence is active (e.g. vendor X only contracts with TCF websites) or passive (e.g. website Y finds it easier to adopt the same standard as their partners).

To shed more light on these relationships, we run logistic regressions with TCF 2.0 adoption as the dependent variable. For each website, we have the following explanatory variables: a binary dummy for the presence of Google ads β_1 (from Figure 6), log of the number of embedded third parties β_2 ⁹ (from Figure 5), and the website category (from Figure 4). We include a full regression table in the Appendix (Table A.1).

As we would expect from the figures, the first regression shows β_1 and β_2 have a positive relationship with adoption:

$$y \approx -4.6^{***} + 0.15^{***}\beta_1 + 0.77^{***}\beta_2 \quad (1)$$

and both effects are statistically significant at the $p = 0.01$ level. This means each variable adds additional explanatory power.

Model 2 adds a fixed effect for each website category and this boosts the Pseudo- R^2 from 0.08 to 0.13 relative to Model 1. The coefficient for News & Entertainment is positive and highly significant. The high adoption rate among such websites exceeds what could be explained by β_1 and β_2 alone.

Finally, Model 3 explores the interaction effect between β_1 and β_2 . The sign of $\beta_1 * \beta_2$ means that the relationships are sub-additive—the increased likelihood of adoption from increasing both variables is less than the sum of increasing each variable independently. Although these regressions have shown that website category and web relationships help explain TCF 2.0 adoption rates, the Pseudo- R^2 shows a lot of the variance remains unexplained. This could be down to our relatively crude statistical design aiming to directly link variables to organisation-level outcomes. A recent systematization of knowledge [41] highlights similar difficulties explaining cybersecurity outcomes via manifest variables and suggests latent variables inferred via reflexive indicators represent a better way forward.

5.2 TCF 2.0 Migration

The release of TCF 2.0 provides an opportunity to observe how actively both vendors and websites adopt these signals.

⁹ We count the first party domain so that $\beta_2 \geq 0$.

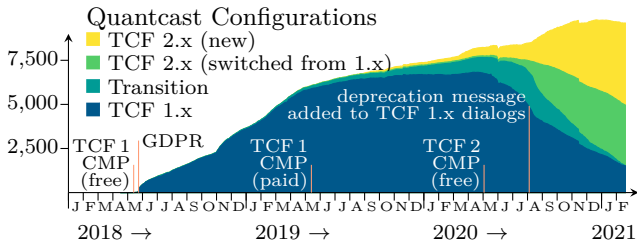


Fig. 7. TCF Adoption by Quantcast customers. Note that the y-axis differs from OneTrust; Quantcast started with a significantly larger number of TCF 1.x customers.

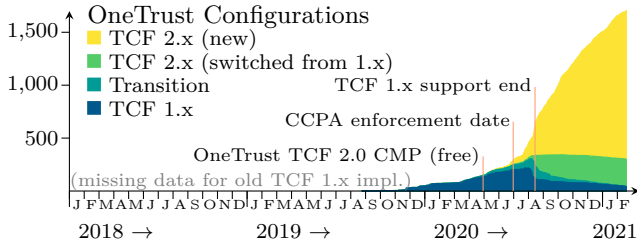


Fig. 8. TCF Adoption by OneTrust customers. Most TCF 1.x customers switched to TCF 2.x around August 2020. Since July 2020, OneTrust gained a large number of new customers which directly started using TCF 2.x. *Transition* marks the unobserved interval during which a switch from TCF 1.x to 2.x occurred.

5.2.1 Websites

Quantcast have the most customers embedding TCF, claim to be a driving force behind its development, and launched a new free TCF 2.0 product in May 2020. Yet Figure 7 shows how a large share of their customers had not adopted the new version when TCF 1.x support by the IAB ended on August 15th. Approaching the IAB’s deadline, Quantcast went as far as embedding a prominent deprecation notice visible to all website visitors into its TCF 1.x consent dialogs (see Figure A.1). Quantcast lost customers while enforcing the switch over, which can be seen in the fall (6%) in old customers who had implemented TCF 1.x from the start of August to end of September. Quantcast’s total customers continue to grow due to new customers who directly adopt TCF 2.0 (the yellow fraction), but the fall in old customers can be seen in the decreasing total of the green and blue lines in Figure 7.

In contrast, OneTrust lost very few customers in transition, which can be seen in the bright green area in Figure 8. OneTrust acquired many new customers from June 2020 and the majority of these immediately adopted TCF 2.0. As a result, OneTrust had a higher fraction of customer implementing TCF 2.0 than Quantcast by the end of September 2020 even though Quant-

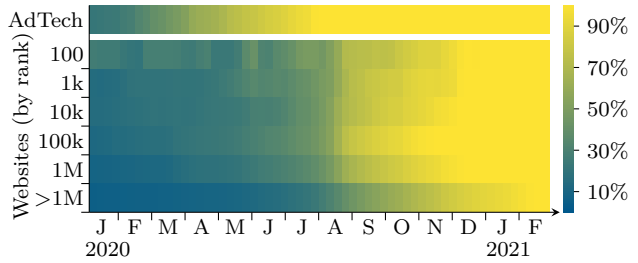


Fig. 9. Share of websites in each segment of the Tranco toplist that use the TCF and have upgraded to version 2.x.

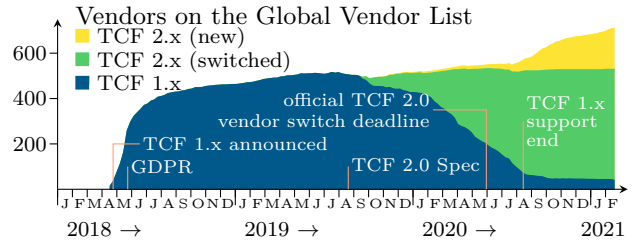


Fig. 10. TCF Adoption by ad-tech vendors.

cast pursued a more assertive transition strategy. However, Quantcast remain comfortably ahead of OneTrust in terms of number of websites embedding TCF (although OneTrust also implements a significant number of non-TCF dialogs [13]).

Returning to the role of top list position, Figure 9 shows that websites in the Tranco top 100 began experimenting with TCF 2.0 migration in the first half of 2020. The experimentation can be seen in how migration went down at various points. The majority had permanently transitioned by July 2020. This suggests the CMP’s announcement about ending support for TCF 1.x were sufficient to lead to migration for popular websites. However, the less popular websites were far less responsive.

