CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

To pay, or not to pay: whether ’tis tastier To indulge in cookies Or guard thy privacy

A look at cookie paywalls, their tracking claims, and implementations.

Master’s thesis in Computer Science and Engineering

Viktor Fredholm

Simon Hansson

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2023

Master's thesis 2023

To pay, or not to pay: whether 'tis tastier To indulge in cookies Or guard thy privacy

A look at cookie paywalls, their tracking claims, and implementations.

Viktor Fredholm

Simon Hansson

UNIVERSITY OF
GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2023

To pay, or not to pay: whether 'tis tastier To indulge in cookies, or guard thy privacy.
A look at cookie paywalls, their tracking claims, and implementations.
Viktor Fredholm
Simon Hansson

Supervisor: Victor Morel, Department of Computer Science and Engineering
Advisor: Per Grundtman, Omegapoint
Examiner: Alejandro Russo, Department of Computer Science and Engineering

Master's Thesis 2023
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Description of the picture on the cover page (if applicable)

Typeset in LaTeX
Gothenburg, Sweden 2023

To pay, or not to pay: whether 'tis tastier To indulge in cookies, or guard thy privacy.
A look at cookie paywalls, their tracking claims, and implementations.
Viktor Fredholm
Simon Hansson
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

The experience of browsing the web and lazily clicking cookie consent go hand in hand for many. However, it has been shown that the majority of these cookie banners do not adhere to the GDPR and ePrivacy Directive. Based on previous findings that there is a new type of cookie banner, cookie paywalls, where denial of consent is replaced by a fee. Our study presents results showing how well cookie paywalls comply with the GDPR. Our results show that the majority of websites implementing cookie paywalls do not contain any major violations of the GDPR and ePD. However, many different implementations exist. Some contain dark patterns, and some misuse legitimate interest. With websites using dark patterns, it is feasible to question if there is any actual interest in being paid.

Keywords: Computer, science, engineering, project, thesis, cookies, cookie paywall, paywalls, legitimate interest.

Acknowledgements

We would like to express our deepest gratitude and appreciation to the individuals who have contributed to the successful completion of our master thesis. Their support, guidance, and expertise have been invaluable throughout this journey.

First and foremost, we extend our heartfelt thanks to our supervisor, Viktor Morel. Your unwavering dedication, insightful advice, and constant encouragement have been instrumental in shaping our research and pushing us to achieve our best. Your expertise and guidance have been invaluable, and we are truly grateful for your mentorship.

We are also indebted to Per Grundtman, our Omegapoint supervisor. Your continuous support, industry insights, and practical guidance have enriched our research experience. Your willingness to share your knowledge and expertise has been immensely valuable, and we sincerely appreciate your contributions.

We would also like to extend our gratitude to Alejandro Russo, our esteemed examiner. Your thorough review, constructive feedback, and valuable suggestions have greatly enhanced the quality of our work. Your expertise in the field has been an inspiration and a guiding light throughout this thesis.

Additionally, we would like to express our gratitude to all the other individuals who have played a significant role in our thesis, including our fellow students, colleagues, friends, and family members. Your encouragement, understanding, and support have been vital in helping us overcome challenges and stay motivated during this demanding endeavour.

Finally, we acknowledge the institutions and organizations that have provided us with the necessary resources and facilities to conduct our research. Your commitment to fostering academic growth and innovation is deeply appreciated.

To everyone mentioned above, as well as those who have contributed in ways that may not be explicitly mentioned, we extend our heartfelt thanks. This thesis would not have been possible without your unwavering support and guidance.

Sincerely,

Viktor Fredholm, Simon Hansson, Gothenburg, 2023-07-01

Contents

List of Figures

List of Tables

1

Introduction

The experience of browsing the web and getting cookie banners that ask for your consent go hand in hand for many. With the introduction of the *General Data Protection Regulation* (GDPR) in May 2018, the occurrence of cookie banners increased [1]. These cookie banners might be annoying for most, but they also give us a choice over what information to share with advertisers.

To address uncertainty within the commercial sector, the *Transparency and Consent Framework* (TCF) was created. The TCF was designed by the *Interactive Advertising Bureau Europe* (IAB Europe), which is "the European-level association for the digital marketing and advertising ecosystem" [2]. The framework proposes a standard for parties collecting and processing personal data to ensure compliance with regulations while pursuing large-scale data collection and processing.

Cookies can be regarded as personal data regulated by the GDPR [3]. The value of cookies without consent is however limited. It is only permitted for technical cookies (required for the functioning necessary for the service) and statistical cookies (only when not shared with third parties and not used for profiling). The GDPR, along with *the ePrivacy Directive* (ePD), specifies requirements for lawful collection of personal data, both with and without consent [3, 4]. However, it is debatable whether cookies respect choices made by users. As demonstrated in a previous work by Matte et al. [1], numerous cookies violate user choices and, therefore, the requirements mandated by the GDPR and ePD. The *ePrivacy Directive* (ePD), although older, it complements the GDPR, acting as a "*cookie law*" [4, 1]. In 2002 the ePD passed and has since been a directive on how electronic communication providers should handle personal data. One requirement of the ePD is that providers have to gather consent before any processing of personal data. Since the ePD is a directive, each country has to have a *data protection authority* (DPA) to interpret the ePD.

Besides traditional *consent* cookie banners that appear in usual websites, a novel subcategory of cookie banners was identified in 2022 by Morel et al. [5]. This new type of banner only allows the user to accept the tracking or pay a fee not to get tracked. We refer to these options as *cookie paywalls*. Cookie paywalls are new enough not to have been part of the analysis of cookie banners, which were previously studied by Matte et al. [1].

1.1 Motivation of this research

Many users are used to interacting with cookie banners, becoming frustrated by their design, and many do not like the idea of being tracked and profiled while browsing [6].

The research of Matte et al. [1] covers an in-depth study of cookie banners where they inspect cookie banners to see whether they adhere to their claims of tracking and how the design of cookie banners may differ. Matte et al. [1] found that over half of the websites they looked at did not handle personal data lawfully and that the fundamental design of the TCF is flawed. Then, more recently, Morel et al. [5] found that the yearly cost for not being tracked by cookie paywalls ranged from 36€ to 75€ per year. The results of Matte et al. [1], along with the preliminary study of Morel et al. [5], pave the way for further research on the analysis of cookie paywalls. Since Matte et al. [1] found that 1 426 out of 28 257 websites implemented the TCF when they looked in 2020, this is what we will look at in the cookie paywalls. Both of the mentioned papers also provide clear guidelines for tracking analysis on cookie paywalls.

1.2 The problem

A manual inspection of all interactions with cookies done by a single user would be infeasible, with each visited page containing approximately ten cookies [7]. Additionally, users are often nudged towards consenting to all cookies [8], further eroding user rights. Due to the uncertainty surrounding the implementation of cookie paywalls, mentioned in the previous section, it would be reasonable to want to know whether websites track our browsing habits even after opting out of tracking. Furthermore, with the introduction of paying to deny personalised ads and tracking, covered by Morel et al. [5], it is even more crucial that the websites do not track the users' habits. We intend to investigate the cookie paywalls and their implementations to deliver an overview of their lawfulness.

1.3 Purpose and goals

In our research, we strive build a understanding of cookie paywalls, their tracking claims, and their implementations. The results of this analysis will aid in answering whether it could be worth paying the fee not to get tracked, or if there still is some degree of tracking whatever the user chooses.

Described at a high level, the first goal is to identify which information can be retrieved by cookies or other means of local storage from cookie paywalls and vetting their compliance against their claims. Part of this goal will also be to investigate whether cookie paywalls correctly implement the technical specifications of the TCF.

This leads to the following questions, which we will address in this project:

1. Is the data subject not tracked if the subscription is paid?

2. Is the TCF consent management correctly implemented on cookie paywalls?
3. How much do cookie paywalls use legitimate interest?
4. Are all CMPs equal?
5. To what extent are dark patterns present on cookie paywalls?

1.4 Outline

The rest of this thesis is organised as follows. We begin by diving into some necessary background inchapter 2, where we explain data privacy regulations and directives, and decisions from authorities on related questions. In the same chapter, there is also an explanation of concepts on cookie banners, their design, and tracking on the web, which is necessary to follow the rest of this thesis. After the background, we present an overview of related work in chapter 3 where we look at other research in the area of GDPR, tracking and cookies. We will also highlight where our research fits into this research space. Later, in chapter 4, we describe our research goals, how we gather cookie paywalls, what tools we use to gather information, and what information we will collect. Then the data is presented in the results, chapter 5. After presenting the results, we will discuss them and lift our opinions. Then, more importantly, what could be done with this in the future and if this research could be used to implement some countermeasure against tracking on the web. Finally, we conclude our work in chapter 7.

The research of this paper will focus on the style of cookies found on websites and applications. For edible cookies, see the cookie recipe in Appendix A.

2

Background

Cookies, the GDPR, and tracking on the web can be daunting to understand. We have collected and explained the necessary parts to help set a good base for understanding the rest of this thesis. This chapter is split into three parts, legal, tracking, and interface. Firstly, in the legal section (section 2.1), we write about the underlying legal frameworks related to tracking and cookie banners. The GDPR, ePD, and legal texts will be explained in this section, along with different court decisions on the lawfulness of cookie paywalls. Then, we provide in section 2.2, information on the implementation of online tracking, how personal data is traded, and the implementation of cookie banners. The largest corporation, IAB, claims to manage tracking and offer implementation guidelines for managing personal data [2]. Lastly, we introduce the main design differences between cookie paywalls and cookie banners in the interface section (section 2.3).

2.1 Legal

Several legal frameworks cover cookies and online tracking, but the most relevant for EU citizens are the GDPR and ePD [9]. In this section, we begin by explaining relevant parts of the GDPR and related regulations and directives. Then ew write about court decisions regarding cookie paywalls and their implementations.

2.1.1 GDPR and data privacy

The GDPR is a data protection legislation with rules on processing personal data belonging to EU citizens [10]. According to the GDPR, the only cookies allowed before a user has given their consent are functional cookies [11]. Consent is defined by the GDPR as *any freely given, specific, informed, and unambiguous indication of an individual's wishes by which they, through a statement or a clear affirmative action, signify agreement to the processing of their personal data.* Users must also be notified of any tracking and be able to act, opt out, or leave, before any tracking is initialized. This means users must opt in to tracking before a website can track them. Additionally, the GDPR requires that withdrawal of consent should be as easy to withdraw as to give consent [11].

Prior to the GDPR, the former law regulating personal data collection and processing in the EU was the EU DPD [12]. The EU DPD has been active since 1995 and has

been the foundation for the minimum requirements for handling personal data on the web. When the GDPR replaced the EU DPD in May 2018, it catalysed numerous organizations to join the data privacy movement, as noted by Kessler [9]. Since then, various agencies and regulations have followed, drawing inspiration from the GDPR. One is the *California Consumer Privacy Act* (CCPA), which went into effect in January 2020. The GDPR and CCPA provide users with rights over their data, such as the right to be forgotten, access, rectification, and the user's ability to opt out of sharing their data [13].

The CCPA and GDPR are both privacy laws designed to protect the personal information of individuals [9]. While the GDPR applies to businesses processing the personal data of EU residents, regardless of business location, the CCPA only applies to businesses operating in California.

Another difference is that the GDPR requires companies to have a legal reason for processing personal data, while the CCPA only requires notice and consent for some types of data sharing [9]. However, in the case of CCPA, this does not apply to children, as there is a separate set of requirements when it comes to the personal data of children.

2.1.2 The ePD and ePR

As mentioned in chapter 1, the ePD complements the GDPR in specifying how electronic communication providers should handle personal data [4]. The ePD supplements the GDPR in many ways, in some even overrides, and the ePD address many important aspects of the confidentiality of electronic communication along with the tracking of internet users. In 2018, the ePD was supposed to be replaced by the *ePrivacy regulation* (ePR). This decision was to expand on the definitions of the directive and act as a central regulation that all EU countries have to follow. However, the EU missed that goal, but document drafts are available online [14]. The draft of the ePR is currently being negotiated, and there is no official date when the new standard will be introduced.

One of the six legal grounds for processing personal data under the GDPR is *legitimate interests*, and it allows vendors to process personal data without the explicit consent of the users [15]. According to the GDPR, processing is lawful if it is:

> *"necessary for the purpose of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child."* [15]

The concept of legitimate interests recognizes that certain types of data processing comply with the GDPR without user consent [15]. Examples could be legitimate business purposes or lawful reasons. However, for vendors utilizing legitimate interests in tracking purposes, it is still essential to provide clear information to the user on the purpose of these legitimate interests. Users also have the right to object to the process of their data based on these legitimate interests, and the vendors have to

respect the user's objections unless the vendor demonstrates compelling legitimate grounds for the processing.

2.2 Technical concepts

In this section, we begin by describing a common framework for implementing consent storage and tracking. Then we dive deeper into how the consent is stored, which is crucial to understanding this thesis. Lastly, we describe one way that personal data is commonly traded online.

2.2.1 TCF by IAB

IAB Europe created the TCF with the vision of helping the digital advertising industry interpret and comply with EU regulations on data protection and privacy [16]. *IAB Europe* is the European branch of the larger corporation *IAB* [2]. IAB represents the digital advertising industry, and the mission of IAB Europe is to promote the development of that industry in Europe. IAB Europe includes over 700 companies from across the digital advertising ecosystem. One of the key initiatives of IAB Europe is to develop standardized formats to obtain user consent and inform users about the processing of their data.

In 2018 the first TCF version was released (TCF v1). It was the first attempt by IAB Europe to standardize the collection of consent for data processing [16]. The first version of the TCF contained a set of technical specifications and policies that website owners and vendors had to follow to comply with the framework.

The second version of the TCF (TCF v2) was released in 2019, but adoption did not commence until 2020 [17]. The TCF v2 was designed to better comply with the GDPR and to provide users with better transparency and control over their personal data. One fundamental improvement of the TCF v2 is the inclusion of a global vendor list, allowing users to see what vendors were collecting and processing their data. TCF v2 also includes new legal bases for data processing, such as legitimate interest.

