Master Thesis

Automatically detect Dark patterns in cookie banners

Rémi Coudert

Supervisors

Prof Claude Castelluccia
PRIVATICS, INRIA Grenoble

Prof Carmela Troncoso
SPRING Lab, EPFL

SPRING Lab, Computer Science Department
EPFL, Switzerland
August 16, 2020

Contents

1 Introduction

Manipulation of human emotions and biases for their own good is used at a larger scale than one could imagine. For example, in 2017 France adopted an opt-out policy regarding organ donation. That is, french citizens are now by default organ donors and need to ask health authorities to no longer be one. This decision will arguably improve the well being of citizens at large by increasing the amount of donors. Because of the default effect bias, the previous system prevented individuals from becoming donors even though in an opt-out system they are less likely to refuse to be donors [1, 2, 3, 4]. This system however raises ethical questions [5].

This kind of manipulation for 'good' is called a Nudge. The term was popularized by Thaler and Sunstein in their book : "Nudge: Improving Decisions About Health, Wealth, and Happiness" [6]. In a later work, Thaler introduced Sludges[7], the inverse concept. A Sludge is a manipulation of human biases that encourages self-defeating behaviour or discourages behaviour that is in the best interest of an individual. Some (most) types of advertisements can be described as sludges, as well as unnecessary complex administrative loads required for important tasks (e.g. tax filling).

Thaler advocates for the use of Nudges and against the use of Sludges even though others disagree with the manipulation all-together. But beyond ethical and philosophical questions that arise when studying biases and Nudges, manipulation of users without them noticing is a well known and widely spread phenomenon even though it happens sometimes with no intend to do so from the individual or organization issuing the Nudge/Sludge.

Dark patterns are a similar concept to Sludges and its definition is sometimes a bit different depending on authors. However, it is Harry Brignull that first coined the term in a blog post [8] warning against "dirty tricks designers use to make people do stuff". Brignull describes Dark patterns as UX designs that are made specifically to manipulate users to take decisions they would not otherwise or to prevent them from it, with some similarity with Sludges.

We describe in section 2.2 a more in depth discussion of the different definitions to prevent confusion.

This Master Thesis focuses on Dark patterns, and specifically on Cookie banner bearing Dark patterns. This work is the result of a 6 month long work at INRIA Grenoble in the PRIVATICS team under the supervision of Claude Castelluccia (INRIA) and Carmela Troncoso (EPFL).

2 Definitions

2.1 Cognitive biases

In behavioural economics, we often make the distinction between decisions made by the System 1 and the System 2, terms coined by Kahneman in his work [9]. He describes both systems as two different ways for the brain to think and make decisions.

When using the System 1, the brain makes fast, unconscious decisions, often based on heuristics and previous experience. We use this system when doing repetitive work, or when taking decisions on problems that seem easy. The System 2 on the other hand, is used by the brain to make, complex, slow conscious and calculated decisions. We typically use System 2 when solving complex problems, computing taxes, focus on specific sounds etc...

Our brain is nonetheless subject to Cognitive biases. This typically happen in situations where we will try to use our System 1 when the System 2 should be used. There exist long lists of cognitive biases, and we can cite several.

Anchoring for example, happens when we tend to overestimate or underestimate numbers based on previous irrelevant ones.

Framing happens when we take different decisions based on the context, for example when using positive wording to present a situation, we can nudge individuals to change their likelihood of taking a decision based on this situation.

The default bias is the tendency of the human brain to stick with the default choice when given multiple ones. Many online contracts or notices have preselected check-boxes, and the human brain will tend to let it selected, and potentially signing up for services users didn't intend to.

Being aware of the existence of these biases can help us make better decisions in our everyday life. However malicious (or unknowing) service providers could exploit human biases to make users take decisions that go against their best interest.

2.2 Nudges, Sludges and Dark patterns

While studying papers to find out the main differences between their usage, it appeared that the definitions of Dark Pattern and Sludge were not always the same. In order for this work to stay clear, I will use definitions slightly adapted from Soman et. al. work [10], depicted in Table 1 below. This table makes the distinction between manipulations that help or harm users and facilitate or impede decision making.

This leaves us with four specific types of user manipulation.

	Facilitate decision making	Impede decision making
Helps	Nudge: Making things easy for end users	Decision points, cooling off periods: prompt vigilance and thoughtfulness
Harms	Nudge-for-bad: Easy to choose welfare reducing options	Sludge: Difficulty to cancel subscriptions, change privacy settings etc...