5.2.2 Vendors

The majority of early adopters were vendors rather than websites. By the start of 2020, more vendors (84) had switched to TCF 2.0 than there were websites (48) embedding either version of TCF using OneTrust’s Consent SDK. Figure 10 shows vendors appear to follow an S-growth pattern with slow uptake, a relatively small window in which the majority adopt, and a stubborn tail. The number of vendors implementing each version of TCF was relatively consistent through to September 2020, which suggests the upgraded TCF was not a major draw for vendors unlike for websites embedding Google Ads (see Figure 6). The growth rate increased

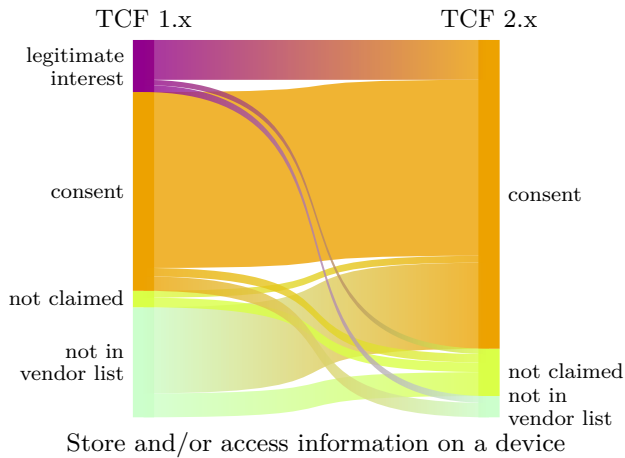


Fig. 11. Removing the option to claim legitimate interest for purpose 1 of the TCF (see Section 2) led more vendors to collect consent for accessing information such as advertising identifiers under TCF 2.x. New vendors that did not adopt TCF 1.x (*not in vendor list*) mostly seek consent as well.

from September 2020 for reasons we do not know, but this is much smaller than the post-GDPR growth.

Comparing time to adoption and migration between vendors and websites speaks to the question of which constituency is driving TCF adoption. Figure 10 shows most vendors had already adopted TCF 1.x by the time GDPR came into effect, whereas OneTrust had no TCF product and only a fraction of Quantcast’s 2020 customers were implementing TCF. The same pattern holds for TCF 2.0 migration. This is consistent with vendors providing an incentive for partner websites towards adoption. While we cannot claim causality, this evidence at least makes it unlikely that websites pushed vendors towards adoption.

5.2.3 Implications

Thus far we have focused on adoption and migration without considering the details or privacy implications of the switch. We illustrate the need for future work by measuring the effect of migrating to TCF 2.0 on the legal basis by which vendors claimed the right to process personal data. We recount some of the background from Section 2. Both versions of TCF define purposes for processing personal data. For each purpose, vendors

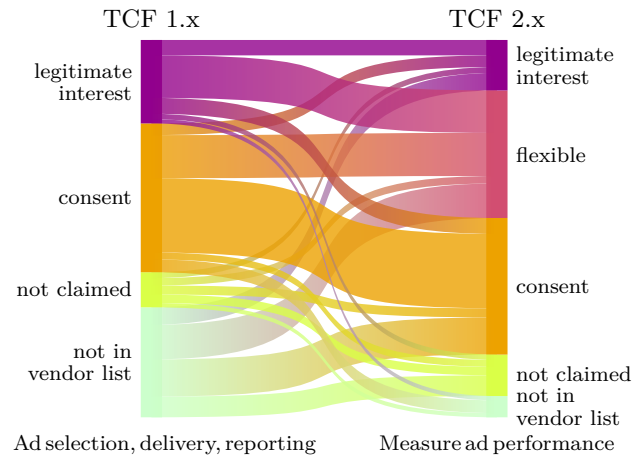


Fig. 12. In migrating from TCF 1.x to TCF 2.x, a large portion of vendors now can claim to be flexible regarding the legal basis; i.e. they will perform the processing based on consent or a legitimate interest.

implementing TCF 1.x can declare either; they do not use personal data for that purpose, need to first obtain consent before doing so, or claim they have a legitimate interest in doing so (which users cannot dispute).

The IAB removed the option to claim a legitimate interest in storing and/or accessing information on a device under TCF 2.x. Figure 11 shows how this shifted the majority of vendors who were previously claiming legitimate interest towards asking for consent. This highlights how standards setters can influence how privacy preferences are communicated at scale by removing the legally questionable options.

Updated standards can also add complexity that makes analyzing impacts difficult to evaluate. For example, the purpose “ad selection, delivery and reporting” was renamed and split into multiple purposes in TCF 2.x. Additionally, vendors had the additional option to declare that they are flexible regarding the legal basis; they can perform the processing based on consent or a legitimate interest. Figure 12 shows how this led to a decrease in both the number claiming legitimate interest and also the number collecting consent, which means it’s unclear whether users lost or gained control under the new standard. These results show just one way in which the design of standards impacts user privacy.

6 Discussion

This section discusses the past, present (as established in the previous section) and future of privacy preference signals.

6.1 Past

Mark Twain’s quip that “history doesn’t repeat itself, but it often rhymes” is also true of privacy preference signals, and identifying these rhymes helps to reason about the present and future. For example, Table 1 shows that signals proposed by AdTech (NAI and TCF) collect user preferences via a web page, whereas the signals proposed by privacy advocates are collected by a browser. As a result, browsers immediately support AdTech signals and could only stop them by actively preventing web content rendering, meanwhile AdTech vendors must actively make the decision to support P3P, DNT and GPC. Consequently, standards developed by AdTech industry bodies have been adopted by browsers by default, whereas AdTech vendors can delay adoption and thus undermine the standard.

Privacy preference signals also vary in terms of the signal’s scope, permanence, and how decision volume scales with web usage. Table 1 highlights how privacy preferences are collected in a single interaction under P3P and DNT/GPC and the browser assumes that this decision applies to the entire Web. Consequently, the user makes a single decision that has long-term signaling implications. In contrast, the NAI’s opt-out cookies only apply to specific forms of tracking [5] and only last until the user loses the cookie or the vendor sets a new one.