Going forth, IAB has released what is called the *Global Privacy Platform* (GPP), which was released in June 2022, to integrate with the already existing TCF v2 [18]. GPP is a single protocol intended to make sending privacy, consent, and consumer choices from sites and apps to advertising providers easier. Additionally, there has just recently been an announcement that a new version of the TCF is coming, version 2.2 [19]. We have yet to find any notions of GPP implemented, and TCF v2.2 has just been announced, not implemented. Because of this, this research will only look at the TCF v2 implementation.

2.2.2 Publishers and advertisers

Publishers and advertisers play crucial roles in the current browsing experience [20]. Publishers are individuals, companies, or organizations that create and distribute

content online, such as websites, blogs, or mobile apps. They commonly generate revenue by attracting an audience and selling advertising space on their platform. Publishers often rely on advertising to fund their content creation and maintenance costs. Advertisers, on the other hand, are individuals, businesses, or agencies who promote their product, service, or brand through online advertising. They seek to reach their target audience by displaying ads on various digital platforms. Advertisers typically pay publishers for ad placements or rely on ad networks to facilitate their advertising campaigns.

In the context of a website, the publisher is the provider of the website, and on that website, the advertiser pays the publisher to be able to display their ad [20]. This can also be because the advertiser wants to buy user data which the publisher has acquired. Then the advertiser could use this data to be able to target the ads more to specific individuals, change the design of their product, or to try and affect the habits of the user.

2.2.3 Vendor and CMP

Two ubiquitous terms in this paper are *Compliance Management Platform* (CMP) and *vendor*. The term *vendor* refers to a person or company who manages, buys, or sells personal data to other companies or individuals. CMPs were first introduced by IAB in 2018 [17]. The purpose of a CMP is to be in charge of collecting and communicating user consent. How the information is intended to flow and how the interaction is done between publishers, CMPs, and advertisers can be seen in Figure 2.1. Gathering and distributing consent typically involves displaying a consent pop-up or a banner informing users about the website or app's data collection and processing practices. Being able to exercise user rights to access data, modifying consent, or being forgotten are also parts of what CMPs manage.

2.2.4 The TCF consent string

The core technical aspect of the TCF is the consent string. The string allows for a standardized approach to transmitting user consent between websites, CMPs, and vendors [16]. The TCF consent string contains all information disclosed by a user and the expression of their preferences for processing their personal data. The information is encoded into a string using the preferences entered by the user in a consent pop-up. By decoding the string, vendors can ascertain whether there is a legal basis for processing personal data. Based on the preferences submitted by the user, the string can contain information regarding a set of different topics. General metadata is always encoded into the string. This includes information such as when the user submitted their preferences and which CMP and Global Vendor List (GVL) version is used. User consent is also always encoded, and the consent preference is encoded in two contexts, relating to specific purposes and specific vendors.

In Figure 2.2, is an example of a decompiled TCF consent string. The TCF consent string can also contain information about legitimate interest, publisher restrictions, publisher transparency and consent, and specific jurisdiction disclosures [16]. However,

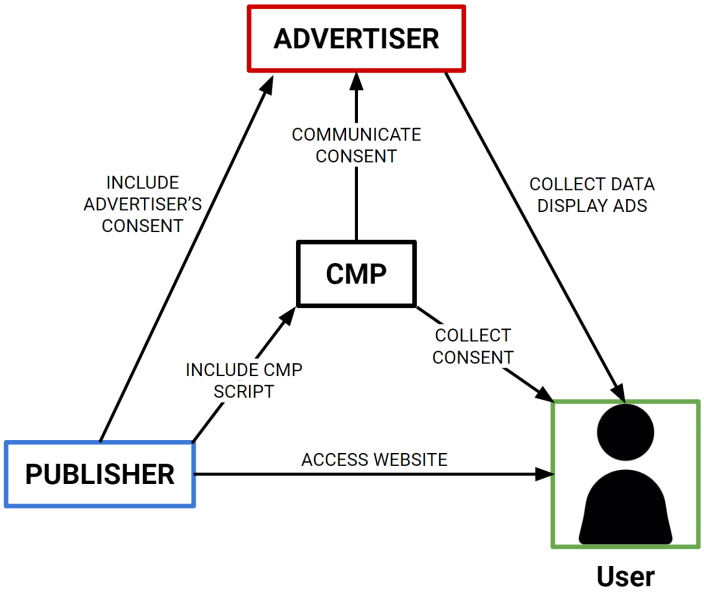


Figure 2.1: The role of a CMP according to the TCF


```
 { "version":  2,
"created":  "2023-05-31 13:37:12",
"lastUpdated":  "2023-05-31 13:37:12",
"cmpId":  12,
"cmpVersion":  1,
"consentScreen":  2,
"consentLanguage":  ENG,
"vendorListVersion":  187,
"policyVersion":  2,
"isServiceSpecific":  true,
"useNonStandardStacks":  false,
"purposeOneTreatment":  false,
"publisherCountryCode":  SE,
"allowedPurposeIds":  [1,2,3,...],
"allowedVendorIds":  [1,2,3,...],
...}
```


Figure 2.2: An example of the data in a decoded TCF consent string.

this thesis will only focus on user preferences regarding consent per purpose, legitimate interests and vendors.

2.2.5 Consent storage

Storing the TCF consent string can be done in various ways. The TCF guidelines provide specific requirements for how this is done. The IAB published a webinar detailing the switch from TCF v1 to TCF v2 and good information on how consent is managed [21]. According to the TCF, the consent string should be stored using their base64url encoding scheme and stored locally in the user's local storage or as a cookie. However, new in TCF v2 is that how the TCF consent string is stored is ultimately up to the CMP to decide. Usually, the string is stored in a cookie with the name *euconsent-v2* or in the local storage with a key value of *"_sp_user_consent_XXXX"*, where *XXXX* represents a number [1, 16]. With freedom for the CMP to implement the consent storage, there could also be a larger risk of a faulty implementation and, therefore, non-GDPR-compliant implementation.

The local storage of a web browser is commonly used in modern websites [22]. A key-value pair is used to store the data, and storing information locally on the user's computer can be helpful in the functionality of more advanced websites. This allows the websites to better keep track of the state and all other things which may be necessary. Unfortunately, anything stored in the local storage will be accessible by any script present on the current webpage. This may lead to others accessing the information more easily [23].

Conversely, cookies are a predetermined set of attributes stored by the user's browser [24]. The attributes of a cookie are:

- Name: *Name of the cookie*
- Value: *Stored value*
- Domain: *The domain that created the cookie*
- Path: *Only sends cookies with requests of this path*
- Expiration: *Expiration date*
- Size: *Size of the cookie*
- Http only: *Whether a cookie can be accessed from client-side scripts*
- Secure: *Cookie only sent with HTTPS requests*
- Same Site: *Cookie only sent with first-party requests*
- Partition key: *Partition key of the cookie*
- Priority: *Priority for re-authentication of cookies*

Unlike local storage, depending on the value of these cookie attributes, each cookie may be sent to the server with each request and can be used to track a user's browsing behaviour.

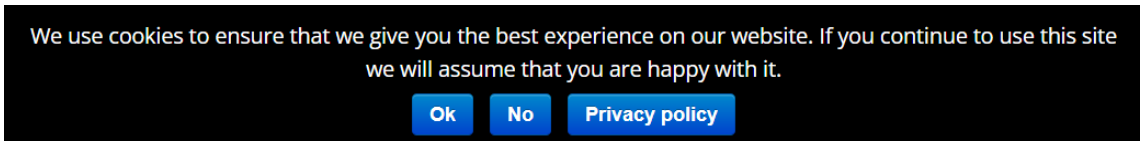


Figure 2.3: Cookie consent banner from *https://www.gdpr.eu/* (accessed 5. Mar. 2023). The user can browse without making a decision, and the options are all presented directly in the banner.

2.2.6 Real time bidding

An important functionality of the TCF is simplifying the transfer of personal data between vendors, and this functionality is often used in real-time bidding [25]. Real-time bidding (RTB) is a live auction of advertising space. When a user visits a web page, an auction is started where multiple vendors can bid. With the increase in the use of RTB by advertising vendors, there has been an increase in the distribution of personal metadata [25]. Hence, the value of personal data increases with the precision of the data, leading to a conflict of interest for web pages. The hosts of web pages will gain financially from collecting a larger amount of data, including before the user has consented, but risk fines for non-compliance with the GDPR. Secondly, personal data is distributed to a wider group of companies, making it harder for the user to understand who has permission to handle the data. Research on RTB is presented in section 3.5.

2.3 Interface

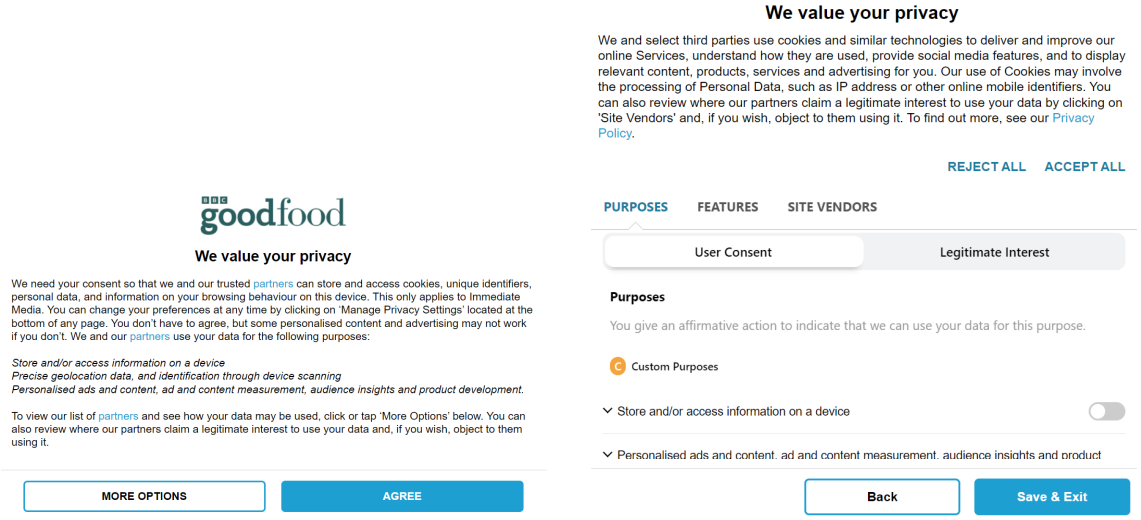
In this section, the interface of cookie banners and cookie paywalls will be described along with examples to clarify further the difference and how they both work. Then we describe what some illegal design choices might be.

2.3.1 Cookie consent Banners

Since the new regulations demanded that users' consent be collected before handling their personal data, cookie consent banners got widely adopted [1]. A cookie consent banner is a graphical element or notification that appears on a website to inform visitors about the usage of cookies and what tracking will be taking place. They also serve to gather user consent as required by regulators, such as specified by the GDPR in Europe. These banners can range from a small amount of text notifying the user to complete web pages with complex menus where the user has to navigate multiple choices. This can be seen in the simple example in Figure 2.3 and the more complex example in Figure 2.4.

2.3.2 Cookie paywalls

A subcategory to the cookie consent banner is the cookie paywall banner, referred to as *cookie paywall*, and can also be colloquially referred to as *pay or okay* [26]. In Figure 2.5, we see two clear choices, and instead of the *deny* button, there is an



(a) Cookies banner view 1

(b) Cookies banner view 2

Figure 2.4: Cookie consent banner from *https://www.bbcgoodfood.com/* (accessed 5. Mar. 2023). Upon loading any page, the user is presented with the first view in 2.4a and must decide before browsing. After selecting '*more options*', page 2.4b appears.

alternative to pay or to log in to an existing account. Either we accept the advertising and tracking, or we pay 2.99€ to have an experience free from advertising, and the website will not track us.

The Austrian DPA (DSB Austria) stated that cookie paywalls could be an acceptable alternative to traditional cookie consent banners [27]. More data protection authorities have deemed cookie paywalls to be lawful under certain circumstances, such as the French DPA (CNIL) and Danish DPA (Datatilsynet) [28, 29]. All the above-mentioned DPAs agree on most of the guidelines for a lawful implementation of a cookie paywall. Some shared descriptions that we find notable are that:

- The fee for declining tracking has to be reasonable.
- The company is not allowed to have a cookie paywall if they have a monopoly or quasi-monopoly position in the market.
- If a user accesses the website through the paid alternative, no personal data may be processed for advertising.

2.3.3 Dark patterns

Different methods exist for retrieving consent from data subjects, including the pop-up. Designs of pop-ups vary wildly between web pages. However, there is a trend toward creating designs which nudge users towards unwanted choices. These designs are so-called dark patterns [8]. Nudging via dark patterns aims to persuade the user to interact in a specific way subliminally. In the context of cookie banners, the goal is to get the user to consent to all cookies. Dark patterns can be making the

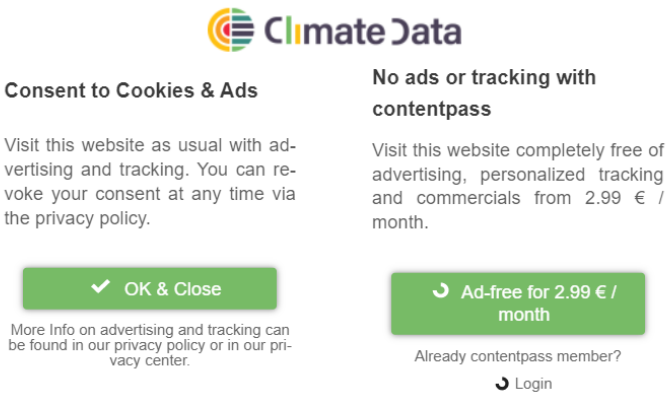


Figure 2.5: Cookie paywall from *https://en.climate-data.org/* (accessed 5. Mar. 2023) where it is possible to agree to tracking, pay not to be tracked, or sign in to an existing account.

user think that there is no other choice, making the option to disagree too complex that users will not bother, or by making the consent option look like the user is opting out of tracking.

Intending to promote cooperation, information sharing and best practices, the *European Data Protection Board* (EDPB) created a cookie banner task force [30]. This task force reviewed complaints about cookie banner practices and released the report in January 2023 [31]. In their report, the EDPB task force commented on design patterns, the lawfulness of these in cookie consent banners, and how they compare to the GDPR consent requirements. Four of these design patterns felt especially interesting for this thesis.