Table 1: Different types of manipulations designed to help or harm individuals

Nudges ("Coup de pouce" in french) aim at helping users by making it easier for them to take decisions that have a beneficial impact on them. For example, the opt-out organ donor policy discussed in the introduction is a Nudge.

Nudge-for-bad are the inverse of Nudges, and are designed to make users take decisions that harms them or are self-defeating. Countdowns displayed on many e-shopping websites count as Nudges-for-bad as they create a sentiment of urgency making customers more likely to buy items they don't want or need. These countdowns often are not even counting down to anything real and are mostly random [11].

Sludges ("Enlisement" in french) impede decision making to discourage behaviour that are in the best interest of users. The concept of roach motel described by Brignull is a Sludge [12]. Premium subscriptions, such as the New York Times subscription[13], that are hard to opt-out of are an example of Sludges (and roach motel).

Decision points and cooling off periods are not discussed in this thesis. These two techniques are designed for users to take their time before making decisions to help them. It is often used to help binge buyers [14], and promote the use of the System 2 by asking individuals to sleep on it for example.

Originally, Soman et. al. classified Dark patterns with Nudge-for-bad (Facilitate decision making and harms users) but Brignull uses a wider definition that includes Sludges in it as well.

We will use the terms Nudge, Sludge and Nudge-for-bad according to the definitions above, and we will be using the term Dark patterns when describing UX designs that harm users regardless of whether it impedes or facilitate decision making. That is, we classify Dark patterns in the bottom half of Table 1 but use the term when talking specifically about user interfaces.

3 Cookie banners, GDPR and the ePrivacy directive

Due to the General Data Protection Regulation (GDPR) [15] and the ePrivacy directive (ePD) [16], website owners are now required to ask for users consent to collect their personal data when they are EU based and to do so in a clear manner.

Cookie banners are the most common way for website owners to ask for this consent. They can take several forms as each website has its own banner, and warn users that their data is or will be collected. The law-fullness of banners has been widely discussed [17, 18, 19], some banners do not provide explicit mean of consent from users, which is illegal under EU directives. In a recent paper Matt et. al. [20] showed that $\approx$10% of banners store consent of users before asking them, and $\approx$ 1.5% even fail to stop collecting data when explicitly told not to. They found out hat 54% of websites had at least one violation of consent, as stated by the ePD.

Appendix B showcases several examples of cookie banners encountered on the internet. Some of them have an explicit accept and decline button or/and fine grained options whereas some only notice users that their data is being collected by continuing on the website.

The Ronaldo7 and Tedibear websites in Appending B Figures 3e and 3i are good examples of what a website provider should aim at displaying in their banners, with clear consent and decline possibilities, no (or very few) hierarchy of clickable elements and clear wording, all on the very first layer of the banner with no need to click on a link to access further features of the banner.

To help website administrators, the Interactive Advertising Bureau (IAB Europe) published the Transparency and Consent Framework [21] to have a common ground that administrators could rely on to prevent the implementation of a cookie banner framework for every website. They work with *Consent Management Providers* (CMPs), actors making the link between users and advertisers.

4 Related Works

4.1 Nudges, Sludges and manipulation

The literature on the impact of manipulation on individuals' behaviours is extensive. Matz et. al. [22] for example, studied the impact of psychological targeting for persuasion. They published several fake Facebook ads designed for specific psychological profiles and found that it is possible to improve ads conversion score with this kind of targeting. Ali et. al. [23] studied the impact of political ads on users.

Thaler [6] was the first to use the terms of Nudges and Sludges and they have been used by researchers since then. Acquisiti et. al. [24] studied ways to use

Nudges (or 'soft paternalistic interventions') to help users make better privacy and security choices for themselves. They present several Nudging dimensions: Information, Presentation, Defaults, Incentives, Reversibility and Timing and argue that they should be used as follows:

- Information: Reduce information asymmetry
- Presentation: Reduce cognitive load on interfaces
- Defaults: Configure default choices according to users' expectations
- Incentives: Motivate users
- Reversability: Prevent mistakes
- Timing: Define the right moment to Nudge

Soman et. al. [10] studied nudges, sludges and Dark patterns and devised the taxonomy we describe in section 2.2.

Defaults have been discussed by Johnson et. al. [25] by comparing opt-in and opt-out choices. They demonstrate the power of the default bias and how easy it is to exploit it with carefully designed interfaces.

In their work, Moser et. al. [14] describe manipulation techniques used by shopping websites in particular and mechanisms that rule individuals to make them impulse buy.

Sunstein published some work regarding Sludges [26, 27] with very precise definitions and examples. In particular, he insists on the links between behavioural biases and Sludges.