Scope and permanence are even narrower under the TCF, which contains asymmetries based on the preferences expressed. The decision not to provide consent¹⁰ only applies to a specific website and only last until the website re-requests consent, whereas positive consent signals may apply to multiple websites [8, 14] and re-requests are less frequent. Table 1 shows history repeating itself in that privacy advocates support a signal that imposes a low decision load on users (P3P, DNT and GPC), whereas AdTech vendors support impermanent signals with a narrow application that force a decision burden on users (NAI and TCF).

Turning to the forum in which signals were designed, we have seen a movement away from development via consensus-based working groups committed to open standards. Initially all parties met in working groups coordinated by the W3C but the clashing political objectives led to splintering. For example, the Digital Advertising Alliance withdrew from the DNT working group in 2012 citing the lack of progress [82].

The second wave of privacy preference signals were developed outside of open, consensus-based groups. TCF was developed via a working group listing 139 participating organizations [83] for which the Interactive Advertising Bureau controlled membership. The resulting TCF signal is closed in that both websites and vendors need the IAB’s permission to implement it, although this authority is delegated to consent management providers. GPC is developed more openly, but lists only 17 supporting organizations with no formal forum to coordinate development. For comparison, the P3P 1.0 specification lists participants from 56 organizations, the DNT working group contained 110 members [82], and the NAI for a long time only included “a fraction of the industry” [5] and now counts 91 members.

In retreating to less consensus-based processes, the Global Privacy Control and the Interactive Advertising Bureau follow (in more than just initials) the governance model of the Internet Advisory Board, which was created in 1984 to incorporate stakeholders beyond Vint Cerf’s “kitchen cabinet” [84, p. 51]:

“The IAB cannot be characterized as a democracy, since nobody voted and the Board only let in the people they wanted . . . Democracy, with its competing factions and its political compromises, was not an appropriate political model for the IAB or the Internet.”

The same could be true of privacy standards given over 10 years was spent drafting P3P and DNT at the W3C. It should be noted that the Internet’s IAB later moved towards more open governance by creating and transferring power to the IETF [84]. It seems unlikely AdTech’s IAB will voluntarily follow suit, which raises the question of regulatory involvement.

The history of privacy preference signals is intertwined with regulation. Do Not Track began as a letter to congress and was re-invigorated by the FTC chairman going off script to mention it years later [85]. The NAI’s opt-out cookies resulted from an agreement with the FTC to self-regulate [5]. The IAB created the TCF in response to the GDPR, and GPC quotes “Do Not Sell” directly from the CCPA. However, none of these signals

¹⁰ Notably, the TCF framework does not even mention the possibility a user can “revoke” a decision [31].

are mandated by law, which means they could become de-facto standards by achieving widespread adoption.

A final lesson from history is that for all the willingness of browser developers to attend working groups, they are reluctant to support privacy preference signals if doing so risks impacting user experience. For example, Microsoft set allow-all cookies as the default for sites who misconfigure P3P presumably because blocking cookies may have affected those websites. This decision on defaults was widely exploited; a misconfiguration described on a Microsoft support page was detected down to the exact typo in 2 756 sites [1]. Similarly, DNT was adopted without sufficient enforcement from browsers, which does little to improve user privacy beyond shifting the blame to AdTech vendors for not respecting the signal.

More encouragingly, history also shows privacy advocates can subvert systems with relatively low-effort browser add-ons. For example, advertising networks expected every user to visit their individual websites to set opt-out cookies [5]. In reality, the TACO browser extension allowed one individual to maintain and share an updated list of cookies with thousands of users [85]. Similarly, the Privacy Bird allegedly helped boost P3P adoption by directly making the user aware of websites' adoption decisions. These two examples point to the importance of designing privacy enhancing technologies that allow users to send low-effort privacy preference signals. This becomes especially urgent given the state of the present, to which we now turn.

6.2 Present

Having surveyed a history in which P3P and DNT were eventually deprecated and NAI membership remains at less than one hundred vendors, our measurements provide an updated picture as of February 2021. TCF is the dominant signal as the GPC was released as an unofficial draft in October 2020 and only six websites in the Tranco top 100k now implement it. Given signals must be adopted by both sender and recipient, we now discuss adoption among each stakeholder.

Websites are arguably the most important stakeholder for the success of TCF since only websites can collect consent signals [14]. We discovered 7,582 TCF implementations in the top 100k. A crude comparison can be drawn with a 2010 sample detecting 19.8k P3P implementations [1]. Turning to estimates that reference a toplist, TCF is more prevalent among both the top 5k (13%) and top 100k (7%) than historic P3P measure-

ments (8% [21] and 2% [22] respectively). Such comparisons are limited by changes in the Web and also research methods; P3P adoption studies relied on commercial rankings, whereas we used a top list designed to be stable over time for research purposes. This should make our measurement more comparable to future work.

Turning to adoption among AdTech vendors, vendors were early adopters of TCF and also the first to migrate to TCF 2.0 (see Figure 9). By October 2020, more than 600 vendors had adopted TCF. For comparison, just 75 vendors were offering opt-out cookies in June 2010 of which only 11 were also implementing P3P [1]. Although AdTech vendors drafted the TCF specification, adoption was not inevitable given the NAI had no more than 6 full members from 2001–2007 [5]. Thus, TCF is the first privacy preference signal to achieve widespread adoption among AdTech vendors.

Our results also speak to why websites are adopting TCF. Numerous pieces of evidence suggest vendors incentivize partner websites to adopt TCF (see Figure 3, Figure 5 and especially Figure 6). An interesting comparison can be made with P3P. Websites embedding more third-party domains are more likely to adopt TCF but less likely to adopt P3P [62, p. 292]. This supports the common sense intuition that TCF was designed to perpetuate privacy eroding business models.

More generally, we provide evidence in support of the general finding that private firms deploy economic resources to ensure the adoption of standards [86]. Figure 3 shows how Quantcast's free consent management solution supported TCF adoption, particularly among less popular sites. The role of institutional support is crucial even to open standards, such as the organization of TCP/IP education events [68] and subsidization of free certificates via Let's Encrypt to support HTTPS adoption [75]. In terms of migrating to updated standards, we show how Quantcast boosted TCF 2.0 adoption by adding prominent deprecation messages into consent dialogs. Thus, Figure 7 suggests that IAB policy (TCF 1.x consent strings becoming invalid) led to Quantcast losing customers.