1. The reject button has to be at the same level as the consent button to comply with the ePD. However, they reflect that the ePD does not explicitly mention a "reject option".
2. If there are pre-ticked boxes to opt in to consent, then this does not count as an action by the user to opt in and thus does not lead to valid consent.
3. There has to be a clear indication of what the banner is about. For the consent to be valid, it should be clear what the user is agreeing or disagreeing to, and the user should understand what choices and alternatives they have.
4. It should always be easy to withdraw one's consent. There should be a small hovering and visible icon or a link placed in a standardized place, easy to find by the user.

3

Related work

While cookie paywalls have yet to be extensively studied, much work has been done on related topics. This work relates to the technical aspect and the user interaction with cookies and cookie banners. In this section, we will present research directly related to the proposal of this thesis. The position of this project is intended to be at the intersection of the areas presented, with a focus on cookie paywalls.

3.1 GDPR compliance and tracking

In previous research by Matte et al. [1], they found that more than 50% of the web pages that the authors analysed did not comply with the GDPR, with all offending web pages implementing the TCF. Matte et al. [1] looked at 28 257 websites, of which 1 426 implemented the TCF. The authors extensively tested these websites to determine whether they and the TCF adhere to the GDPR and the ePrivacy directive.

To analyse the cookie banners in depth, Matte et al. [1] performed automatic, semi-automatic, and manual testing. Firstly, the automatic approach was made by crawling the web. Then, for the semi-automatic approach, the authors manually navigated menus and recorded consent strings using automatic scripts. Lastly, the authors had three individuals navigate the cookie banners, noting down the details on design and tracking present on those web pages. The results of Matte et al. [1] show that there were many violations of the GDPR and ePrivacy directive, more than 50%, and that different websites that implement the same CMP did not always behave the same. The authors found that compliance with the TCF would not necessarily mean that the owners and vendors would comply with the ePrivacy directive and the GDPR. Additionally, while only looking at the first version of the TCF, the base specification did not comply with the GDPR.

Three websites that Matte et al. [1] looked at stored a full consent of user tracking and sent the information to several other vendors before there had been any interaction from the user. What this meant was that if they visited one of the websites from one of those vendors, they had stored a full consent of user tracking before any interaction as well.

Matte et al. [1] noted an issue with the TCF where it assumes that all the CMPs that set the string are doing it correctly. There is also the issue that anyone can

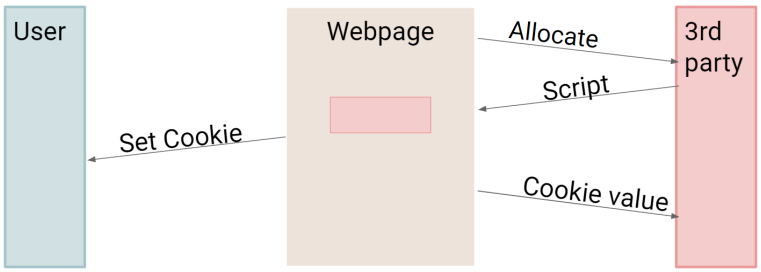


Figure 3.1: Illustration of a third party setting first-party cookies.

create a consent string, meaning that CMPs who are not originally present on a webpage or who are not the original CMP, can set consent strings. What this means is that any CMP can come along and set a new user consent, leading to a website not complying to the GDPR because of a non-compliant CMP.

Consent banners should assign cookies to users after they have selected which data will be tracked or not. However, this is only sometimes the case [1]. These cookies could be first- or third-party cookies. First-party cookies are handled by the website currently visited, whereas an external party handles third-party cookies. First-party cookies are often used to keep track of shopping carts, log-in status, and other usability features on modern websites. As third-party cookies are used by vendors other than the one currently visited, these cookies are often used to track user habits across multiple websites [32].

3.2 More on tracking

The research by Chen et al. [32] discovered that a third party can access first-party cookies. With the value of this first-party cookie, this third party could track the user over multiple websites without the need for persistent third-party cookies. Chen et al. [32] also found that a third-party script would have to be present to access a first-party cookie, which then sends the cookie value back to their database.

To use first-party cookies for tracking, Chen et al. [32] describes that third-party companies often embed code or scripts on the first-party website. This could be in ads or any dedicated space on the website. In the sequence diagram of figure 3.1, we can see the flow of a third party using the allocated space of a website to load their content. The sequence diagram starts with the webpage communicating to the third party that space is allocated on the webpage. Then the third party sends their script, which could be an ad, to the webpage to load. In this script, they have also added code to set a cookie in the user's browser in the next step. Finally, after setting a cookie in the user's browser, the third-party script will send a message home with the cookie value that has been set. With this code, when the third-party script sets or accesses cookies, the browser will see it as if the first-party webpage does it, and the cookie can be tracked in subsequent requests made from the user within this webpage.

A find by Munir et al. [23] was that third-party scripts could use the value of known

first-party cookies to track users. An example of this could be the Google Analytics tracking ID cookie, which is present on many websites. Instead of setting a cookie and returning the value of their cookie, the third-party script could search for a common name for one of these cookies, fetch the value of the cookie and send that value home to the third party's server. Then tracking would continue as if the third party had created their cookie.

In the study by Munir et al. [23], the authors develop a machine learning-based approach, called *CookieGraph*, to detect first-party cookies used for third-party tracking. The accuracy of *CookieGraph* is good, but the most exciting part for our thesis is the research that led to the development of *CookieGraph*. Munir et al. [23] found that some third-party scripts present on websites use the local storage to store a user ID. Storing the user ID like this can be a way to get around modern browsers and their automatic blocking of third-party cookies.

Cookies are not the only way to track users. Papadogiannakis et al. [33] discovered that techniques such as first-party ID leakage, ID synchronisation, and browser fingerprinting could be used for tracking. These methods are harder to detect and sometimes even impossible for users to check for themselves, as some are not performed on the client side. Fortunately, this type of tracking still falls under the GDPR as it is still personal data. However, this adds lots of complexity to analysing tracking, and this will not be included in the scope of this thesis.

3.3 Dark patterns and legitimate interest

In the study by Nouwens et al. [8], some dark patterns were considered non-compliant examples of the GDPR. When the authors scraped a set of 10 000 websites, they found that only 11.8% met the minimal requirements according to European law. They highlighted how the prevalence of dark patterns in consent pop-ups undermines the intended goals of the regulation and limits users' ability to make informed choices about their personal data. The experiments examine how people not used to consent banners interact with dark patterns [8]. The results of the research by Nouwens et al. [8] show that users gravitate towards the easy bulk options, such as *"accept all", "reject all"* and *"submit preferences"*, instead of taking the time to navigate the complex submenus and reading the meaning of those options. There has also been a further investigation into dark patterns by Hausner and Gertz [34], where they developed an automatic detection system for dark patterns.

According to the findings of Kyi et al. [6], using legitimate interest does not align with the user's expectations. However, the *legitimate interest* part of tracking can be a complicated subject, and misuse of legitimate interests counts as a dark pattern according to the EDPB task force [30]. Kyi et al. [6] mention two studies in their report. First, they identify various deceptive designs when legitimate interests are implemented. Second, they found that how legitimate interests are used did not align with the users' expectations.

3.4 Paywalls

Morel et al. [5] conducted exploratory research on cookie paywalls, analysing websites published by companies active in Europe. In this research, the authors explored websites in Europe and assessed whether they had a cookie paywall, what type of paywall, the website category, and additional notes. Seven different classifications of user access restrictions are introduced, with five representing a paywall. These five paywalls are different variations requiring payment for accessing the website's content. The three classifications most relevant for this thesis are **paywall**, **cookie wall** and **cookie paywall**. **Paywall** demands that the user pay to access some content or functionality. **Cookie wall** denies user access if they do not consent to all vendors present on that website, regardless of payment. **Cookie paywall** provides the user with two choices, either consent to tracking or payment for tracking free access. With 2800 websites analysed, 1000 Austrian and 150 from 12 different EU countries, Morel et al. [5] found that 13 contain some cookie paywall. For these 13 websites, the annual fee ranged between 36€ and 75€.

Furthermore, the research presents an overview of the current legal status of cookie paywalls, a summary of which is found in subsection 2.3.2. Morel et al. [5] conclude that the variation in legality creates uncertainty for all parties involved and that more research is needed in the area of cookie paywalls.

Searching for *paywall* at `https://dblp.uni-trier.de/` results in 14 papers about paywalls which restrict content if not paid, and one paper which mentions cookie paywalls. As one of the predecessors to the cookie paywall, the traditional paywall has been researched from 2012 to 2022. The traditional paywall has more papers released in 2020 and 2021, focusing on profitability, than any other year. With the decline in the popularity of printed news, mainly due to easy access to alternative online channels, the advertising revenue from these media has become less profitable [35]. Therefore, many news publishers have focused on online advertising and paywalls to monetise their content [35]. These news paywalls are not a new phenomenon. Paying to see the content of an online newspaper has been typical since before 2013. In the study by Myllylahti [35], the author examined whether news publishers benefited from having paid walls to access their content.

Myllylahti [35] found that these subscriptions represented around 10% of the total income of the news provider. An argument was made about whether paywalls would be a viable business model, Myllylahti [35] also argued that smaller providers were more likely to be harmed by traffic and ad revenue loss.

3.5 Research on RTB

RTB (or *real time bidding*), described technically in subsection 2.2.6, has been covered by Liu et al. [36]. In their paper, Liu et al. [36] examined what bidding took place depending on users' browsing habits. The authors found a significant difference in the bidding behaviour of some CMPs depending on the user's choice. However, some advertisers would process personal data even if the users opted out of being tracked.

Liu et al. [36] also found that automated services, such as the National Advertising Initiative's (NAI) opt-out, could be equally effective as manually opting out.

To carry out their research, Liu et al. [36] captured the bids and synchronisation of the cookies from the advertisers. This information was gathered on a set of 16 fake personas, and the authors could simulate a persona by visiting websites and filling in information similar to those of predefined imaginary people (different personas). Each persona was designed to behave differently to one another. Some allowed cookies, some did not, and different types of websites were visited. All of this allowed the vendors to build up a profile of the persona before the testing occurred. Then Liu et al. [36] measured the targeting for the data bidding when users opt out of processing and selling of data. Targeting was measured by looking at header bidding on the client side. Finally, these results were compared to what happened if the personas opted into tracking or used state-given tools like NAI's central opt-out. The results showed that even though users opt out of tracking, there is a clear difference in bidding behaviour, showing that users' choices are likely not fully respected.

3.6 Lawfulness of TCF

The *Belgian data protection agency* (GBA) issued a decision concerning IAB Europe and the TCF in February 2022 [37]. The ruling concludes that IAB Europe is acting as a data controller for the TCF string. Furthermore, it is also concluded that IAB Europe is a joint controller with CMPs, publishers, and ad-tech vendors. Classification of IAB Europe as a data controller stems from the ruling that the TCF string should be considered personal data. According to the GBA, the TCF's proposed legal grounds are invalid. Most important for this thesis is that collecting consent and classifying legitimate interests is non-compliant with the GDPR. The consent collected by CMPs implementing the TCF is invalid for several reasons, and the two relevant for cookie paywalls are presented here. Firstly, the standardised purposes provided by the TCF are too vague, not permitting the user to make an informed decision. Second, too many actors are involved, requiring the user to spend a disproportionate amount of time understanding their choices. Similarly, the GBA rejects the use of legitimate interest as a legal basis for data processing. Both issues with the consent collection are also present in establishing legitimate interest.

4

Methods

In this section, we present the methodology employed in our project. We begin by outlining our research approach. After which, we detail the dataset of cookie paywalls. Followed by an explanation of the methods used to extract data from these paywalls. Finally, we explain the specific data we will examine, our intended presentation format, and the analytical techniques we will employ to analyse the gathered data.

What we aim to do in this research is similar to what Matte et al. [1] did in their research. However, instead of looking at cookie banners, we will look at cookie paywalls, a sub-genre of cookie banners, and similarly look at if they behave as they should. Matte et al. [1] mentioned that the second iteration of the TCF had been released, but it had yet to be widely adopted and was therefore not part of their research. When we look at the cookie paywalls, we expect that the new iteration of the TCF will be implemented and the tools available from Matte et al. [1] will be outdated, but we will look to see whether they are.

4.1 Research approach

The research approach for the research questions is a quantitative approach; this involves collecting and analysing numerical data to evaluate and answer the research questions. To build a good understanding of the current state of cookie paywalls, we will look at as broad a set as possible. Descriptive statistics will be used to summarise the characteristics of cookie paywalls. Using this method, we aim to analyse the tracking characteristics of cookie paywalls.

The tools used to accomplish this are a Python script using the Selenium framework, a clean installation of a Chromium browser, and a Jupyter Notebook. To do large-scale analysis, we used the Selenium framework. Then to get more detailed statistics, the Chromium browser is used to investigate the implementation of cookie paywalls manually. Finally, to manage statistics and display the relevant information, in an easily readable way, the Jupyter Notebook Deepnote is used.

We have collected some cookie paywalls by browsing CMPs, news websites, and looking at paywalls found in previous research. Fortunately, we collaborated with another group of master students working on a connected topic, i.e. a quantitative measurement of cookie paywalls. After one of their early runs, they got a set of

230 cookie paywalls, which we deem is enough to get a view of how cookie paywalls behave.

4.2 Data extraction

Two main sets of cookie paywalls are analysed, one containing the cookie paywalls that use `https://contentpass.net/`, and another where the *contentpass* sites have been removed from our set of 230 cookie paywalls. Firstly, an automated approach analysing cookie paywalls utilising the *contentpass* paywall. This approach is possible due to all *contentpass* websites utilising a standardised pop-up and login process, and only one payment is required, more information in subsection 4.2.1. Secondly, a semi-automated approach towards the remaining cookie paywalls. We do not look at what happens after the user pays; otherwise, it is the same check as the fully automated test, and for more information, see subsection 4.2.2. Lastly, there is a manual approach on a random subset of the set of cookie paywalls which does not include *contentpass*. The randomly selected subset is limited in size by budget and time constraints, which we explain more in subsection 4.2.3.

If all websites were to be fully analysed, the cost would be too high and not proportional to the information collected; the same is true for the time required. For the automated and manual approaches, the focus will be on collecting the TCF consent string, before any interaction, after full consent, and after payment. Then for the semi-automated approach, the TCF consent string will only be collected before interaction and after full consent.