4.2 Dark patterns

Dark patterns have first been desribed under that name by Brignull[8] for which he dedicates a website [12]. According to this website, Dark patterns are 'tricks used in websites and apps that make you do things that you didn't mean to'. The term is mostly used in UX design but is often applied to any technique that nudges users in a direction detrimental to them. You can refer to section 2.2 for a more in depth discussion of the different definitions that exist.

Luguri et. al. [28] describe a taxonomy of dark patterns based on previous ones separated in 7 categories each containing one or more variants. The nagging dark pattern for example refers to repeated requests from firms, the roach motel dark pattern consists in an asymmetry between signing up and canceling to offers whereas privacy zuckering is the act of tricking consumers into sharing personal information.

Mathur et. al. [11] showed that at least one dark pattern is present on almost every e-commerce website and that some exhibit several of them. Moser et. al [14] demonstrate a similar result and highlight the potent effects of dark patterns on impulse buying.

Dark patterns are persuasive and effective. Luguri et. al. set up an experiment to test their potency by designing a questionnaire that contains dark patterns. At the end of the questionnaire, participants to the experiment are told that they have been signed up to an offer and that they can opt out of it at any time. The amount of effort required to opt out is what makes it a dark pattern or not. They used three different version of this questionnaire, a control, one containing "mild dark patterns" and one with "aggressive dark patterns". They found out that the presence of mild dark patterns doubles the odds of staying signed up to the offer whereas it quadruples it for the aggressive version. The authors point out that the aggressive dark patterns create backlash from participants and that the mild one does not. This makes mild dark patterns particularly pernicious as they are both effective and not too upsetting for users.

Overall, Dark Patterns are mechanisms using cognitive biases of online users against their best interest. In the specific topic of privacy and privacy laws, it is used to nudge them towards accepting privacy invading settings they would not accept if given full objective disclosure. They lie on a grey space regarding to law, when not totally illegal, hence the importance to detect them.

In their 2018 report [29], the Norwegian Consumer Council describe several types of Dark patterns with numerous examples and demonstrate their influence.

Conti et. al. [30] published their work on malicious interfaces and present a detailed table with a taxonomy of malicious designs. For example *Coercion* can be found in mandatory form fields or *Obfuscation* can happen with low contract color schemes or partially hidden information. Using this taxonomy, they assessed the impact of each category on users. In particular they measured user frustration coming from this malicious interfaces.

Chivukula et. al. [31] looked at the "/r/assholedesign" subreddit, an internet forum where users share bad or malicious designs they encountered on the internet or in life (see [13] for an example). They present several types of design and conclude that a subset of them are indeed Dark patterns.

Westin et. al. [32] studied the impact of Dark patterns in users privacy behaviour, and how these patterns can promote self sabotaging behaviour in term of privacy.

4.3 Cookie banners

The literature on Cookie banners, and in particular Dark patterns in cookie banners is more recent and less extensive that on cognitive biases or Dark patterns.

In a technical report for the european commission, Van Bavel et. al. [33] showed the effect of different types of messages banners on users' behaviour and whether they accept cookie or not.

Degeling et. al. [34] studied the impact of the GDPR on cookie banners. They monitored EU websites to see whether they changed their privacty policies and consent forms (cookie banners) and conclude that the GDRP indeed had some impact with 60% of their dataset exhibiting cookie consent banners.

The effect of nudging on user interfaces has been the subject of several studies [18, 17] as well as their law-fullness [19].

Dark patterns in cookie banners have been studied in [35, 36] and Matt et. al. [37] showed that banners don't even respect users choices in some cases and contain some kind of violation in about 50% of their dataset. These two studies used CMPs to detect cookie banners.

5 Reproducing previous work and adapting it

We based most of the first part of our work on Mathur et. al. [11] paper *Dark Pattern at Scale: Findings from a Crawl of 11K Shopping Websites*. In their work, the team from Princeton describe their analysis of Dark patterns from a large dataset of shopping websites and we tried to apply a similar methodology to french websites.

5.1 Mathur et. al. paper on Dark patterns

The authors created a corpus of websites using the Alexa Top Sites list [38] that records and maintains a ranking of the most visited websites on the internet and is often used in studies. To classify these websites into categories, they used Webshrinker [39], a paid online service. Then, they kept only English speaking websites by using the python language processing library polyglot [40].

Their approach was to segment websites content into separate HTML elements using a deterministic algorithm they devised, by going through the HTML elements tree and keep leaves interesting to the application as segments. Based on these segments, they used the HDBSCAN clustering algorithm [41] and manually labelled the resulting clusters to extract useful parts of shopping websites (Add to Cart buttons, clothing size options etc…). According to their paper, HDBSCAN

creates less noise in its clustering than other clustering algorithms, hence their choice.