Finally, we can quantify the relative decision volume of users relative to vendors. Quantcast boast of processing 25 billion consent signals [87], whereas we observed just 2,103 changes in vendor purposes since 2018. This means users have made at least 11 million times more decisions than vendors since TCF was launched. At 3.2s per decision [13], this means users have spent at least 2,500 years since 2018 expressing their privacy preferences through Quantcast dialogs alone.

6.3 Future

Given this startling time investment in sending TCF signals, it is worth considering what the future holds for pro-privacy signals. Releasing the GPC specification in an unofficial draft [30] over two years after GDPR came into effect and ten months after CCPA provided TCF with a first-mover advantage. However, we have few concerns that privacy aware users will adopt the GPC in the future. Pro-privacy browsers like Firefox supported the design, additionally the Brave browser¹¹ and add-ons like Privacy Badger¹² already turn the GPC signal on by default.

We are less optimistic that the intended recipients, namely AdTech vendors, will adopt the GPC signal. Much like with DNT [4], AdTech vendors are likely to claim that on-as-default makes the signal meaningless. However, privacy advocates can now rely on privacy laws like the CCPA, which was not available when DNT was first adopted by browsers.

Fighting legal cases to establish a favorable precedent is a likely strategy. One of the GPC's participating organizations, Brave Browser, has already lodged complaints under the GDPR against rival browsers [88], national regulators [89], and even the IAB Europe's website [90]. We anticipate similar actions under the CCPA, especially given California's attorney general tweeted about the GPC in January 2021¹³. Multiple publishers adopting the same standard and out-sourcing implementation to dominant CMPs creates the potential for auditing at scale [p. 10][13], as evidenced by an NGO's threat of automated complaints against publishers¹⁴.

Regulatory interventions may begin to undermine the adoption of TCF. For example, the Danish regulator ruled that the Danish Meteorological Institute could not claim a legitimate interest in collecting personal data [91]. Possibly preempting such a ruling, the option to declare a legitimate interest in storing and/or accessing information on a device was removed in TCF 2.0 (see Figure 11). The case also ruled that opt-out must be as easy as opt-in. Many websites collecting TCF signals do not follow this ruling [9, 13]. The leading provider of TCF dialogs distances itself from ambiguity in privacy law [92] by making the design choice a configuration that

websites select, with one CMP warning “with great customizability comes great responsibility” [13]. This indicates that AdTech vendors perceive liability risk related to TCF.

This discussion raises the question of what happens when two signals co-exist. Whereas standards usually have a definitive winner, such as DVD over DIVX or VHS over Betamax [93], GPC and TCF signals can be sent simultaneously because they are defined on different network layers (see Table 1). Encouragingly, one could imagine a future in which browsers exploit control over what is rendered to the user to block dialogs from loading, whereas AdTech cannot stop browsers from sending GPC headers as part of HTTP requests. Signals co-existing is more troublesome when it comes to interpretation. A TCF opt-in signal could be sent in an HTTP request with GPC opt-out headers. We leave it to legal scholars and future court cases to ponder which signal has priority.

Arguably this back and forth over privacy preference signals has been a distraction for over 20 years. Regardless of the adoption of privacy preference signals, there is little basis to trust that expressed preferences will be respected. In terms of what we can observe: vendors ignoring the DNT signal was public policy [4], P3P was intentionally misconfigured by websites [1], TCF consent signals misreport the user's expressed preferences [8], tracking remains ubiquitous in a post-GDPR world [78] and there is growing evidence firms use dark patterns to manipulate users' expressed preferences [94–96]. More fundamentally, there is no way of auditing whether AdTech vendors respect expressed signals.

7 Conclusion

Privacy preference signals must be adopted by both senders (users) and recipients (AdTech vendors) who have differing requirements. Vendors want to receive positive consent signals in order to comply with privacy laws, and prefer not to receive negative signals that undermine the vendor's business model. This reasoning helps to explain why hundreds of vendors adopted TCF [13, 64], which represents a historical anomaly given vendors reluctance to adopt P3P [1], DNT [28] and NAI opt-out cookies [5]. Our evidence that vendors were early adopters of TCF 2.0 (Figure 9) underlies the AdTech vendors' commitment to receiving these signals.

History reveals two approaches to collecting users' privacy preferences that are represented in the signal, namely via the user agent (as in P3P and DNT) or a webpage (as in NAI opt-out). As with the previous sig-

¹¹ <https://brave.com/global-privacy-control/>

¹² <https://www.eff.org/gpc-privacy-badger>

¹³ <https://digiday.com/media/why-a-tweet-from-californias-ag-about-a-global-privacy-tool-has-companies-scrambling/>

¹⁴ <https://noyb.eu/en/noyb-aims-end-cookie-banner-terror-and-issues-more-500-gdpr-complaints>

nal designed by AdTech [5], TCF collects user preferences via dialogs embedded in a web page but this requires adoption among websites. Our results show website adoption varies from 5% to 12% across sections of the Tranco top 100k (Figure 3) and is most prevalent among News & Entertainment websites (Figure 4). We also show that the presence of Google Ads (Figure 6) and the number of embedded parties (Figure 5) are both associated with greater TCF adoption rates.

Adoption is further supported by AdTech actors like Quantcast lowering the cost of adopting TCF by providing free dialogs marketed as compliant with GDPR (although legality has been called into question [8, 9]). The increase in adoption following May 2018, which can be seen in Figure 7, shows how AdTech capitalised on the passage of the GDPR. This means AdTech firms now not only draft the TCF, but also actively manage and configure it. This market power facilitated the swift transition to TCF 2.0 (see Figure 8 and Figure 7), which is remarkable when contrasted against the time to migrate to HTTPS [75] or IPv6 [72].