When gathering the information on the cookie paywalls, for all websites not available in Swedish or English, `https://www.deepl.com/` will be used to translate the information on the cookie paywall. The browser used to inspect these paywalls is a clean installation of Chromium [38]. A clean installation means the Chromium browser was installed just before gathering the information. This ensures that there are no cookies or other cache from previous browsing and no browser extensions that will affect the results. Additionally, to ensure that no cookies are missed, all automatic third-party blocking is disabled in the Chromium browser. Data will be collected from an IP address originating in Gothenburg, Sweden.

4.2.1 Contentpass

Data collection from websites implementing the *contentpass* paywall is automated using Python and Selenium. A list of all websites implementing *contentpass* is created by scraping the marketing webpage of *contentpass*. Every website in the list is visited, and the consent string is stored as it appears in all three relevant states. For the transition between the states, two different functions are implemented, one for finding and clicking the accept button and one for login in to *contentpass*. The accept button element is located by finding a button or link containing the text *Accept* in English, French, German, or Danish. These languages were selected as these were the only observed languages in cookie paywalls. The logging-in process is done similarly, firstly the link containing *login* is located and clicked, and then

the process of filling in the email and password is hard coded since all *contentpass* websites utilise the *contentpass* website for login in. In parallel with the interactions mentioned above, the TCF consent string is extracted; this is done by searching the cookie and local storage for cookies stored per the naming standards specified in the TCF.

4.2.2 Semi-automated

The semi-automated approach is to get data from most cookie paywalls. This step will focus on getting an overview of how the cookie paywalls behave before interaction with the cookie paywall. We will utilise Python to automate checking the consent recorded for websites, not part of *contentpass*. In some cases, there will be manual intervention and analysis.

4.2.3 Manually

There will be a manual analysis of a set of websites. This will make it possible for us to look at a sample set of cookie paywalls, pay the fee of that cookie paywall, and look at them from the perspective of our questions in 1.3. For the manual analysis, a selection of websites must be made. To do this, we will use the dataset of cookie paywalls, and then we will randomly select 20 websites for manual inspection using pythons built-in random generator. However, we will not select any website contained in *contentpass*, and if we find that any websites do not implement a cookie paywall, a new random website will be selected. If, for some other reason, we cannot pay to access one of the random websites, a different website will be randomly selected.

If we come across a website where the same subscription gives access to several cookie paywalls, we will look at these websites. We will see if the implementation of the cookie paywall differs from these websites and if they all behave the same. To not skew the results of the random set of websites' results, these extra analyses will be collected under their categories, separated by CMPs or subscription managers. Additionally, for this project, we have been supplied with a budget by Omegapoint which will cover the cost of all cookie paywalls we will have time to look at.

4.3 Potential violations

When analysing the cookie paywalls, we will look for potential violations. The violations we are looking for are stated below. If any of these are present on a website, we will classify that website as not compliant with the GDPR. If a website contains multiple violations, all will be recorded.

4.3.1 Tracking consent

The websites that are deemed to be non-compliant with the GDPR are those that fail to fulfil one or more of the following:

1. Before any interaction by the user, there should not be any recorded consent.

2. The recorded consent does not match the expectations of the user.
3. We, as the user, still have to manually opt-out from tracking after payment.

What we mean with the user's expectations is that when a subscription has been paid, there should not be any tracking. This is also part of the demands for cookie paywalls to be legal, covered in subsection 2.3.2.

With the claims noted, we carry on with the payment for the service; if we have an account for this service or any related services which applies to the current cookie paywall, we will use the access of that account to navigate through the cookie paywall. After getting past the cookie paywall, we will look for TCF consent strings in the cookie and local storage. The TCF consent string is then stored to be decoded into a human-readable format and used. If we are presented with a menu where we can manage consent levels before we browse the site, we will select the lowest possible consent and continue until we can browse the site, where we then record the stored consent.

The data extracted from the TCF consent string is as follows:

- Number of legitimate interests agreed to
- Number of vendors consented to

4.3.2 Dark patterns

We will also record the design when we look at the cookie paywalls. In order to identify the presence of dark patterns in cookie paywalls, we will employ a systematic review approach. We will examine various websites and their cookie paywall designs, specifically focusing on elements in the design which could potentially manipulate user choices or obscure the true nature of the cookie paywall.

During the analysis of cookie paywalls, we will carefully observe the design elements and interactions to identify any instances of the following dark patterns.

1. *Imbalanced choices*: We will assess whether the design emphasizes one option, such as accepting cookies, while making the alternative, such as paying, less apparent or less visually prominent.
2. *Misleading price*: We will compare the price displayed in the main pop-up to the actual amount drawn to identify any discrepancies or inconsistencies.
3. *Unclear intention*: We will examine the design elements to determine if the purpose of the cookie paywall is clear that the user is able to pay not to get tracked.
4. *Not opted out*: We will assess whether users are required to manually opt out of tracking or additional services after making the payment.

It is important to note that the list of dark patterns mentioned above differs from those discussed by the EDPB task force in subsection 2.3.3, which primarily focuses on cookie banners. The patterns discussed by the EDPB task force, as mentioned in

subsection 2.3.3, focus on cookie banners, which serve a different purpose and may have different designs compared to cookie paywalls. One of the clearest differences is the payment option, and we have therefore adapted these patterns to suit what we observed in previous research and some initial look at cookie paywalls. Some patterns might therefore be more relevant to paywalls, such as the *misleading price*. As traditional cookie banners do not contain payments, this would not be a relevant pattern for the task force. Since this is something that could differ in cookie paywalls, we decided that this is something which is relevant to look at.

Due to the loose definitions of dark patterns, we will not judge whether these dark patterns violate the GDPR. Along with the sources from this paper's dark patterns section 2.3.3, we will leave the decision to the reader or possible further work.

4.4 Data analysis

Our primary interest with the data collected from the cookie paywalls is to see how many of our known cookie paywalls violate the GDPR. In addition, it would be interesting to see the difference between websites with the same CMP and compare the differences between CMPs. Also, we will note any common design patterns, or possible new dark patterns, that we find on the cookie paywalls.

A few different metrics will be examined to answer our research questions.

1. *Is the data subject not tracked if the subscription is paid?*

   This depends on the TCF consent string, and there might be different levels of tracking present. If any vendor consent is registered after the payment, then we count that as *the subject being tracked if the subscription is paid*. The reason for counting vendor consent for tracking is the requirements for a lawful cookie paywall presented in subsection 2.3.2. If, on the other hand, there are Legitimate interests registered after payment, then it could depend on the number of legitimate interests and the purpose of those.

2. *Is the TCF consent management correctly implemented on cookie paywalls?*

   Much like in the research by Matte et al. [1], this question can be answered by looking at the registered consent before any interaction and the registered consent after the payment. If any consent is registered before the user has interacted with the cookie banner, then the implementation of the TCF consent management is invalid. Also, if there is any consent other than functional cookies and legitimate interests after paying, we count the implementation as invalid.

3. *How much do cookie paywalls use legitimate interest?*

   By looking at the TCF consent string, we can see how many legitimate interests have been registered. This can be done before consent, after accepting, and after paying.

4. *Are all CMPs equal?*

By separating the data based on the CMP, we can look for patterns and determine if there are differences in how the CMPs implement cookie paywalls and whether some CMPs are better than others. The differences between the CMPs will be examined on consent and design levels. For the consent level, we look at the different numbers of vendors and legitimate interests consented to in the interactive steps of a paywall. For the design level, we look for different dark patterns of the cookie paywalls and if there are stand-out differences between paywalls implemented by different CMPs.

5. *To what extent are dark patterns present on cookie paywalls?*

   To answer the question, there will be a manual check on 20+ websites where we look for the dark patterns described in subsection 4.3.2. Additional patterns and descriptions will also be noted.

5

Results

In this section, we will present the information gathered when we analysed the cookie paywalls. Then the tools we developed were used to analyse the cookie paywalls and determine whether they adhere to what they claim as well as the GDPR, and the results will be presented here. As discussed in previous sections, previous studies have concluded that many websites do not fulfil requirements set out by the GDPR, both in the collection of consent and in implementing the TCF. However, there is no research that explores if the same patterns of noncompliance also exist in websites implementing a cookie paywall. We aim to present data showing trends among different websites and CMPs.

5.1 Datasets

Three main sets of quantitative data have been collected

- websites implementing contentpass
- websites where an account has been created and paid for
- websites where only data for before and after consent has been collected

A complete list of websites and the collected data can be found in Appendix B; only the summarized data is presented here.

The tables are structured to separate the data from before user interaction, after giving full consent, and after paying and logging in. For Table 5.2, there is no *After login* since we did not pay for those cookie paywalls. Within the different consent levels, there are a number of recorded consents for vendors and legitimate interests. These are also separated to show the mean value, standard deviation, and max recorded value of the number of consented vendors and legitimate interests. In Table 5.3, are the results from the manual inspection and payment of the cookie paywalls. It can be seen that there are, on average, more legitimate interests consented to compared to the contentpass results in Table 5.1.

	Before interaction		After consent		After login	
	Vendors	Legitimate interest	Vendors	Legitimate interest	Vendors	Legitimate interest
Mean	0	0.71	49.25	49.26	0	0.72
std	0	5.22	93.32	93.33	0	5.42
max	0	73	361	361	0	73

Table 5.1: Summary of results for websites that implement Contentpass, $n = 189$

	Before interaction		After consent	
	Vendors	Legitimate interest	Vendors	Legitimate interest
Mean	0	13.49	29.19	29.19
std	0	34.79	78.44	78.44
max	0	173	356	356

Table 5.2: Summary of results for websites where no payment was made, $n = 67$

	Before interaction		After consent		After login	
	Vendors	Legitimate interest	Vendors	Legitimate interest	Vendors	Legitimate interest
Mean	0	75.43	130.93	130.93	0	6.86
std	0	40.88	88.80	88.80	0	25.66
max	0	96	345	345	0	96

Table 5.3: Summary of results for websites were payment was made, $n = 14$

5.2 Dark patterns

The cookie paywalls investigated have, in some ways, contained what we deem to be dark patterns. Following are our observations of patterns presented in subsection 4.3.2 on websites implementing a cookie paywall.

We looked at 24 cookie paywalls, and the resulting number of dark pattern occurrences can be seen in Table 5.4. Out of these 24 websites, eight of them did not match any of these definitions. In Table 5.4, to get the number of occurrences of a pair of dark patterns appearing on the same page, look at one pattern name on each edge, and then the number at the intersection between those patterns is the number of occurrences. To see the number of occurrences of a single dark pattern, look at the numbers along the diagonal, which have the same pattern described in the corresponding horizontal and vertical edges. No website contained more than two different dark patterns. Screenshots of more cookie paywalls can be found in Appendix C.

	Imbalanced choices	Misleading price	Unclear intention	Not opted out
Imbalanced choices	5	2	1	1
Misleading price	2	6	3	0
Unclear intention	1	3	7	0
Not opted out	1	0	0	4

Table 5.4: Observed dark patterns in 24 randomly selected cookie paywalls.

A common pattern in the webpages which had a *misleading price* was that there was a marketed cost per week which had to be paid on a monthly basis, along with a cancellation period which also was monthly. One example is `https://www.swp.de` where a trial is marketed as 1€, but the cost is per week and the trial period is eight weeks with the whole sum of 8€ due at once, then while subscribed, the cancellation is monthly. There was on a website we found that advertised a fee of 2€ until we got to the PayPal checkout, then the fee was 1€. This does not match with the advertised price, but we do not count it as a *misleading price* since it is not an increase.

For four of the webpages which were marked as having *unclear intention*, that was because the banner was designed to be interpreted as a traditional paywall, and the information about tracking was hidden on the second page. All of them were implemented by the same CMP. For two pages, the text describing the tracking of consent was cut off, making it impossible to read the full purpose of the banner without using the inspect option of the browser. For the last page, the option to pay to opt was hidden behind a traditional cookie consent banner, The cookie paywall only appeared when denying tracking, then the user is presented with the option to subscribe to the newspaper, and finally, at checkout, is the option to add 2€ to the fee to opt out of tracking.

In the group of webpages that were *not opted out* after payment, three of them have the same CMP and publishing house. The last one of the four had a vendor select screen after the payment where there were pre-ticked boxes, meaning that the users are not opted out by default and have to de-select the purposes, and then continue to deny tracking and continue to the webpage. This was marked as an *imbalanced choice*.

For the imbalanced choices, it was mostly the case that the button to pay was asymmetrical to the accept button, not in the same colour, or no background colour. In one case, there was a link which was placed outside the pop-up, it did not look like a button and was not in the same format as the links in the text, see top right corner of Figure 5.1.

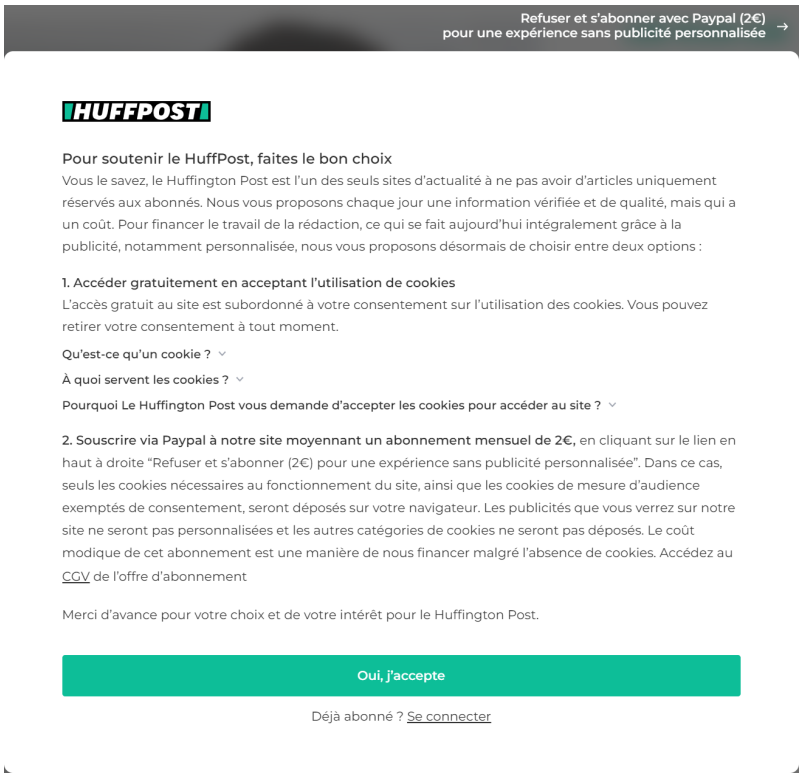


Figure 5.1: Cookie banner from `www.huffingtonpost.fr`.