Next, using logistic regression, Mathur et. al. built a dataset of articles from the website corpus using their urls. With this model, they were able to classify links most likely to be articles from main pages of websites. The final dataset is a set of article URLs. Then, they manually selected clusters corresponding to Dark patterns, and adapted previous definitions to categorize them. To do so, the team made several passes at defining cluster types with several of their team members and later agreed on a final decision based on it.

Finally, using OpenWPM [42], a tool based on Selenium [43] designed for automatic privacy experimentation, they crawled the dataset of articles, mimicking a human behaviour (e.g. by clicking on links) to detect Dark patterns based on the clusters they defined before.

With this work, they discovered 1.818 dark pattern instances, representing 15 types and 7 broader categories.

5.2 First work on french websites

As a result we decided to try to apply this study to french websites. To this end, we first created a corpus of french websites and our first approach was to keep websites with known french TLDs (e.g. .fr or .immo) and remove resulting websites not written in french by using polyglot. Polyglot analyzes a given text to input the most likely language it is written in, and we used the whole text of websites as input.

In order to gather a website ranking, we first used Alexa's 1M Top sites list freely available online [44] dated from 14/02/2020. We later decided to use Tranco Top list [45] for its better stability, robustness and multiplicity of sources. The Tranco list is an aggregate of four regularly used rankings : Alexa, Cisco Umbrella [46], Majestic [47] and Quantcast [48]. We decided to remove Quantcast from the aggregate as it only takes into account US traffic, and cookie banners usually appear on EU based websites, or ones accessed from an EU IP address. After removing non responding websites the final list contained 7.420 websites written in french.

We could not use the Webshrinker service to reproduce website categorization because it is a commercial service. After some research, we chose to use Fortiguard [49] categorization. To validate Fortiguard categorization, we manually labelled a sample of 336 websites and checked the frequency of erroneous categorization. We observed an accuracy of 94.6%, a false positive rate of 0.89% and a false negative rate of 4.46% (Mathur et. al. had 94%, 18% and 0.4% respectively using Alexa).

In the end, after filtering the list to only keep shopping websites, the list contained 351 elements, which is arguably not enough.

6 Focus on cookie banners

After discussion with part of the PRIVATICS team working on consent banners, we decided to shift the focus to Dark patterns and cookie banners. Our plan was to use a similar methodology as Mathur et. al. to automatically detect Dark patterns in banners.

To this end we worked together with Nataliia Bielova and her team to coordinate our efforts. Her team previously published a paper on cookie banners and the new GDPR laws [37], pointing out that many websites don't respect users' consent even after a clear decision from them. We decided to stick with the Tranco top sites ranking to build a new dataset and use it to detect cookie banners, then to crawl these banners and find Dark patterns within them.

6.1 Creating the dataset

We first decided to work on a subset of the dataset to find out whether it was doable. We used a crawler with python's Scrapy [50] along with Splash [51]. The former is a tool aiming at automatically scrap websites and the latter computes and renders Javascript code on webpages before returning the HTML. Javascript rendering is mandatory since some cookie banners only appear after it. This is done to prevent simpler robots from crawling websites or to choose whether to display the banner or not, based on the user location (inside or outside the EU).

After some issues with the crawler and especially the interaction between Scrapy and Splash not good enough, and issues with the local internet connection, including rate limitation from the internal network, we settled on a dataset of 638 websites along with their HTLM content. After filtering by language and only keeping English written websites we ended up with a dataset of 483 elements.

It is important to note that, since the scrapping was done on a French IP, a lot of international websites (such as Google or Microsoft) display their page in French and hence are out of the dataset. To prevent this, we initially planned for the final scrapping to use a UK VPN.

6.2 Web pages segmentation and textual analysis

Using this dataset, we adapted Mathur et. al. segmentation and clustering to detect cookie banners from HTML content. To do so we reapplied their dataflow with some differences (e.g. PCA is not useful here, because the number of features

is significantly lower) to the dataset. The new segmentation algorithm is based on a manual and iterative study of its effectiveness to take into account as many corner cases as possible. This algorithm is shown in Algorithm 1 and the segments dataset is composed of 55.957 segments from the 483 websites.

The main goal after segmentation was to use a NLP approach to create a Bag of Words matrix based on segments from the websites. In such a matrix, rows correspond to segments and columns correspond to words whereas the content of the matrix depending on the algorithm used later and contains information on the occurence and/or frequency of appearance of words. The most used algorithms are Count Vectorizer and Tf-idf Vectorizer. The first one simply counts the number of occurrences of words in segments whereas the second one takes into account their frequency across the whole corpus. Mathur et. al. chose to use Count Vectorizer because of the noise that Tf-idf creates in HDBSCAN clustering.