Thus, our measurements of the present reveal TCF is now the dominant privacy preference signal. Further, its adoption among *both* senders and recipients is a significant historical development (see Table 1). Adoption among recipients is unsurprising given the working group who designed TCF was controlled by the Interactive Advertising Bureau and contained no privacy advocates. However, websites appear to have sided with their business partners over users. Consequently, users are forced to send signals via time consuming dialogs. Our back-of-the-envelope calculation on p. 262 suggests over two thousand years of user time has been spent on sending TCF consent signals since 2018. All stakeholders should ask to what extent the TCF's fine-grained, site-by-site signal clarifying privacy preferences has materially changed how recipients process personal data? A second question is whether a revised signal would lead to better outcomes, or can the problems only be resolved by the technical constraints of *hard privacy*?

Acknowledgements

We would like to thank Aldo Cortesi for his continuous support and the generous access to the Netograph API and capturing technology. We thank Anelia Kurteva, Jérémie Bernard Glossi, Dennis Jackson, our shepherd Christo Wilson, and the other anonymous reviewers for many constructive comments. The second author is funded by the European Commission's call H2020-MSCA-IF-2019 under grant number 894700.

References

- [1] Pedro Giovanni Leon, Lorrie Faith Cranor, Aleecia M McDonald, and Robert McGuire. Token attempt: The misrepresentation of website privacy policies through the misuse of P3P compact policy tokens. In *ACM Workshop on Privacy in the Electronic Society*, pages 93–104, 2010.
- [2] Electronic Privacy Information Center and Junkbusters. Pretty Poor Privacy: An Assessment of P3P and Internet Privacy. <https://epic.org/reports/prettypoorprivacy.html>, 2000.
- [3] Tracking Protection Working Group. WG closed. <https://github.com/w3c/dnt/commit/5d85d6c>, 2019.
- [4] Interactive Advertising Bureau. "Do Not Track" set to "On" by Default in Internet Explorer 10—IAB Response. <https://www.iab.com/news/do-not-track-set-to-on-by-default-in-internet-explorer-10iab-response/>, 2012.
- [5] Pam Dixon. *The Network Advertising Initiative: Failing at Consumer Protection and at Self-Regulation*. World Privacy Forum, 2007. http://www.worldprivacyforum.org/wp-content/uploads/2007/11/WPF_NAI_report_Nov2_2007fs.pdf.
- [6] Martha K. Landesberg, Toby Milgrom Levin, Caroline G. Curtin, and Ori Lev. *Privacy online: A Report to Congress*. US Federal Trade Commission, 1998.
- [7] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. (Un)informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS '19*, pages 973–990. ACM, 2019.
- [8] Célestin Matte, Nataliia Bielova, and Cristiana Santos. Do Cookie Banners Respect my Choice? Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework. In *IEEE Symposium on Security and Privacy*, pages 791–809. IEEE, 2020.
- [9] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*. ACM, 2020.
- [10] Dominique Machuletz and Rainer Böhme. Multiple Purposes, Multiple Problems: A User Study of Consent Dialogs after GDPR. *Proceedings on Privacy Enhancing Technologies*, (2):481–498, 2020.
- [11] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. "It's a Scavenger Hunt": Usability of Websites' Opt-Out and Data Deletion Choices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*. ACM, 2020.
- [12] Sean O'Connor, Ryan Nurwono, and Eleanor Birrell. (Un)clear and (In)conspicuous: The right to opt-out of sale under CCPA, 2020.
- [13] Maximilian Hils, Daniel W Woods, and Rainer Böhme. Measuring the Emergence of Consent Management on the Web. In *Proceedings of the Internet Measurement Conference 2020, IMC '20*. ACM, 2020.
- [14] Daniel W Woods and Rainer Böhme. The commodification of consent. In *20th Annual Workshop on the Economics of*