5.3 Other findings

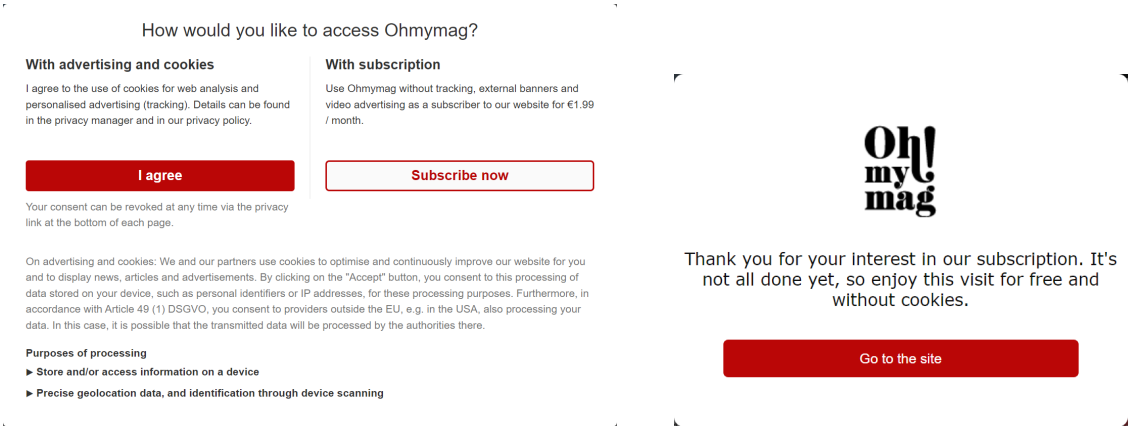
Here we present other findings which we have come across. They do not fully serve to answer our research questions, but might be interesting for anyone investigating further on the topics of tracking and cookies.

5.3.1 JSON storage

When looking for consent stored in the local storage of the browser, specifically at the webpage `www.voici.fr`, there were legitimate interests stored separately from the TCF consent string. These legitimate interests were stored under *gdpr -> customVendorsResponse -> legIntPurposes*. Some of the legitimate interests that were found are *"Select basic ads"*, *"Create a personalised ads profile"*, *"Select personalised ads"*, *"Create a personalised content profile"*, and *"Select personalised content"* This collection of legitimate interests were found in twelve of the manually inspected websites. Eleven of these websites have the same CMP and the same publishing house, which is *Prisma Media*.

5.3.2 Remote consent string storage

Another observation we have made on the consent management is where a unique ID is stored in the browser and the TCF consent string is communicated to a server. The cases when this has been observed have been when there is a cookie named



(a) ohmymag cookie paywall initial view

(b) ohmymag cookie paywall second pop-up

Figure 5.2: Cookie paywall from `www.ohmymag.co.uk` (accessed 21. May. 2023). The user is presented with the first view in 5.2a. After selecting "*Subscribe now*", the pop-up in 5.2b appears.

OptanonConsent, and inside this cookie there is a *consentId*. A TCF consent string was then observed in the network traffic of the browser upon accepting tracking.

5.3.3 Didomi token

While manually decompiling TCF consent strings, we found confusing behaviour in the `https://iabtcf.com/#/decode` tool. Upon entering the below string in to the tool, the date seems to be accurate, and it displays that there are 62 registered vendor consents. This turned out that the date is set to the present time if the date in the string could not be read, causing the confusion of whether the base 64 string was a legitimate TCF consent string.

In six of the manually inspected websites we found consent stored in the format of a cookie named "*didomi_token*", the format of this string was base 64. After decoding the string, there was a JSON format with the consent information. All of these six websites did also have a TCF consent string present after accepting or paying, but not initially, and the content of both the *didomi_token* and the TCF consent string did not register any unlawful consent.

5.3.4 Not implemented paywall

There was even one case where we found that the website had not yet implemented the cookie paywall that was displayed. Upon selecting the agreement option, there is another pop-up, notifying the user that this feature has yet to be implemented and that they are welcome to browse free from ads and tracking without paying; this can be seen in Figure 5.2.

6

Discussion

Here in this chapter, we reflect on the results previously presented. In this, we attempt to answer the research questions which are the backbone of this research. We question whether cookie paywalls should be a dark pattern, and then we discuss possible countermeasures to cookie paywalls and unwanted tracking. Finally, we will discuss ethics, limitations, and possible future works to continue on this.

6.1 Results

In this section, we will discuss the results of our study. Firstly, there are some comments on the resulting data of the automatic and semi-automatic crawls. Then we look at the research questions, followed by a deeper dive into the dark patterns.

6.1.1 The resulting data

Looking at the results, most websites seem to behave, and even though we did not manage to analyse all the cookie paywalls in our set, we managed to cover most of them. During the period between the collection of the cookie paywalls and the data collection, some websites switched from implementing a cookie paywall to implementing a cookie banner. No information was posted on the websites about this change, so we do not have any information about why this switch was made. The tool we developed for automatically collecting data was not compatible with all websites, and the decision was made to omit these websites from the dataset. //Some were switched from having cookie paywalls to having cookie consent banners, and some did not work to investigate using our automated tool. Furthermore, there might be some flaws in our data and the tools we used to analyse the cookie paywalls, more on that in section 6.3.

6.1.2 Answering the research questions

Is the data subject not tracked if the subscription is paid? We could look at the results of the TCF consent string to see whether there has been recorded consent on tracking. None of the recorded consents contained positive consent for vendor tracking in the steps before interacting with the cookie paywalls and after paying. However, we noticed that in some cases, there were legitimate interests for user tracking and personalised ads. If a user pays not to get tracked on one of these

websites, then they would still be tracked. Looking at the *contentpass* websites, 167 of them did not register any legitimate interests before interaction, and 160 did not register any legitimate interests after paying. On these websites, the user would not be tracked if they pay the subscription. However, we found websites that were not part of *contentpass* where there was registered consent for personalised ads and tracking in the form of legitimate interests. In these cases, the user would be tracked even after paying not to get tracked.

Is the TCF consent management correctly implemented on cookie paywalls? We believe that the TCF is correctly implemented in most cookie paywalls. Comparing the results of this study with the results of Matte et al. [1], it seems like a larger percentage of cookie paywalls seem to behave compared to the cookie banners. This is, however, only considering the technical implementation as ascertained from the stored TCF string. Other aspects of the TCF were not part of the evaluation for this question.

How much do cookie paywalls use legitimate interest? Legitimate interest is often used by websites, even before any user interaction. However, there is a big difference in the number of vendors with legitimate interests before user interaction and after consent. This is interesting because of the premise that the GDPR specifies legitimate interests as lawful if necessary for the purpose pursued by the controller. In the case of consent given for tracking purposes, since the user has given consent to tracking, we do not believe that there is anything wrong with an exaggerated amount of legitimate interests. Still, it would be interesting to know whether these legitimate interests are indeed legitimate interests or examples like the ones found in section 5.3.

Are all CMPs equal? In the testing, we found that websites implementing the same CMP behaved similarly. However, looking at the results separated by CMP, the websites with the same CMP did not always behave the same. Looking at the *contentpass* webpages, we believed that contentpass acted as the CMP, but it proved that different CMPs are used by different websites. However, they behaved very similarly, except for some outliers. Additionally, looking at the design, there were little to no differences in cookie paywalls implemented by the same CMP, and looking at differences between CMPs there were differences but not many. It is hard to specify if specific differences are due to different CMPs being used, and if there is a difference, if the difference is due to the technical implementation by the CMP. As an example, websites implementing Contentpass use different CMPs, however, the banners use the same design. This shows that popup design can be completely independent of the CMP used. What could be a difference is what CMP's specify as defaults when no preference is submitted.

To what extent are dark patterns present on cookie paywalls? The short answer is, unfortunately, that there are dark patterns present on cookie paywalls as well. Dark patterns exist on cookie paywalls, and patterns such as *pre-ticked boxes* were found, which we did not expect would be present on cookie paywalls due to the take it all or pay idea. On the cookie paywalls, the feeling of being nudged towards a choice was not close to what is experienced when interacting with normal cookie banners. However, there were many websites which fit into at least one of the four categories

we had specified.

6.1.3 Are cookie paywalls a dark pattern?

The results show plenty of websites where dark patterns were used. During the work, the question arose, whether cookie paywalls should be classified as a dark pattern has emerged. Due to the ruling of reasonable fees mentioned in subsection 2.3.2, one might assume that cookie paywalls present a fair offering to the user. However, with 99.9% of users consenting to tracking at contentpass websites, it shows that cookie paywalls greatly increase the consent rate compared to the 55.2% found by Nouwens et al. [8] when they looked at cookie banners.

If cookie paywalls were to be implemented like the example in Figure 5.2, then by the definition of Morel et al. [5], they would be consent banners, but they could hide behind the acceptance of cookie paywalls. Considering the dramatic increase in the consent rate from the consent banner to the cookie paywall, we believe that cookie paywalls are either too expensive or too annoying to navigate.

An assumption, however, has to be made regarding the number of users choosing to exercise the option of not using the website.

6.2 Cookie banner countermeasures

One way to avoid being tracked by cookie paywalls could be by applying countermeasures. In this section, we will discuss what could be done to avoid paying as well as getting tracked. To expand on this research, other than the suggested further work in 6.5, these are areas which we also see as possible further work. The countermeasures are not exclusive to cookie paywalls, and they should be equally applicable to normal cookie banners as well.

6.2.1 Freezing the browser tab

The first countermeasure could be that the tab of the browser is frozen whenever there is a consent string detected which do not correspond to the user's choices. In doing this, it means that there would be no further tracking, and the user can be safe in knowing that, unless unfrozen, there will be no possibility of the website tracking them. This, however, could become very cumbersome, and if there are many websites where there is a violation, then it would become a hassle for the user.

6.2.2 TCF consent string spamming

The second countermeasure involves the TCF consent string. It is specified in the TCF that the latest communicated consent is the one to store, as mentioned in 2.2.1. This means it would be possible to override a consent string with a manufactured and minimal consent string. Two different countermeasures are then proposed, either we could send a minimal consent string whenever there is positive consent registered, or we could set an interval to send a minimal consent string continuously. In the

latter case, if there would be positive consent registered, then it would not take long until it gets overwritten by the manufactured minimal consent string.

6.2.3 Removing the cookie banner

It might be possible to work around the cookie pop-ups without looking at the TCF consent string. Ad blockers remove incoming website requests for ad placements Garimella et al. [39]. Removing the ads in this way reduces the number of ads that can track the user. By developing a similar extension tho Garimella et al. [39], but which instead targets cookie pop-ups, it would be possible to block cookie paywalls. This means that there would be no need to accept tracking, pay not to get tracked, or navigate a menu to find the denial button. However, if there is registered consent before any user action, this method would not work to get rid of that consent. It would only serve to hide the pop-up, and potentially block ads with loaded tracking scripts. Accessing and isolating the pop-up has been done by Nouwens et al. [8] and their development of Consent-O-Matic.

6.2.4 CMP traffic blocking

The last countermeasure that we could think of is to stop any traffic to and from CMPs. Stopping traffic to and from a CMP could be possible by dropping any traffic going to and from the CMPs servers. This would be similar to sending the minimal consent string in that it tampers with the management of the TCF consent string. Blocking any traffic going to and from any CMP means that the TCF consent string cannot be stored and updated, and if there is no consent string at the CMP to begin with, then there would not be any consent communicated to the ad providers.

6.3 Ethics and Limitations

There are several limitations in the collection of user actions and the tracking analysis. Considering these limitations, along with ethics in user tracking and automated website scanning, we will present the reasoning behind some of our choices in this section.

When we looked at what numbers there were after accepting tracking. In our automated and semi-automated analyses of the cookie paywalls, there were several instances where we found that there was no recorded consent after accepting. This might be an indicator that the way in which we collected the recorded consent was not fully working on all webpages. However, the coverage of the TCF consent strings that we did collect is still large enough that it covers most of the set of cookie paywalls.

While collecting the Consent string, as we noticed in the manual analysis, there were different ways to handle consent and sometimes multiple TCF consent strings. We coded the automatic and semi-automatic scripts to search for different consent storage methods. In the cases where there were several ways to store user consent, we do not know which one is the "*correct*" one. Therefore, there might be some

incorrect data due to the wrong string being collected. However, it might be that having multiple stored TCF consent strings violate the guidelines by IAB and might therefore be something worth investigating further.

The manual analysis covered a set of 24 cookie paywalls. This is a small subset of the total set of cookie paywalls which we know of. Therefore, there could be a misrepresentation in the data, and what we found in this set could show to be an edge case of the market.

Cookie walls are bad at conforming to a unified layout [34]. Therefore, a program for automatically checking the implementation of cookie paywalls will be a challenge to implement. What we could do to make this work is to look at cookie paywalls that have a similar layout to each other, but by doing this, we might get a bias in the results as they might be implemented by the same company or individuals. Instead, what we decided was to try to automate a solution for those websites which are similar and manually check for the rest.

An interesting result would be to know exactly who is currently tracking the user. As described in the research by Chen et al. [32], it is possible to see if a cookie set by one vendor is being used by another vendor. To see all the vendors who have access to the user's data could, however, be challenging. If there is any sharing of personal data, which does not occur locally, then we will not be able to see what other vendors have access to that data. Additionally, the website visited could allow some vendors to access the personal data, but their attempts at tracking are prevented by a modern browser or ad blocker.

We decided to look at stored consent in the form of the TCF consent string. Since examining what is tracked compared to the consent string has been investigated by Liu et al. [36]. In addition, whether the websites that implement the TCF follow the guidelines would be fitting for another research topic.

When a cookie is created, this happens through an HTTP request to the target browser [32]. This allows surveying traffic to and from the web browser to identify when cookies are set and when they are accessed. Other forms of tracking, investigated by Papadogiannakis et al. [33], can include fingerprinting and ID sharing. These other methods of tracking cannot be easily detected and will therefore be omitted in this investigation.