We first tried to let the algorithm create the vocabulary to cluster segments based on their word appearance frequency. This approach created a lot of noises and the result of the HDBSCAN clustering was volatile and varied a lot depending on the subset of the data. After discussion, we decided it would be easier to give a score to segments based on specific words we know appear often in cookie banners, such as 'cookie', 'policy' or 'data' for example. In the end, this is the same as forcing the vocabulary of the BoW to be cookie banner related. Previously, we let the BoW algorithm create the vocabulary based on words found in the segment list.

The final cookie banner vocabulary can be found in Appendix A, we built it in an interative manner when analyzing websites and during the manual tests.

6.3 Clustering and scoring

From there, we worked on finding the best mix of segmentation algorithm, parameters selection, dictionary selection, tokenization method, clustering algorithm selection, and other parameters with a manual check of websites to validate. This process was highly iterative. The main parameters were the cluster size (2-5) and the distance metric (manhattan/euclidean) and the vocabulary must be precise but still take into account as many banners as possible. Tokenization is simple but we tested it with and without stemming.

After some iterations, we deemed the segmentation algorithm good enough and the vocabulary selection gave much less noise than before. However after checking the content of the clusters, it was still not usable enough, clusters did not always mapped to a specific type of banner nor made a good distinction between banner segments and non-banner segments. Following reflections on what to do, including considering manual labelling, we settled on a much simpler solution

and chose to associate segments with scores loosely based on the segment textual content and its similarity to the cookie banners vocabulary.

6.4 Caveats

This gives a good estimate of cookie banners segments on web pages, however this method has 3 main caveats :

1. It is dependent on the quality of the segmentation
2. Even though we tested numerous sites the vocabulary may not be exhaustive enough
3. Some websites don't have cookie banners at all

It is hard to evaluate the quality of the segmentation, other than by picking a random sample and checking by hand whether one of the segment correspond to a cookie banner.

To make the vocabulary exhaustive is not trivial either as in theory we would need to analyze every single website of the web, but manual evaluation showed that the vocabulary was actually pretty robust for its size (i.e. we almost always ended up with a segment corresponding to a banner with a relatively simple vocabulary. An interesting thing to note here, is that the vocabulary is language dependant but not the cookie banner detection, this means that provided a vocabulary in another language corresponding to cookie banners, the whole analysis can be done with this language. We decided to stick with English, mostly for the readability of the final report, and to be easier to evaluate.

As for cookie banner existence, a significant number of websites do not have cookie banners at all but still often have footer links such as 'privacy policy', this means we could get false positives, i.e. segments classified as banners when they are not.

To get around this there are multiple options:

- Stick to websites using a CMP
- Add some more complex tests
- Define a score lower threshold on segment scores

Sticking to websites that use CMPs is easier because they call specific predefined Javascript procedures that we can catch. However, even if these calls are supposed to be IAB standards, some websites depending on CMPs do not use

them. In addition, this options forces the analysis to focus on the subset of websites using CMPs.

We could have added more tests to check whether segments classified as banner were indeed cookie banners, but this requires more NLP knowledge that the author did not posses, and is still not guaranteed to work, likely requiring even more tests that could never really be exhaustive given the amount of different websites on the internet.

The last option, setting a lower threshold, is the simplest option and yet is effective. This entails setting a specific score that should be met by a segment to be considered as a cookie banner. This can be very effective in weeding out false positives, but at the cost of creating false negatives (e.g. banners with few key-words). In the end we chose this option, with a threshold as low as possible. We agreed that if the sample data was big enough and the threshold low enough, the amount of usable data should still be sufficient.

6.5 Crawling with OpenWPM and information extraction

After trying to work with Scrapy/Splash for crawls, we decided to switch to OpenWPM to automate the process of finding Dark patterns. The main 5 steps of crawling with OpenWPM are as follows:

1. Crawl sites using OpenWPM
2. Identify segments on the page
3. Identify segments corresponding to cookie banners when they exist
4. Relevant information extraction on cookie banners
5. Look for Dark patterns in banners

Note that step 5 can be done in the end, after crawling whereas the first 4 steps are to be done sequentially during the crawl, this is what we decided to do.

Due to the complexity and diversity of websites, we decided to focus on the first layer of cookie banners. Most banners indeed have a layer with an 'accept' button or at least a cookie policy notice and many of them have subsequent and more complex layers that mostly provide tailoring of cookie usage from the website.