- Information Security, WEIS*, 2020.
- [15] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, CCS '14*, pages 674–689. ACM, 2014.
- [16] Steven Englehardt, Dillon Reisman, Christian Eubank, Peter Zimmerman, Jonathan Mayer, Arvind Narayanan, and Edward W. Felten. Cookies That Give You Away: The Surveillance Implications of Web Tracking. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15*, pages 289–299, Republic and Canton of Geneva, CHE, 2015. International World Wide Web Conferences Steering Committee.
- [17] Steven Englehardt and Arvind Narayanan. Online Tracking: A 1-Million-Site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, CCS '16*, pages 1388–1401. ACM, 2016.
- [18] Pierre Laperdrix, Natalia Bieleva, Benoit Baudry, and Gildas Avoine. Browser Fingerprinting: A Survey. *ACM Trans. Web*, 14(2), April 2020.
- [19] T. Bujlow, V. Carela-Español, J. Solé-Pareta, and P. Barlet-Ros. A Survey on Web Tracking: Mechanisms, Implications, and Defenses. *Proceedings of the IEEE*, 105(8):1476–1510, 2017.
- [20] Simon Byers, Lorrie Faith Cranor, and David Kormann. Automated analysis of P3P-enabled web sites. In *Proceedings of the 5th International Conference on Electronic Commerce*, pages 326–338, 2003.
- [21] Patricia Beatty, Ian Reay, Scott Dick, and James Miller. P3P adoption on e-commerce web sites: a survey and analysis. *IEEE Internet Computing*, 11(2):65–71, 2007.
- [22] Ian Reay, Patricia Beatty, Scott Dick, and James Miller. Privacy policies and national culture on the internet. *Information Systems Frontiers*, 15(2):279–292, 2013.
- [23] Riva Richmond. A loophole big enough for a cookie to fit through. *New York Times*, 2010. <https://nyti.ms/2mDvTBQ>.
- [24] Lorrie Faith Cranor, Manjula Arjula, and Praveen Guduru. Use of a P3P user agent by early adopters. In *Proceedings of the 2002 ACM Workshop on Privacy in the Electronic Society*, pages 1–10, 2002.
- [25] World Wide Web Consortium. Tracking Protection Working Group. <https://www.w3.org/2011/tracking-protection/>, 2011.
- [26] Julia Angwin. Microsoft's "Do Not Track" Move Angers Advertising Industry. <https://www.wsj.com/articles/BL-DGB-24506>, 2012.
- [27] Chrome Blog. Longer battery life and easier website permissions. <https://chrome.googleblog.com/2012/11/longer-battery-life-and-easier-website.html>, 2012.
- [28] Future of Privacy Forum. Companies that have implemented Do Not Track. <https://allaboutdnt.com/companies/>, 2020.
- [29] Alex Fowler. Mozilla's new Do Not Track dashboard: Firefox users continue to seek out and enable DNT. <https://blog.mozilla.org/netpolicy/2013/05/03/mozillas-new-do-not-track-dashboard-firefox-users-continue-to-look-out-and-enable-dnt/>, 2013.
- [30] Robin Berjon, Sebastian Zimmeck, Ashkan Soltani, David Harbage, and Peter Synder. Global Privacy Control (GPC) Unofficial Draft 15 October 2020. <https://globalprivacycontrol.github.io/gpc-spec/>, 2020.
- [31] IAB Europe. What is the Transparency and Consent Framework (TCF)? <https://iab europe.eu/transparency-consent-framework/>, 2020.
- [32] J. R. Mayer and J. C. Mitchell. Third-party web tracking: Policy and technology. In *2012 IEEE Symposium on Security and Privacy*, pages 413–427. IEEE, 2012.
- [33] Balachander Krishnamurthy and Craig E Wills. On the leakage of personally identifiable information via online social networks. In *Proceedings of the 2nd ACM workshop on online social networks*, pages 7–12, 2009.
- [34] Gunes Acar, Steven Englehardt, and Arvind Narayanan. No boundaries: data exfiltration by third parties embedded on web pages. *Proceedings on Privacy Enhancing Technologies*, 2020(4):220 – 238, 2020.
- [35] Shehroze Farooqi, Maaz Musa, Zubair Shafiq, and Fareed Zaffar. Canarytrap: Detecting data misuse by third-party apps on online social networks. *Proceedings on Privacy Enhancing Technologies*, 2020(4):336 – 354, 2020.
- [36] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman. "Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale. *Proceedings on Privacy Enhancing Technologies*, 2018(3):63 – 83, 2018.
- [37] Hamza Saleem and Muhammad Naveed. SoK: Anatomy of Data Breaches. *Proceedings on Privacy Enhancing Technologies*, 2020(4):153 – 174, 2020.
- [38] Sébastien Henri, Gines Garcia-Aviles, Pablo Serrano, Albert Banchs, and Patrick Thiran. Protecting against Website Fingerprinting with Multihoming. *Proceedings on Privacy Enhancing Technologies*, 2020(2):89 – 110, 01 Apr. 2020.
- [39] Miti Mazmudar and Ian Goldberg. Mitigator: Privacy policy compliance using trusted hardware. *Proceedings on Privacy Enhancing Technologies*, 2020(3):204 – 221, 2020.
- [40] Martino Trevisan, Stefano Traverso, Eleonora Bassi, and Marco Mellia. 4 Years of EU Cookie Law: Results and Lessons Learned. *Proceedings on Privacy Enhancing Technologies*, 2019(2):126 – 145, 2019.
- [41] Daniel W. Woods and Rainer Böhme. SoK: Quantifying cyber risk. In *IEEE Symposium on Security and Privacy*, May 2021.
- [42] Laura Shipp and Jorge Blasco. How private is your period?: A systematic analysis of menstrual app privacy policies. *Proceedings on Privacy Enhancing Technologies*, 2020(4):491 – 510, 2020.
- [43] Ryan Amos, Gunes Acar, Elena Lucherini, Mihir Kshirsagar, Arvind Narayanan, and Jonathan Mayer. Privacy Policies over Time: Curation and Analysis of a Million-Document Dataset. *arXiv preprint arXiv:2008.09159*, 2020.
- [44] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. In *26th Annual Network and Distributed System Security Symposium, NDSS '19*. The Internet Society, 2019.