6.4 Significance of the study

Previous research on the TCF has seen that it did not comply with the GDPR. Later, the TCF was updated to better comply with the GDPR. There has also been plenty of research that looked at the tracking by online vendors, which could be the reason that modern browsers have been updated to block more tracking. In addition to examining cookie paywalls, the TCF has also been a large part of this study. Both aspects have been thoroughly researched previously. However, there has not been a deep investigation of cookie paywalls. With our results, we show that this area is in much need of more research.

By just looking at the registered consent, we can say that it is common to use legitimate interests in cookie paywalls and some which have legitimate interests which violate the GDPR. We have conducted a thorough analysis of 230 websites, making this investigation a representative examination of the present cookie paywalls. The results obtained from our research partners' crawl indicate that the majority of the cookie paywalls currently found in Europe are included in this dataset. Recording positive consent for personalised ads and tracking, even after payment, could be seen as a new violation due to not complying to the service paid for by the user. Additionally, we have found a set of websites that do not violate the GDPR in the stored consent, these could be seen as safe websites to visit for those who wish to read their news and pay not to be tracked, and they could be interesting for researchers looking to see whether the consent stored in the TCF string is respected.

6.5 Future work

In this research, we have witnessed several things with the cookie paywalls and their consent management. All of which could be part of their own specified research. In this section, we will present what we believe will be some valuable and interesting future work from what we have discovered.

The suggested future work is not all about cookie paywalls, and the TCF is also something that should be investigated. Furthermore, an investigation regarding the countermeasures discussed in section 6.2 and if they are effective. We will first go through the specific cookie paywall-related future work, then what we think could be good future work on the TCF.

6.5.1 Cookie paywalls as dark patterns

Regarding the subject of dark patterns, it would be worthwhile to conduct an investigation into the underlying concept of cookie paywalls. This investigation would aim to determine whether cookie paywalls themselves can be considered as another form of a dark pattern designed to manipulate users into giving up their personal data. By examining the strategies employed by cookie paywalls, the investigation would shed light on their potential alignment with the characteristics of dark patterns and their impact on user privacy.

6.5.2 Cookie paywalls and their actual tracking

Given that websites can be implemented in various ways, it would be valuable to explore the use of cookie paywalls and investigate scenarios where the TCF may be ignored. In such cases, the focus would shift towards analysing the cookies used on a website, identifying the entities responsible for setting those cookies, and determining if any tracking cookies are present. This investigation would provide insights into the practices employed by websites that deviate from or overlook the TCF. Additionally, a tool similar to *CookieGraph* by Munir et al. [23] could be used to see differences in the bidding taking place, and from that, assumptions on tracking could be made.

6.5.3 Further investigation of the TCF

Since our research looked at the TCF consent string, it could be beneficial with more research on whether the new TCF version will be compliant with the GDPR. If there would be discoveries similar to what the GBA found, described in section 3.6, it could point to the TCF needing more changes to be compliant. Additionally, if the TCF is again found to be non-compliant, it would further highlight the need to look at the actual tracking of cookie paywalls. During our research, we have not come across any alternatives to the TCF, but research on other standards than the TCF could be important, or if there even exist alternatives to the TCF.

6.5.4 Dark patterns in cookie paywalls

Investigating the dark patterns present on cookie paywalls would be interesting. Not only by individuals with a legal background, who could make a fair judgement on the grounds of the GDPR, but also by some with a better understanding of user interface design. It would be interesting to see differences between what dark patterns are common on cookie paywalls compared to cookie banners. Also, if there are some new dark patterns which have not been present on traditional cookie banners before.

6.5.5 More cookie paywalls

One possible future work is to look at how the cookie paywalls behave after payment. With the set of 24 paywalls which we paid for, there are plenty more cookie paywalls which we have not been able to look at. What they do after payment, and looking at a larger set, could give a better representation of the state of cookie paywalls. Additionally, near the end of this thesis, we learned that there was a new dataset of 431 cookie paywalls discovered by the group of master's students that we collaborate with. As we did not have time to look at this set, it would be beneficial to have said list of cookie paywalls as part of future looks at cookie paywalls in scale.

7

Conclusion

We have looked at a set of 230 cookie paywalls, paid for access to 24 of them, and got access to a set of 189 of them. There were not many cases where consent was registered for vendor tracking. However, there were an alarming amount of legitimate interests used, some of which were in direct violation of the GDPR and ePD. The design of cookie paywalls also raises the question if a cookie paywall with a consent rate of 99.9% actually should be considered a dark pattern in the world of cookie banners. It is very hard to answer whether users should pay not to be tracked. With some websites relying heavily on legitimate interests, even after paying, and the possibility to limit tracking by means of browser extensions and not pay, it seems to us that the payment is not worth it. Ultimately, the decision to pay cookie paywalls or not should be based on an individual's comfort level with data tracking and their assessment of the value they receive in return for their payment. Continuing this research, it would be interesting to see what tracking is actually taking place instead of looking at registered consent. Also, it would be interesting to see an investigation on whether cookie paywalls should be considered lawful, or if there is anything that could be done to make it less inconvenient for the end-user to opt out of tracking, similar to tools such as *Consent-O-Matic* which exist for normal cookie banners.

Bibliography

[1] C. Matte, N. Bielova, and C. Santos, "Do Cookie Banners Respect my Choice? Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework," Feb. 2020, arXiv:1911.09964 [cs]. [Online]. Available: http://arxiv.org/abs/1911.09964

[2] "IAB Europe about us," https://iabeurope.eu/about-us/, accessed: 2022-12-10.

[3] Council of European Union, "Council regulation (EU) no 2016/679," 2016, https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[4] "Directive 2009/136/EC of the European Parliament and of the Council." [Online]. Available: https://edps.europa.eu/sites/default/files/publication/dir_2009_136_en.pdf

[5] V. Morel, C. Santos, Y. Lintao, and S. Human, "Your Consent Is Worth 75 Euros A Year - Measurement and Lawfulness of Cookie Paywalls," in *Proceedings of the 21st Workshop on Privacy in the Electronic Society*. Los Angeles CA USA: ACM, Nov. 2022, pp. 213–218. [Online]. Available: https://dl.acm.org/doi/10.1145/3559613.3563205

[6] L. Kyi, S. Ammanaghatta, F. Roesner, and C. Santos, "Investigating Deceptive Design in GDPR's Legitimate Interest," 2023.

[7] D. Bollinger, K. Kubicek, C. Cotrini, and D. Basin, "Automating Cookie Consent and GDPR Violation Detection."

[8] M. Nouwens, I. Liccardi, M. Veale, D. Karger, and L. 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*. Honolulu HI USA: ACM, Apr. 2020, pp. 1–13. [Online]. Available: https://dl.acm.org/doi/10.1145/3313831.3376321

[9] J. Kessler, "DATA PROTECTION IN THE WAKE OF THE GDPR: CALIFORNIA'S SOLUTION FOR PROTECTING "THE WORLD'S MOST VALUABLE RESOURCE"," *SOUTHERN CALIFORNIA LAW REVIEW*, vol. 93.

[10] H. Li, L. Yu, and W. He, "The Impact of GDPR on Global Technology Development," *Journal of Global Information Technology Management*, vol. 22, no. 1, pp. 1–6, Jan. 2019. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/1097198X.2019.1569186

[11] G. D. P. R. (GDPR), "Art. 6 gdpr – lawfulness of processing," https://gdpr-info.eu/art-6-gdpr/, accessed: Apr. 13, 2023.

[12] E. D. P. Directive, "Directive 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data," https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31995L0046, accessed: Apr. 25, 2023.

[13] G. D. P. R. (GDPR), "Art. 28 gdpr – processor," art-28-gdpr, accessed: Apr. 13, 2023.

[14] P. Church, "Eu - status of the proposed eprivacy regulation: Tighter cookie rules and more," https://www.linklaters.com/en/insights/publications/tmt-news/tmt-news---june-2017/eu---status-of-the-proposed-eprivacy-regulation-tighter-cookie-rules-and-more, accessed: Apr. 13, 2023.

[15] "Art. 6 GDPR – Lawfulness of processing." [Online]. Available: https://gdpr-info.eu/art-6-gdpr/

[16] "IAB Europe transparency and consent framework," https://iabeurope.eu/transparency-consent-framework/, accessed: Dec. 12, 2022.

[17] "GDPR-Transparency-and-Consent-Framework/TCFv2 at master · InteractiveAdvertisingBureau/GDPR-Transparency-and-Consent-Framework." [Online]. Available: https://github.com/InteractiveAdvertisingBureau/GDPR-Transparency-and-Consent-Framework

[18] "IAB Tech Lab Unveils Global Privacy Platform (GPP) to Consolidate Domestic and Global Privacy Signals for Digital Advertising -." [Online]. Available: https://iabtechlab.com/press-releases/iab-tech-lab-unveils-global-privacy-platform/

[19] I. Europe, "Tcf 2.2 launches! all you need to know," https://iabeurope.eu/all-news/tcf-2-2-launches-all-you-need-to-know/, accessed: May 22, 2023.

[20] W. Ma and H. Xu, "A study of the partnership between advertisers and publishers," in *Passive and Active Measurement - 22nd International Conference, PAM 2021, Virtual Event, March 29 - April 1, 2021, Proceedings*, ser. Lecture Notes in Computer Science, O. Hohlfeld, A. Lutu, and D. Levin, Eds., vol. 12671. Springer, 2021, pp. 564–580. [Online]. Available: https://doi.org/10.1007/978-3-030-72582-2_33

[21] I. europe, "The full tcf workshop webinar recordings – switch over support from tcf v1.1 to tcf v2.0," https://iabeurope.eu/events/tcf-workshop-webinars-switch-over-from-tcf-v1-0-to-tcf-v2-0-support/, accessed: Apr. 13, 2023.

[22] Mozilla, "Window: localstorage property," https://developer.mozilla.org/en-US/docs/Web/API/Window/localStorage, accessed: Apr. 13, 2023.

[23] S. Munir, S. Siby, U. Iqbal, S. Englehardt, Z. Shafiq, and C. Troncoso,

"COOKIEGRAPH: Understanding and Detecting First-Party Tracking Cookies," 2022. [Online]. Available: https://www.semanticscholar.org/reader/97a14cd2f8fb2b50a4ae4db2f514c07c4110ab2b

[24] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz, "The Web Never Forgets: Persistent Tracking Mechanisms in the Wild," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. Scottsdale Arizona USA: ACM, Nov. 2014, pp. 674–689. [Online]. Available: https://dl.acm.org/doi/10.1145/2660267.2660347

[25] L. Olejnik, T. Minh-Dung, and C. Castelluccia, "Selling Off Privacy at Auction," Dec. 2013, working paper or preprint. [Online]. Available: https://hal.inria.fr/hal-00915249

[26] D. (Austria), "Faq zum thema cookies und datenschutz," https://www.dsb.gv.at/download-links/FAQ-zum-Thema-Cookies-und-Datenschutz.html, accessed: Apr. 13, 2023.

[27] ——, "GZ___2023-0.174.027_(pseudonymisierte_kopie)," Mar. 2023. [Online]. Available: https://www.dsb.gv.at/dam/jcr:6608c6c7-8576-4e21-aa71-89073657ab6d/GZ___2023-0.174.027_(pseudonymisierte_Kopie).pdf

[28] CNIL, "Cookie walls : la cnil publie des premiers critères d'évaluation," https://www.cnil.fr/fr/cookies-et-autres-traceurs/regles/cookie-walls/la-cnil-publie-des-premiers-criteres-devaluation, accessed: Apr. 13, 2023.

[29] DATATILSYNET, "Cookie walls," https://www.datatilsynet.dk/hvad-siger-reglerne/vejledning/cookies/cookie-walls, accessed: Apr. 13, 2023.

[30] edpb, "Edpb establishes cookie banner taskforce," https://edpb.europa.eu/news/news/2021/edpb-establishes-cookie-banner-taskforce_en, accessed: Apr. 13, 2023.

[31] edpd, "Report of the work undertaken by the Cookie Banner Taskforce," Jan. 2023.

[32] Q. Chen, P. Ilia, M. Polychronakis, and A. Kapravelos, "Cookie Swap Party: Abusing First-Party Cookies for Web Tracking," in *Proceedings of the Web Conference 2021*. Ljubljana Slovenia: ACM, Apr. 2021, pp. 2117–2129. [Online]. Available: https://dl.acm.org/doi/10.1145/3442381.3449837

[33] E. Papadogiannakis, P. Papadopoulos, N. Kourtellis, and E. P. Markatos, "User Tracking in the Post-cookie Era: How Websites Bypass GDPR Consent to Track Users," in *Proceedings of the Web Conference 2021*. Ljubljana Slovenia: ACM, Apr. 2021, pp. 2130–2141. [Online]. Available: https://dl.acm.org/doi/10.1145/3442381.3450056

[34] P. Hausner and M. Gertz, "Dark Patterns in the Interaction with Cookie Banners," Mar. 2021, arXiv:2103.14956 [cs]. [Online]. Available: http://arxiv.org/abs/2103.14956

[35] M. Myllylahti, "Newspaper Paywalls—the Hype and the Reality," Jul. 2013, iSSN: 2167-0811. [Online]. Available: https://www.tandfonline.com/doi/epdf/10.1080/21670811.2013.813214?needAccess=true&role=button

[36] Z. Liu, U. Iqbal, and N. Saxena, "Opted Out, Yet Tracked: Are Regulations Enough to Protect Your Privacy?" Feb. 2023, arXiv:2202.00885 [cs]. [Online]. Available: http://arxiv.org/abs/2202.00885

[37] M. Veale, M. Nouwens, and C. Santos, "Impossible Asks: Can the Transparency and Consent Framework Ever Authorise Real-Time Bidding After the Belgian DPA Decision?" *Technology and Regulation*, vol. 2022, pp. 12–22, Feb. 2022. [Online]. Available: https://techreg.org/article/view/11594

[38] Google, "Chromium," https://www.chromium.org/Home/, accessed: Apr. 25, 2023.

[39] K. Garimella, O. Kostakis, and M. Mathioudakis, "Ad-blocking: A study on performance, privacy and counter-measures," in *Proceedings of the 2017 ACM on Web Science Conference*, ser. WebSci '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 259–262. [Online]. Available: https://doi.org/10.1145/3091478.3091514

A

Cookie recipes

Chocolate Nut Cookies

Simple but delicious!