Using a dataset of websites used by the team in some earlier work, we manually validated our dataflow and algorithms, by checking whether cookie banners were correctly detected. This led to subsequent improvements in the cookie banner detection (segmentation and extraction). In the first versions of the segmentation algorithm, we split the page source into segments based on several features

such as text, type of element, or structure and derived a score for each of these segments. From this score, we decided whether the current segment was a cookie banner as described earlier.

In a later approach, we mixed the first detection algorithm, with a new one that scans elements starting from the root of the html tree and picks the best ones according to textual scores again. This top down approach mixed with the bottom down one from the previous algorithm gives much better results, and returns a correct banner for most of the websites we tested it on that have a banner. In particular, this algorithm is much better at finding the top-most html element containing the banner, while still being exclusively a cookie banner. This new version is better in cases such as on the etsy website banner in Appendix B Figure 3h. We describe this approach further in the following section.

The next step after locating a cookie banner on a website is to extract relevant features from it, depending on the dark patterns detection that comes after. This means that feature extraction and dark pattern detection are done and improved in an iterative way depending on one another.

We settled on features from clickable elements such as buttons or links. To extract these element, we first check whether they are not obstructed and then find out if the element can be clicked on.

we chose the following features to extract from the selected elements :

- Font
- Font color
- Color
- Background color
- Size
- Visibility
- Button types (e.g. accept, decline etc...)

The reasoning behind this set of features, is that most first layer Dark patterns seem to be either from the absence or dissimulation of choice, or from a difference in visibility of buttons.

6.6 Link with Dark patterns

For our Dark patterns analysis, we chose to use the taxonomy from Gray et. al. [52] as it is one of the main work on which we based ours.

There are multiple Dark pattern taxonomies that have been defined, but none really apply perfectly to cookie banners. The one defined by Gray et. al. is derived from Brignull [8] but focuses on *strategies* instead of Dark pattern categories. These strategies contain the original Dark patterns and are defined as follows :

- Nagging: Repeated forced interactions with the interface (e.g. pop-ups that reopen)
- Obstruction: Making processes more difficult to dissuade users from taking specific actions
- Sneaking: Hide or delay relevant information
- Interface interference: Manipulation of the interface to privilege some actions over others
- Forced action: Forcing users to perform some actions, for example by blocking functionalities

Nagging in cookie banner could happen if the consent was asked multiple times on the same websites until the user accepted it for example.

Forced action happens on some websites that don't let you access it before accepting or declining cookies. On the website `https://www.healthline.com/` for example, you can access to more information on how cookies are collected, but it is not possible to deactivate any single vendor, hence forcing users to accept cookies to visit the website. In some cases, it is not even possible to access a website without accepting cookies, and the website redirects you to a liter version of it or even outside the domain (e.g. back to google).

Sneaking can be observed in cookie banners in the delaying of information concerning vendors to layers beyond the first one for example.

Obstruction happens when you cannot take specific actions, for example when a decline button is not directly accessible (and is in further layers) or not at all. Same for option buttons.

Interface interference is prominent in cookie banner. Almost every banner users encounter on the internet have this Dark pattern strategy embedded in it. It can often be observed when accept buttons are bigger, with background color or more visible in general than their decline and option counterparts, like in Figure 1.

Our work focuses mainly on Obstruction and Interface interference. Other Dark pattern strategies involve either further analysis of the text with at least sentiment analysis or/and full analysis of further layers in banners. We chose to target interface elements that are easier to analyse, namely buttons and links along with their corresponding information (size, color, font etc...)

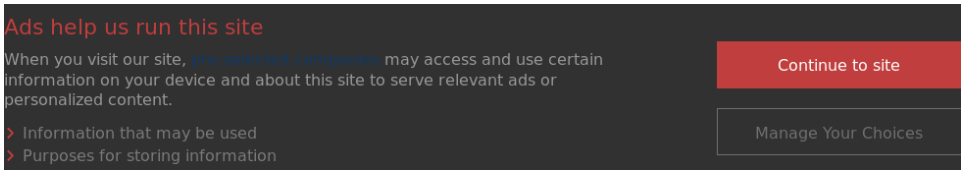


Figure 1: Example of a cookie banner with Interface interference: The consent button ('continue to site') is designed to be more visible than the option button ('Manage your choices')

When a banner does not provide direct access (i.e. in first layer) to a decline button and/or an option button we say that the banner showcases an Obstruction Dark pattern strategy. When a banner makes the accept button more visible than decline or option button, we say that it showcases an Interface interference Dark pattern strategy.