- [45] Thomas Linden, Rishabh Khandelwal, Hamza Harkous, and Kassem Fawaz. The Privacy Policy Landscape After the GDPR. *Proceedings on Privacy Enhancing Technologies*, 2020(1):47 – 64, 01 Jan. 2020.
- [46] Judith S Olson, Jonathan Grudin, and Eric Horvitz. A study of preferences for sharing and privacy. In *CHI'05 extended abstracts on Human factors in Computing Systems*, pages 1985–1988, 2005.
- [47] Mark S Ackerman, Lorrie Faith Cranor, and Joseph Reagle. Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM Conference on Electronic commerce*, pages 1–8, 1999.
- [48] Ben Weinschel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L. Mazurek, and Blase Ur. Oh, the Places You've Been! User Reactions to Longitudinal Transparency About Third-Party Web Tracking and Inferencing. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS '19*, pages 149–166. ACM, 2019.
- [49] Sarah Spiekermann, Jens Grossklags, and Bettina Berendt. E-Privacy in 2nd Generation E-Commerce: Privacy Preferences versus Actual Behavior. In *Proceedings of the 3rd ACM Conference on Electronic Commerce, EC '01*, pages 38–47. ACM, 2001.
- [50] Susanne Barth and Menno DT De Jong. The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and informatics*, 34(7):1038–1058, 2017.
- [51] Nina Gerber, Paul Gerber, and Melanie Volkamer. Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & Security*, 77:226–261, 2018.
- [52] Lorrie Faith Cranor. P3P: Making privacy policies more useful. *IEEE Security & Privacy*, 1(6):50–55, 2003.
- [53] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, and Yirong Xu. XPref: a preference language for P3P. *Computer Networks*, 48(5):809 – 827, 2005. Web Security.
- [54] Johnson Iyilade and Julita Vassileva. P2U: a privacy policy specification language for secondary data sharing and usage. In *2014 IEEE Security and Privacy Workshops*, pages 18–22. IEEE, 2014.
- [55] Jean Yang, Kuart Yessenov, and Armando Solar-Lezama. A language for automatically enforcing privacy policies. *ACM SIGPLAN Notices*, 47(1):85–96, 2012.
- [56] Monir Azraoui, Kaoutar Elkhyaoui, Melek Önen, Karin Bernsmed, Anderson Santana De Oliveira, and Jakub Sendor. A-PPL: An Accountability Policy Language. In *Data Privacy Management, Autonomous Spontaneous Security, and Security Assurance*, pages 319–326, Cham, 2015. Springer.
- [57] Lalana Kagal, Chris Hanson, and Daniel Weitzner. Using dependency tracking to provide explanations for policy management. In *2008 IEEE Workshop on Policies for Distributed Systems and Networks*, pages 54–61. IEEE, 2008.
- [58] Ponnurangam Kumaraguru, Lorrie Cranor, Jorge Lobo, and Seraphin Calo. A survey of privacy policy languages. In *Workshop on Usable IT Security Management (USM 07): Proceedings of the 3rd Symposium on Usable Privacy and Security, ACM*, 2007.
- [59] Jun Zhao, Reuben Binns, Max Van Kleek, and Nigel Shadbolt. Privacy languages: Are we there yet to enable user controls? In *Proceedings of the 25th International Conference Companion on World Wide Web, WWW '16 Companion*, pages 799–806. International World Wide Web Conferences Steering Committee, 2016.
- [60] Saffija Kasem-Madani and Michael Meier. Security and privacy policy languages: A survey, categorization and gap identification. *CoRR*, abs/1512.00201, 2015.
- [61] Victor Morel and Raúl Pardo. SoK: Three facets of privacy policies. In *WPES'20: Proceedings of the 19th Workshop on Privacy in the Electronic Society, Virtual Event, USA, November 9, 2020*, pages 41–56. ACM, 2020.
- [62] Lorrie Faith Cranor, Serge Egelman, Steve Sheng, Aleecia M McDonald, and Abdur Chowdhury. P3P deployment on websites. *Electronic Commerce Research and Applications*, 7(3):274–293, 2008.
- [63] Ian Reay, Scott Dick, and James Miller. An analysis of privacy signals on the World Wide Web: Past, present and future. *Inf. Sci.*, 179(8):1102–1115, 2009.
- [64] Célestin Matte, Cristiana Santos, and Nataliia Bielova. Purposes in IAB Europe's TCF: which legal basis and how are they used by advertisers? In *Annual Privacy Forum*, 2020.
- [65] Yee-Lin Lai and Kai-Lung Hui. Internet opt-in and opt-out: Investigating the roles of frames, defaults and privacy concerns. In *Proceedings of the 2006 ACM SIGMIS CPR Conference on Computer Personnel Research, SIGMIS CPR '06*, pages 253–263. ACM, 2006.
- [66] Rainer Böhme and Stefan Köpsell. Trained to accept? A field experiment on consent dialogs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10*, pages 2403–2406. ACM, 2010.
- [67] Idris Adjrid, Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. Sleights of privacy: Framing, disclosures, and the limits of transparency. In *Proceedings of the Ninth Symposium on Usable Privacy and Security, SOUPS '13*. ACM, 2013.
- [68] Barry M Leiner, Vinton G Cerf, David D Clark, Robert E Kahn, Leonard Kleinrock, Daniel C Lynch, Jon Postel, Larry G Roberts, and Stephen Wolff. A brief history of the internet. *ACM SIGCOMM Computer Communication Review*, 39(5):22–31, 2009.
- [69] Mehdi Nikkha, Aman Mangal, Constantine Dovrolis, and Roch Guérin. A statistical exploration of protocol adoption. *IEEE/ACM Transactions on Networking*, 25(5):2858–2871, 2017.
- [70] Jakub Czumak, Mark Allman, Jing Zhang, Scott Iekel-Johnson, Eric Osterweil, and Michael Bailey. Measuring IPv6 adoption. *SIGCOMM Comput. Commun. Rev.*, 44(4):87–98, August 2014.
- [71] Xuequn Wang and Sebastian Zander. Extending the model of internet standards adoption: A cross-country comparison of IPv6 adoption. *Information & Management*, 55(4):450 – 460, 2018.
- [72] M. Nikkha and R. Guérin. Migrating the Internet to IPv6: An Exploration of the When and Why. *IEEE/ACM Transactions on Networking*, 24(4):2291–2304, 2016.
- [73] Ralph Holz, Lothar Braun, Nils Kammenhuber, and Georg Carle. The SSL Landscape: A Thorough Analysis of the x.509 PKI Using Active and Passive Measurements. In *Pro-*

- ceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, IMC '11*, pages 427–444. ACM, 2011.
- [74] Andy Ozment and Stuart E Schechter. Bootstrapping the adoption of internet security protocols. In *5th Annual Workshop on the Economics of Information Security, WEIS*, 2006.
- [75] Adrienne Porter Felt, Richard Barnes, April King, Chris Palmer, Chris Bentzel, and Parisa Tabriz. Measuring HTTPS adoption on the web. In *Proceedings of the USENIX Security Symposium (USENIX Security 17)*, pages 1323–1338, 2017.
- [76] Victor Le Pochat, Tom van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczynski, and Wouter Joosen. Tranco: A research-oriented top sites ranking hardened against manipulation. In *26th Annual Network and Distributed System Security Symposium, NDSS '19*. The Internet Society, 2019.
- [77] Symantec. Symantec RuleSpace: URL categorization database, 2020.
- [78] Iskander Sanchez-Rola, Matteo Dell'Amico, Platon Kotzias, Davide Balzarotti, Leyla Bilge, Pierre-Antoine Vervier, and Igor Santos. Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security, Asia CCS '19*, pages 340–351. ACM, 2019.
- [79] Pelayo Vallina, Victor Le Pochat, Álvaro Feal, Marius Paraschiv, Julien Gamba, Tim Burke, Oliver Hohlfeld, Juan Tapiador, and Narseo Vallina-Rodriguez. Mis-shapes, Mistakes, Misfits: An Analysis of Domain Classification Services. In *Proceedings of the Internet Measurement Conference 2020, IMC '20*. ACM, 2020.
- [80] Mozilla Foundation. Public suffix list. <https://publicsuffix.org/>, 2007–2020.
- [81] Yana Dimova, Gunes Acar, Lukasz Olejnik, Wouter Joosen, and Tom van Goethem. The CNAME of the Game: Large-scale Analysis of DNS-based Tracking Evasion. *Proceedings on Privacy Enhancing Technologies*, 2021.
- [82] Inside Privacy. Digital Advertising Alliance Leaves Do Not Track Group. <https://www.insideprivacy.com/advertising-marketing/digital-advertising-alliance-leaves-do-not-track-group-2/>, 2013.
- [83] IAB Tech Lab. Global Privacy Working Group. <https://iabtechlab.com/working-groups/global-privacy-working-group/>, 2011.
- [84] Andrew L Russell. 'Rough consensus and running code' and the Internet-OSI standards war. *IEEE Annals of the History of Computing*, 28(3):48–61, 2006.
- [85] Christopher Soghoian. The History of the Do Not Track Header. <http://paranoia.dubfire.net/2011/01/history-of-do-not-track-header.html>, 2011.
- [86] Carl Shapiro, Shapiro Carl, Hal R Varian, et al. *Information rules: a strategic guide to the network economy*. Harvard Business Press, 1998.
- [87] Kochava Inc. Quantcast and Kochava Partnership Delivers Combined Web and Mobile App Solution for CCPA. <https://www.businesswire.com/news/home/20200207005054/en/Quantcast-and-Kochava-Partnership-Delivers-Combined-Web-and-Mobile-App-Solution-for-CCPA>, 2018.
- [88] Johnny Ryan. Regulatory complaint concerning massive, web-wide data breach by Google and other “ad tech” companies under Europe's GDPR. <https://brave.com/adtech-data-breach-complaint/>, 2018.
- [89] Natasha Lomas. Brave Accuses European governments of GDPR resourcing failure. <https://techcrunch.com/2020/04/27/brave-accuses-european-governments-of-gdpr-resourcing-failure/>, 2020.
- [90] Johnny Ryan. Formal GDPR complaint against IAB Europe's “cookie wall” and GDPR consent guidance. <https://brave.com/iab-cookie-wall/>, 2019.
- [91] Tue Goldschmieding. New important decision on cookies from the Danish Data Protection Agency. <https://gorrissenfederspiel.com/en/knowledge/news/new-important-decision-on-cookies-from-the-danish-data-protection-agency-2020>.
- [92] Aaron Ceross and Andrew Simpson. Rethinking the Proposition of Privacy Engineering. In *Proceedings of the New Security Paradigms Workshop, NSPW '18*, pages 89–102. ACM, 2018.
- [93] Carl Shapiro and Hal R Varian. The art of standards wars. *California Management Review*, 41(2):8–32, 1999.
- [94] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. Tales from the dark side: Privacy dark strategies and privacy dark patterns. *Proceedings on Privacy Enhancing Technologies*, 2016(4):237–254, 2016.
- [95] Arunesh Mathur, Gunes Acar, Michael J Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. Dark patterns at scale: Findings from a crawl of 11k shopping websites. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–32, 2019.
- [96] Arvind Narayanan, Arunesh Mathur, Marshini Chetty, and Mihir Kshirsagar. Dark Patterns: Past, Present, and Future. *ACM Queue*, 18(2):67–92, 2020.

A Appendix

Table A.1. Regression Coefficients for TCF 2.x Adoption

	<i>Dependent variable:</i>		
	TCF 2.x Adoption		
	(1)	(2)	(3)
Google Ads	0.150*** (0.043)	0.151*** (0.044)	3.502*** (0.140)
log(# contacted SLDs)	0.765*** (0.020)	0.596*** (0.020)	1.787*** (0.053)
Category: Business		-0.436*** (0.055)	-0.430*** (0.055)
Category: Education		-1.384*** (0.098)	-1.385*** (0.098)
Category: Government		-2.479*** (0.303)	-2.506*** (0.304)
Category: News & Entertainment		0.954*** (0.031)	0.994*** (0.031)
Category: Shopping		-0.885*** (0.067)	-0.826*** (0.067)
Category: Technology		-0.487*** (0.055)	-0.469*** (0.055)
Google Ads * log(# contacted SLDs)			-1.517*** (0.058)
Constant	-4.614*** (0.045)	-4.189*** (0.048)	-6.558*** (0.121)
Observations	92,001	82,326	82,326
McFadden's Pseudo- R^2	0.08	0.13	0.14

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.2. Summary Statistics

Variable	N	Min	Mean	Max
TCF 2.x Adoption	92,475	0	0.072	1
Google Ads	92,538	0	0.570	1
log(# contacted SLDs)	92,538	0	2.189	5.004
Category: Business	88,269	0	0.114	1
Category: Education	88,269	0	0.078	1
Category: Government	88,269	0	0.034	1
Category: News & Entertainment	88,269	0	0.210	1
Category: Shopping	88,269	0	0.086	1
Category: Technology	88,269	0	0.132	1

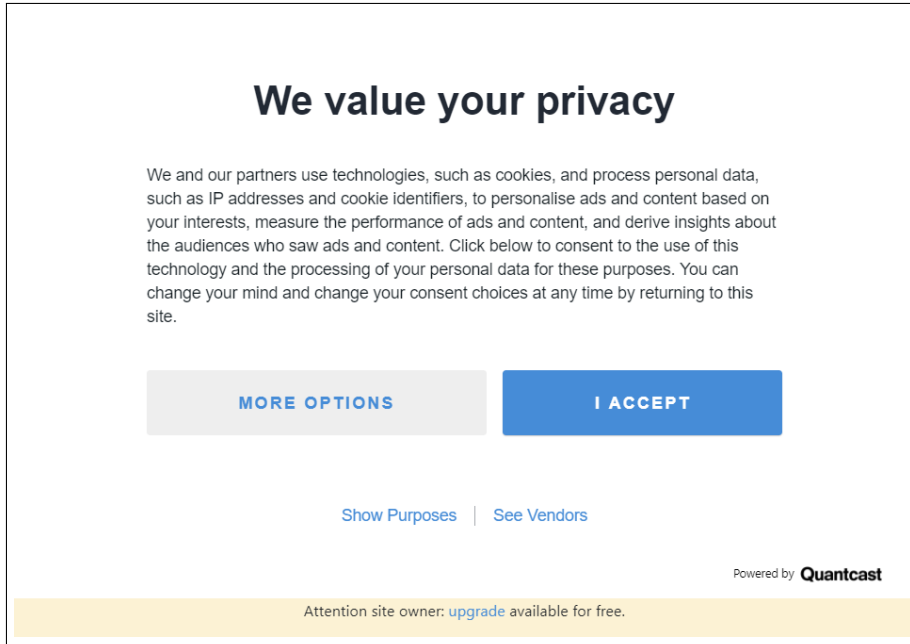


Fig. A.1. Starting August 5th 2020, Quantcast added a prominent deprecation message at the bottom of all its customers' TCF 1.x consent dialogs, prompting them to switch to TCF 2.0.