Ingredients:

- 100g Butter
- 1dl Molasses or brown sugar
- 1/2 tsp Vanilla sugar
- 1 Egg
- 2dl Flour
- 1 tsp Baking soda
- 100g Dark chocolate
- 50g Nuts (Hazelnut recommended)
- pinch of salt

Instructions:

1. Preheat the oven to 200°C.
2. Mix the butter and sugars until combined.
3. Coarsely chop the chocolate and the nuts and add them to the mix.
4. Put several golf ball-sized chunks of the mix onto a plate and put them in the oven, remembering to leave space for the cookies to spread. Wait for about 10 minutes.
5. Let the cookies cool before serving.

Toscakaka

Toscakaka, a Swedish cake, perfect for a cup of coffee. Begin with the cake, then add the glaze.

Cake Ingredients:

- 100g Butter
- 2 Eggs
- 1 1/2 dl Sugar
- 120g Flour
- 1 tsp Baking soda
- 1/2 dl Milk

Cake Instructions:

1. Preheat the oven to 175°C.
2. Clamp a piece of baking paper in a mold with a removable edge, about 24 cm in diameter (for 10 pieces), or grease the mold with cooking fat.
3. Melt the butter and let it cool.
4. Beat the eggs and sugar until fluffy.
5. Mix the flour and baking soda.
6. Add the flour mixture, butter, and milk to the batter.
7. Pour the batter into the mold.
8. Bake the cake in the lower part of the oven for 20-25 minutes. In the meantime, prepare the glaze.

Glaze Ingredients:

- 100g Butter
- 1 dl Sugar
- 2 tbsp Flour
- 2 tbsp Milk
- 100g Peeled and shredded almond

Glaze Instructions:

1. Mix all the glaze ingredients in a saucepan.
2. Heat while gently stirring until everything is mixed and thickened.
3. Pour the glaze over the cake and spread it out.
4. Move the rack to the center of the oven.
5. Bake for another 15 minutes or until the glaze is colored.
6. Allow the cake to cool before serving.

B

Websites

Note: for websites were no TCF string was collected NaN is displayed.

	Before interaction		After consent		After login	
	Vendors	Legitimate interest	Vendors	Legitimate interest	Vendors	Legitimate interest
0180.info	0	0	0	0	0	0
50plus.de	0	0	0	0	0	0
additive-manufacturing-industry.de	0	0	3	3	0	0
aerztezeitung.de	0	0	0	0	0	0
allesbeste.de	0	0	0	0	0	0
analog-praxis.de	0	0	26	26	NaN	NaN
anime2you.de	0	0	0	0	0	0
aufunsere.art	0	0	0	0	0	0
auszeit.bio	0	1	1	1	0	1
autlook.at	0	0	0	0	0	0
autoflotte.de	0	0	199	199	0	0
autofrage.net	0	1	1	1	0	0
autoguru.de	0	0	0	0	0	0
autohaus.de	0	0	197	197	0	0

	Before interaction		After consent		After login	
	Vendors	Legitimate interest	Vendors	Legitimate interest	Vendors	Legitimate interest
automobil-industrie.vogel.de	0	0	26	26	0	0
autoservice.co.at	0	0	0	0	0	0
autoservicepraxis.de	0	0	199	199	0	0
backenmachtgluecklich.de	0	0	0	0	0	0
bahnblogstelle.com	0	0	1	1	0	0
basic-tutorials.de	0	0	0	0	0	0
baugewerbe-magazin.de	0	0	3	3	0	0
bba-online.de	0	0	0	0	0	0
bigdata-insider.de	0	0	26	26	NaN	NaN
blechnet.com	0	0	26	26	0	0
boomer.at	0	0	0	0	0	0
braunschweiger-zeitung.de	0	0	134	134	0	1
calconi.com	0	0	0	0	0	0
christliche-partnersuche.de	0	1	359	359	NaN	NaN
climate-data.org	0	0	0	0	0	0
cloudcomputing-insider.de	0	0	26	26	0	0
computer-automation.de	0	0	18	18	0	0
computerfrage.net	0	1	1	1	0	0
connect-professional.de	0	0	16	16	0	0
connect.de	0	1	332	332	0	1
das-pta-magazin.de	0	0	0	0	0	0
datacenter-insider.de	0	0	26	26	0	0
derwesten.de	0	0	134	134	NaN	NaN
dev-insider.de	0	0	26	26	0	0
devicemed.de	0	0	26	26	0	0
die-frau-am-grill.de	0	0	0	0	0	0
dispo.cc	0	0	0	0	0	0

	Before interaction		After consent		After login	
	Vendors	Legitimate interest	Vendors	Legitimate interest	Vendors	Legitimate interest
donnerwetter.de	0	0	0	0	0	0
dumontreise.de	0	3	3	3	0	3
eat.de	0	0	0	0	0	0
egovernment-computing.de	0	0	26	26	NaN	NaN
elektroniknet.de	0	0	21	21	0	0
elektropraxis.at	0	0	0	0	0	0
elektrotechnik.vogel.de	0	0	26	26	NaN	NaN
embedded-software-engineering.de	0	0	26	26	0	0
emtb-news.de	0	0	0	0	0	0
en.kunststoffe.de	0	0	0	0	0	0
erbrecht-ratgeber.de	0	0	0	0	0	0
erklaer-es-oma.de	0	1	360	360	NaN	NaN
esports.com	0	0	0	0	0	0
etmm-online.com	0	0	26	26	0	0
factorynet.at	0	0	0	0	0	0
fahrschule-online.de	0	0	199	199	0	0
falk.de	0	3	3	3	0	3
fernsehserien.de	0	0	0	0	0	0
filext.com	0	0	0	0	0	0
finanzfrage.net	0	1	1	1	0	0
firmenwagen.co.at	0	0	0	0	0	0
form-werkzeug.de	0	0	0	0	0	0
formel1.de	0	0	128	128	0	0
forum.clubalfa.it	0	0	0	0	0	0
fraenkische-rezepte.de	0	4	302	302	0	4
free-fonts.com	0	0	0	0	0	0
freenet.de	0	0	0	0	0	0

	Before interaction		After consent		After login	
	Vendors	Legitimate interest	Vendors	Legitimate interest	Vendors	Legitimate interest
funandnews.de	0	0	0	0	0	0
fussballfieber.de	0	0	0	0	0	0
gabler-banklexikon.de	0	0	0	0	0	0
gartendialog.de	0	0	0	0	0	0
gartenjournal.net	0	0	0	0	0	0
gartenlexikon.de	0	0	0	0	0	0
gesund-verlieben.de	0	4	302	302	0	4
gesundheitsfrage.net	0	1	1	1	0	0
gifhorner-rundschau.de	0	0	134	134	0	1
goslarsche.de	0	0	0	0	0	0
gutefrage.net	0	1	1	1	0	0
gutekueche.at	0	5	5	5	0	5
gutekueche.de	0	2	2	2	0	2
hallo-eltern.de	0	0	0	0	0	0
hanser-automotive.de	0	0	357	357	0	0
harzkurier.de	0	0	134	134	0	1
hasepost.de	0	0	0	0	0	0
haus-garten-test.de	0	1	1	1	0	1
hausgarten.net	0	0	0	0	0	0
healthcare-computing.de	0	0	26	26	0	0
helmstedter-nachrichten.de	0	0	134	134	0	1
hildesheimer-allgemeine.de	0	0	0	0	0	0
hlk.co.at	0	0	0	0	0	0
hoerzu.de	0	0	134	134	0	0
hortica.de	0	0	0	0	0	0
iban-rechner.de	0	1	1	1	0	1
ibancalculator.com	0	1	1	1	0	1

	Before interaction		After consent		After login	
	Vendors	Legitimate interest	Vendors	Legitimate interest	Vendors	Legitimate interest
industrial-production.de	0	0	3	3	0	0
industriemagazin.at	0	0	0	0	0	0
industry-of-things.de	0	0	26	26	0	0
infranken.de	0	3	234	234	0	3
inrlp.de	0	5	298	298	0	5
inside-digital.de	0	0	0	0	0	0
ip-insider.de	0	0	26	26	0	0
it-business.de	0	0	26	26	0	0
it-daily.net	0	0	0	0	0	0
juraforum.de	0	0	0	0	0	0
karrierefragen.de	0	73	73	73	0	73
kem.industrie.de	0	0	0	0	0	0
klack.de	0	0	134	134	NaN	NaN
korrekturen.de	0	12	12	12	0	12
kunststoff-magazin.de	0	0	3	3	0	0
kunststoffe.de	0	0	357	357	0	0
lab-worldwide.com	0	0	26	26	0	0
labo.de	0	0	3	3	0	0
laborpraxis.vogel.de	0	0	26	26	0	0
lanline.de	0	0	18	18	0	0
likehifi.de	0	1	1	1	0	1
linux-community.de	0	0	0	0	0	0
linux-magazin.de	0	0	0	0	0	0
logistik-heute.de	0	1	116	116	NaN	NaN
logistra.de	0	1	116	116	0	1
lonelyplanet.de	0	3	3	3	0	3
macwelt.de	0	0	0	0	0	0

	Before interaction		After consent		After login	
	Vendors	Legitimate interest	Vendors	Legitimate interest	Vendors	Legitimate interest
maennerseite.net	0	0	0	0	0	0
marconomy.de	0	0	26	26	0	0
marcopolo.de	0	3	3	3	0	3
maschinenmarkt.ch	0	0	26	26	0	0
maschinenmarkt.international	0	0	26	26	0	0
maschinenmarkt.vogel.de	0	0	26	26	NaN	NaN
materialfluss.de	0	0	3	3	0	0
medical-design.news	0	0	18	18	0	0
meineorte.com	0	0	0	0	0	0
mission-additive.de	0	0	26	26	0	0
mm-logistik.vogel.de	0	0	26	26	0	0
motorradfrage.net	0	1	1	1	0	0
motorsport-total.com	0	0	128	128	0	0
mtb-news.de	0	0	0	0	0	0
netzwelt.de	0	0	333	333	0	0
next-mobility.de	0	0	26	26	0	0
oekotest.de	0	0	0	0	0	0
omnibusrevue.de	0	0	0	0	0	0
pc-magazin.de	0	2	335	335	0	2
pcwelt.de	0	0	0	0	0	0
peiner-nachrichten.de	0	0	134	134	0	1
plantopedia.de	0	0	0	0	0	0
prisma.de	0	0	303	304	0	0
process-worldwide.com	0	0	26	26	0	0
process.vogel.de	0	0	26	26	NaN	NaN
profi-werkstatt.net	0	1	117	117	NaN	NaN
qz-online.de	0	0	357	357	0	0

	Before interaction		After consent		After login	
	Vendors	Legitimate interest	Vendors	Legitimate interest	Vendors	Legitimate interest
radsport-news.com	0	1	200	200	0	1
raspberry-pi-geek.de	0	0	0	0	0	0
reisefrage.net	0	1	1	1	0	0
rennrad-news.de	0	0	0	0	0	0
saechsische.de	0	0	0	0	0	0
schifffahrtundtechnik.de	0	0	0	0	0	0
schulferien.org	0	0	0	0	0	0
security-insider.de	0	0	26	26	0	0
selbstversorger.de	0	0	0	0	0	0
selfies.com	0	0	134	134	NaN	NaN
smart-rechner.de	0	0	0	0	0	0
smarterworld.de	0	0	21	21	0	0
smarthouse-pro.de	0	0	20	20	0	0
solidbau.at	0	0	0	0	0	0
spielfilm.de	0	1	361	361	0	1
sportlerfrage.net	0	1	1	1	0	0
springermedizin.de	0	0	0	0	0	0
springerpflege.de	0	0	0	0	0	0
springerprofessional.de	0	0	0	0	0	0
sprit-plus.de	0	0	200	200	0	0
stimme.de	0	0	0	0	0	0
storage-insider.de	0	0	26	26	0	0
stylevamp.de	0	0	0	0	0	0
sudoku-aktuell.de	0	0	0	0	0	0
sudoku-topical.com	0	0	0	0	0	0
tageblatt.de	0	0	0	0	0	0
tagesspiegel.de	0	0	0	0	0	0

	Before interaction		After consent		After login	
	Vendors	Legitimate interest	Vendors	Legitimate interest	Vendors	Legitimate interest
talu.de	0	0	0	0	0	0
teltarif.de	0	0	0	0	0	0
teslamag.de	0	0	0	0	0	0
tga.at	0	0	0	0	0	0
tierfans.net	0	0	0	0	0	0
tomaten.de	0	0	0	0	0	0
traktuell.at	0	0	0	0	0	0
trucker.de	0	0	199	199	0	0
tvdigital.de	0	0	134	134	0	0
tvdirekt.de	0	0	134	134	NaN	NaN
unnuetzes.com	0	0	0	0	0	0
unsere-helden.com	0	0	0	0	0	0
unterwegs-auf-der-autobahn.de	0	1	113	113	0	1
utopia.de	0	0	0	0	0	0
verkehrsrundschau.de	0	0	199	199	0	0
versicherungsmagazin.de	0	0	201	201	0	0
vision-mobility.de	0	1	116	116	NaN	NaN
watson.de	0	0	0	0	0	0
we-go-wild.com	0	0	0	0	0	0
weihnachtsmarkt-deutschland.de	0	0	0	0	0	0
werkstatt-betrieb.de	0	0	0	0	0	0
werstreamt.es	0	0	134	134	0	0
wetter.com	0	0	0	0	0	0
winfuture.de	0	0	0	0	0	0
wir-in-der-praxis.de	0	0	201	201	0	0
wirtschaftslexikon.gabler.de	0	0	0	0	0	0
wunderbunt.de	0	0	0	0	0	0

	Before interaction		After consent		After login	
	Vendors	Legitimate interest	Vendors	Legitimate interest	Vendors	Legitimate interest
wunschliste.de	0	0	0	0	0	0