Note that both strategies are not mutually exclusive, as we could have the absence of a decline button in the first layer, and an option button that is less visible than the accept button. This is visible in Appendix B Figure 2c for example, in this banner the 'I accept' button is red whereas the options button is grayed even though both have the same size.

7 Aside: Work with the cookie banner team

Near the end of the internship, Ruba Abu-Salba and Cristiana Santos from the part of the team working on cookie banners, asked me to generate a dataset for their work. We settled on a dataset of 2.000 websites, along with the screenshot of each site, its cookie banner, the text of the banner and a screenshot of every page linked to by elements contained in the banner.

Ruba and Cristiana will manually analyze this dataset to look for Dark patterns in ways that are harder to do automatically and are working on a paper on this subject.

8 Methodology of final dataset creation

We ran our cookie banner extraction tool on 100.000 websites taken from the Tranco top list. We generated the list to only contain one website per domain, ones marked as safe, and we excluded Quantcast from the list aggregate as it focuses on US traffic. The crawl was made on a machine based in the Netherlands over ten days.

We kept only websites where both the main page **and** the cookie banner were displayed in English, relying as before on the polyglot [40] library to do so. We focused on English-speaking websites because it is easier for the analysis, but the tool can be adapted for any language by changing the vocabulary used to detect banners for a given language, as described earlier.

The crawl was made using OpenWPM as stated. We chose OpenWPM because it is efficient, permits full page rendering before analysis, and has built-in logging and data collection capabilities. As opposed to Scrappy/Splash [50, 51] used in [36] and earlier by us, OpenWPM provides crash recovery, bot mitigation workarounds, and full JavaScript rendering, as well as allows to take screenshots of whole pages or specific elements (such as cookie banners) on websites. As discussed earlier, some websites only display their banners with JavaScript rendering, making full rendering capabilities mandatory.

8.1 Detection of cookie banners.

Cookie banner detection is done in three steps: segmentation, scoring, and tree traversal. The first step is to segment the web page into meaningful small segments and build a segment tree [11], based on their HTML tag and text.

The segmentation algorithm depicted in Algorithm 1 is a modified version of the one used by Mathur et. al [11] to target cookie banner instead of more general segments. After building the segment tree, we assigned a score to each segment based on its inner text using a vocabulary that we built by analysing typical cookie banner content. This vocabulary is used without stemming and is provided in Appendix A.

We ranked tree leaf segments according to their scores. Finally, using the highest scoring segments, we traversed the segment tree both bottom up and top down to narrow down the HTML element that is as close as possible to the root of the tree but only contains the banner. This is done in Algorithm 2.

It is important to note that, the extraction tool does not have a way to know whether a website actually contains a banner. On websites that do not contain cookie banners, the extraction often outputs footers with links to privacy policies of the web page. Since we do a manual analysis, these false positives can be filtered out. However, for a more automated and easier analysis we use the segment scores and decide whether a cookie banner exists based on a threshold that we set. This threshold is high enough to avoid as much false positives as possible (outputting 'yes' to cookie banner existence when the websites does not have one) but creates some false negatives.

We ran the tool with a threshold of 4, meaning that a segment can only be accepted as a segment corresponding to a banner if it has a score of 4 or higher.

In practice it means that the corresponding HTML element must be selected as a potential banner, and it should have at least 4 cookie banner keywords from the vocabulary of Appendix A.

Algorithm 1 Segmentation algorithm used to create a segment tree from HTML elements on websites.
Elements to ignore are predefined and $blocks$ is a set of html elements of interest.
$elem$ is a HTML element.
A separate algorithm is used to define what is a *possible leaf*, and mostly checks how deep the element tree goes and whether it has text in it.

Input: $parent, elem$
  if $elem$ in elements to ignore **then**
    return None
  end if
  if $elem$ is a possible leaf **then**
    if $elem$ not in $blocks$ and is not root and $parent$ is not root **then**
      Set $parent$ to leaf
    else
      Set $elem$ to leaf
    end if
    return $\{elem, parent\}$
  end if
  if $elem$ is not root, $elem$ (and children) not in $blocks$ **then**
    return None
  end if
  $children =$ recursively apply algorithm to each child of $elem$
  if $children$ is empty **then**
    return None
  end if
  if $children$ has a single element and is a leaf **then**
    Set $elem$ to leaf
  end if
  return $\{elem, parent, children\}$

Algorithm 2 Find cookie banners.
The algorithm first finds the segment in the segment tree with highest score then go down the tree from the root in the direction of this segment until the score drops.
Then, we go back up the tree until the score goes back up again (to keep the banner as general as possible)

$root =$ segment tree from Algorithm 1
$leaves =$ set of elements in segment tree marked as leaves
$highest =$ element from $leaves$ with highest score
$elem =$ immediate root child that contains $highest$
while $elem$ has children **do** ▷ Go down the tree until score drops
    $new\_elem =$ element from $elem.children$ with highest score
    if new_elem score < $elem$ score **then**
        $elem =$ highest element in $elem$ ancestor tree that still has the same score
        return $elem$
    end if
    $elem = new\_elem$
end while
return $elem$

9 Results

We analyzed 100k websites from the Tranco list we described previously. The analysis was run over the span of about ten days from 07/07/2020 to 17/07/2020.