	Before interaction		After consent	
	Vendors	Legitimate interest	Vendors	Legitimate interest
vienna.at	0	0	356	356
vol.at	0	0	356	356
kurier.at	0	0	0	0
krone.at	0	0	0	0
derstandard.at	0	17	17	17
lachainemeteo.com	0	0	0	0
lemonde.fr	0	0	0	0
allocine.fr	0	0	0	0
sciencesetavenir.fr	0	0	0	0
huffingtonpost.fr	0	0	0	0
capital.fr	0	96	96	96
le10sport.com	0	0	0	0
geo.fr	0	96	96	96
voici.fr	0	96	96	96
gala.fr	0	96	96	96
femmeactuelle.fr	0	96	96	96
programme-tv.net	0	96	96	96
eklablog.com	0	0	0	0
liberation.fr	0	173	173	173
jeuxvideo.com	0	0	0	0
over-blog.com	0	0	0	0
faz.net	0	4	4	4
spiegel.de	0	2	2	2
bild.de	0	0	0	0
t-online.de	0	0	0	0
welt.de	0	0	0	0
zeit.de	0	0	0	0

	Before interaction		After consent	
	Vendors	Legitimate interest	Vendors	Legitimate interest
heise.de	0	0	0	0
stern.de	0	0	0	0
rp-online.de	0	0	0	0
computerbild.de	0	0	0	0
wunderweib.de	0	0	0	0
praxisvita.de	0	0	0	0
cosmopolitan.de	0	0	0	0
maennersache.de	0	0	0	0
bravo.de	0	0	0	0
liebenswert-magazin.de	0	0	0	0
express.de	0	0	0	0
swp.de	0	6	6	6
gamestar.de	0	0	0	0
rheinpfalz.de	0	73	73	73
volksstimme.de	0	25	25	25
capital.de	0	0	0	0
pcgames.de	0	1	1	1
kino.de	0	0	0	0
gala.de	0	0	0	0
wallstreet-online.de	0	23	23	23
auto-motor-und-sport.de	0	0	0	0
computerbase.de	0	0	0	0
brigitte.de	0	0	0	0
geo.de	0	0	0	0
nordbayern.de	0	4	343	343
bz-berlin.de	0	0	0	0
autobild.de	0	0	0	0

	Before interaction		After consent	
	Vendors	Legitimate interest	Vendors	Legitimate interest
t3n.de	0	0	1	1
giga.de	0	0	0	0
stiften.dk	0	0	0	0
fyens.dk	0	0	0	0
jv.dk	0	0	0	0
hsfo.dk	0	0	0	0
amtsavisen.dk	0	0	0	0
frdb.dk	0	0	0	0
viborg-folkeblad.dk	0	0	0	0
dagbladet-holstebro-struer.dk	0	0	0	0
vafo.dk	0	0	0	0
gazzetta.it	0	0	0	0
ansa.it	0	0	0	0

	Before interaction		After consent		After login	
	Vendors	Legitimate interest	Vendors	Legitimate interest	Vendors	Legitimate interest
huffingtonpost.fr	0	0	336	336	0	0
le10sport.com	0	0	345	345	0	0
gentside.com	0	0	96	96	0	0
gala.fr	0	96	96	96	0	0
voici.fr	0	96	96	96	0	0
cesoirtv.com	0	96	96	96	0	0
caminteresse.fr	0	96	96	96	0	0
ohmymag.com	0	96	96	96	0	0
programme-tv.net	0	96	96	96	0	0
capital.fr	0	96	96	96	0	0
geo.fr	0	96	96	96	0	0
neonmag.fr	0	96	96	96	0	0
femmeactuelle.fr	0	96	96	96	0	0
ohmymag.co.uk	0	96	96	96	0	96

C

Cookie paywalls found

Here are examples of cookie paywall implementations that we have found in the wild. Notable deviations are pointed out in the description.

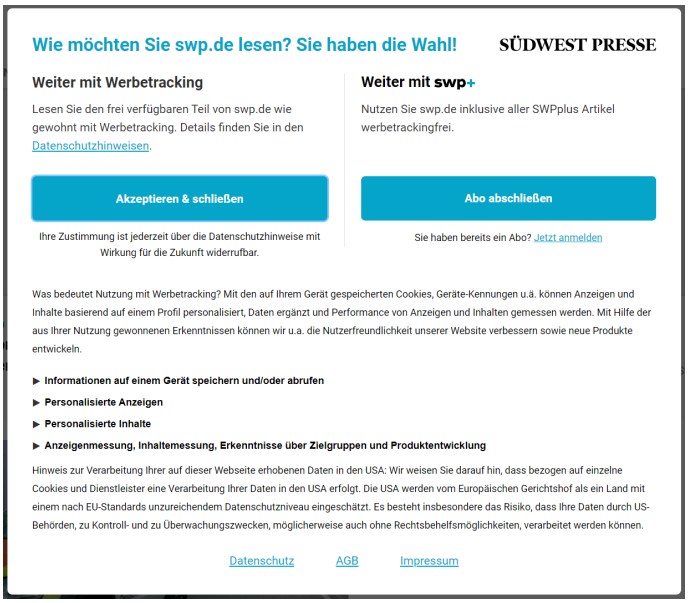


Figure C.1: Cookie paywall from `http://swp.de/` (accessed 21. May. 2023).

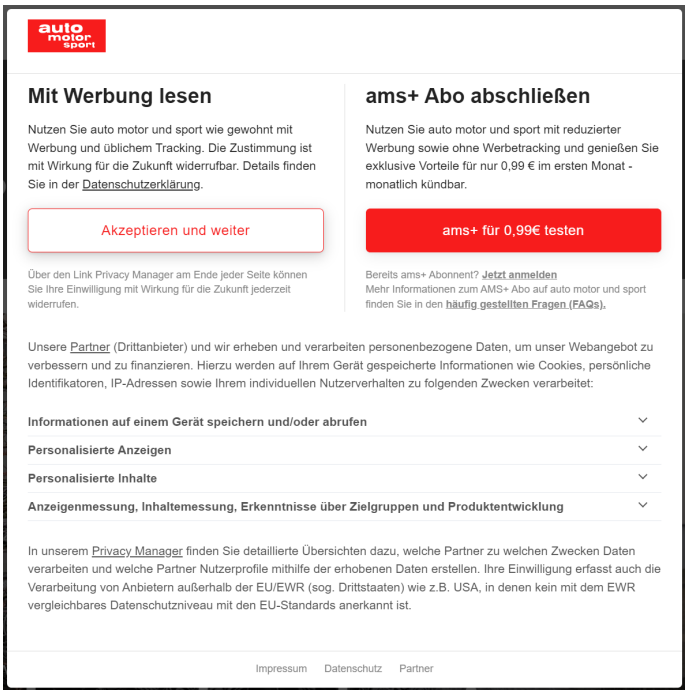


Figure C.2: Cookie paywall from `http://automotorsport.de/` (accessed 21. May. 2023).

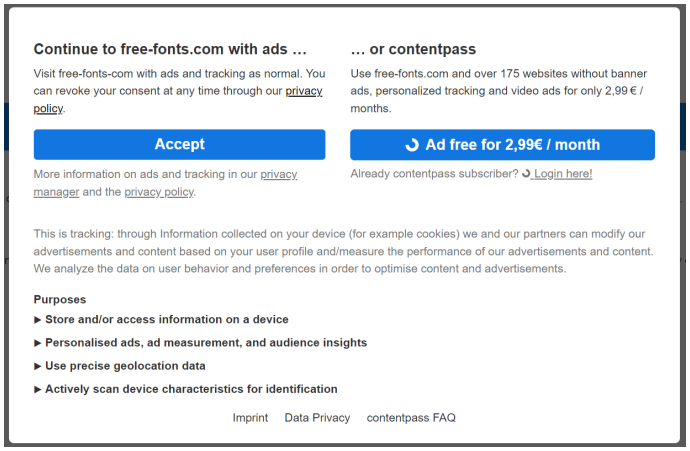
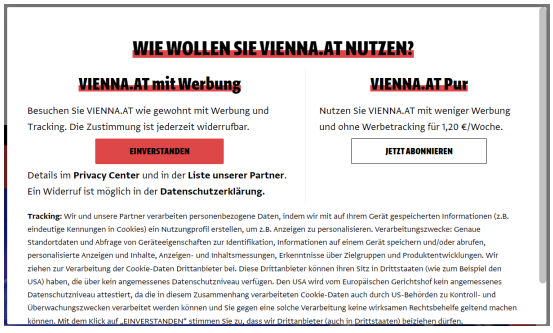
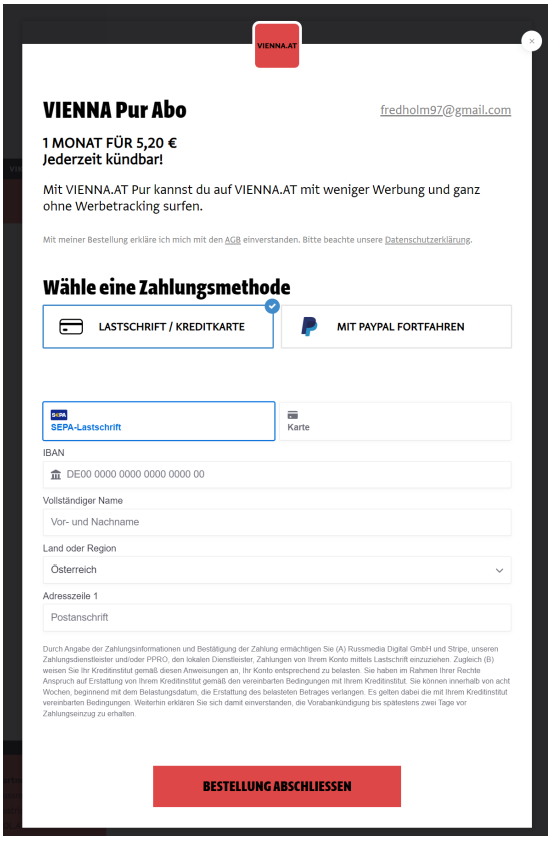


Figure C.3: Cookie paywall from `http://free-fonts.com/` (accessed 21. May. 2023). Note: This is one of the contentpass connected webpages, the design is the same for all contentpass connected webpages

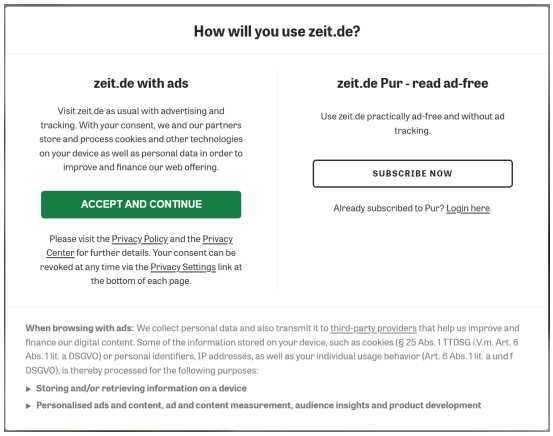


(a) Vienna cookie paywall view 1

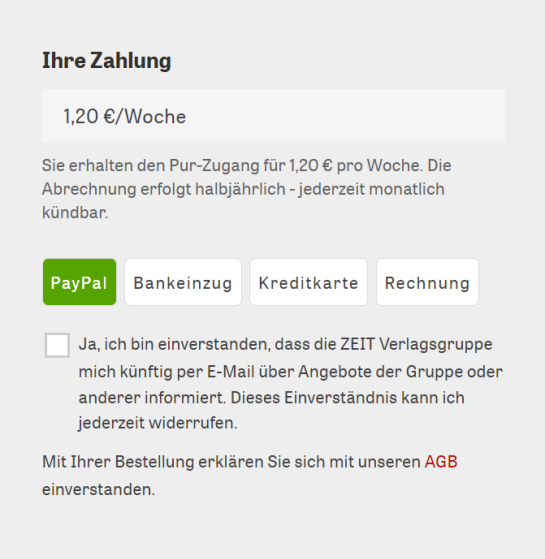


(b) Vienna cookie paywall, payment page

Figure C.4: Cookie paywall from `http://vienna.at/` (accessed 21. May. 2023). The user is presented with the first view in C.4a. After selecting '*JETZT ABONNIEREN' (subscribe now)*, page C.4b appears. Note: The lowest subscription amount changes from €1,2/week to €5,2/month.

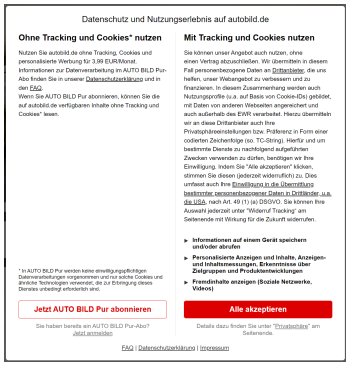


(a) zeit cookie paywall view 1

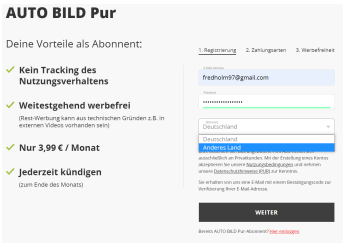


(b) zeit cookie paywall, payment page

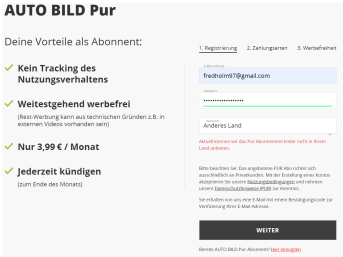
Figure C.5: Cookie paywall from `http://zeit.de/` (accessed 21. May. 2023). The user is presented with the first view in C.5a. After selecting '*SUBSCRIBE NOW*', page C.5b appears. Note: The lowest subscription amount is €1,2/week, but it is only possible to pay monthly, leading to €4,8/month.



(a) Autobild cookie paywall, View 1



(b) Autobild registration and country select



(c) Denying non german payment

Figure C.6: Cookie paywall from `http://autobild.de/` (accessed 21. May. 2023). The user is presented with the first view in C.6a. After selecting '*Jetzt AUTO BILD Pur abonnieren' (Subscribe to AUTO BILD Pur now)*, page C.6b appears. Note: It is not possible to continue to payment if user is not in Germany, see C.6c.

(a) Gazetta cookie paywall, View 1

(b) Gazetta view 2

(c) Gazzetta payment screen

Figure C.7: Cookie paywall from `http://Gazetta.it/` (accessed 21. May. 2023). The user is presented with the first view in C.7a. After selecting '*Rifiuta e abbonati' (Reject and subscribe)*, page C.7b appears. Selecting to access consentless leads to C.7c. Note: refusing consent is only available as an addition to the subscription, and it is not default, users need to opt in to withdraw consent.