In the end, we gathered 9.344 different cookie banners (or marked as cookie banners by our tool). After filtering the data from banners with no usable information and some of the most obvious false positives, the final dataset contained 6.146 banners. Finding so much false positives means that the final threshold score was likely too low, this false positives filter post-generation was done by looking at button classification and weeding out banners that had many unclassified buttons. We chose this criteria over the fact that most false positives seemed to output most the the web page according to the corresponding screenshots, hence the reasoning is to remove banners so big that they might actually be full pages. This filtering is not ideal and might have removed some true positives.

Each banner in the dataset is represented by its screenshot along with a json file containing useful information such as its text, the website url, the visit id, and information about buttons and their classification.

We mostly look at two type of information: existence/absence of buttons that

should appear in cookie banners and difference in visibility between them when they exist. As stated earlier, difference in visibility is checked with simple tests, we look at the difference is size, shape, font and background color of buttons. This visibility test is simple and could let through some false negatives, hence visibility figures are lower bounds on the reality of cookie banners.

We look at three types of buttons, accept buttons (e.g. 'I accept', 'OK', 'Continue to site'), decline button (e.g. 'I decline', 'reject') and what we will call options buttons that encompass privacy policy and preference management types of buttons.

9.1 Presence and absence of buttons

We found that 3.758 of banners in our dataset (61.14%) had an accept button. Besides accept buttons there are several points of interest to notice. First, we observe that 1655 banners have no button at all among the three types that interest us and 625 only have an option button. This is due to a lot of banners being actually a subset of banner we will call consent notices, that do not let users interact much more than by closing it. The banner from Figure 2e is a good example of this kind of cookie notice, that you can only close. These consent notices sometimes have links to privacy policies that are classified as options buttons.

We can also look at the few cases where there is no accept button but there is a decline button. This happens in 108 (1.76%) of the banners we gathered and is pretty counter-intuitive. We reviewed manually these cases and it turns out that they are often due to corner cases in our segmentation, i.e. these banners use keywords that we did not include in our vocabulary (such as 'Got it' to accept).

Among the 3.758 banners that do have an accept button, only 629 had a decline button (16.4%) and 1.953 (51.16%) had at least one options button. These numbers are important, as they suggest that most banners that provide explicit consent to cookies do not let users choose to not consent.

In the end only 200 banners, that is 3.25% of the dataset or 5.21% of banners with an accept button, provide all three of buttons which is arguably low, and even more so provided that we use a loose definition of options buttons.

Table 2 and Table 3 summarize the different configurations of existence of buttons.

In Figure 2, we present several examples of cookie banners with different configurations of buttons. Figure 2a, from BuzzFeed News shows a banner containing an accept, decline and options button. We can also see a slight difference in visibility between the accept and decline buttons due to the background color.

Figure 2b shows a banner from the Rockstar games website with an accept and decline button with no options button, and again the accept button is slightly

Table 2: Principal results of quantity and frequency of different buttons

Has accept button	Has decline button	Has options button	Amount and Frequency
False	False	False	1655 (26.93%)
False	False	True	625 (10.17%)
False	True	False	67 (1.09%)
False	True	True	41 (0.67%)
True	False	False	**1376 (22.39%)**
True	False	True	**1753 (28.52%)**
True	True	False	429 (6.98%)
True	True	True	**200 (3.25%)**

(a) Amount and frequency of the different possible configurations of existing buttons (accept, decline and options)

Has accept button	Has decline button	Amount and Frequency
False	False	2280 (37.1%)
False	True	108 (1.76%)
True	False	**3129 (50.91%)**
True	True	**629 (10.23%)**

(b) Amount and frequency of the different possible configurations of accept and decline buttons

Has accept button	Has options button	Amount and Frequency
False	False	1722 (28.02%)
False	True	666 (10.84%)
True	False	1805 (29.37%)
True	True	1953 (31.78%)

(c) Amount and frequency of the different possible configurations of accept and options buttons

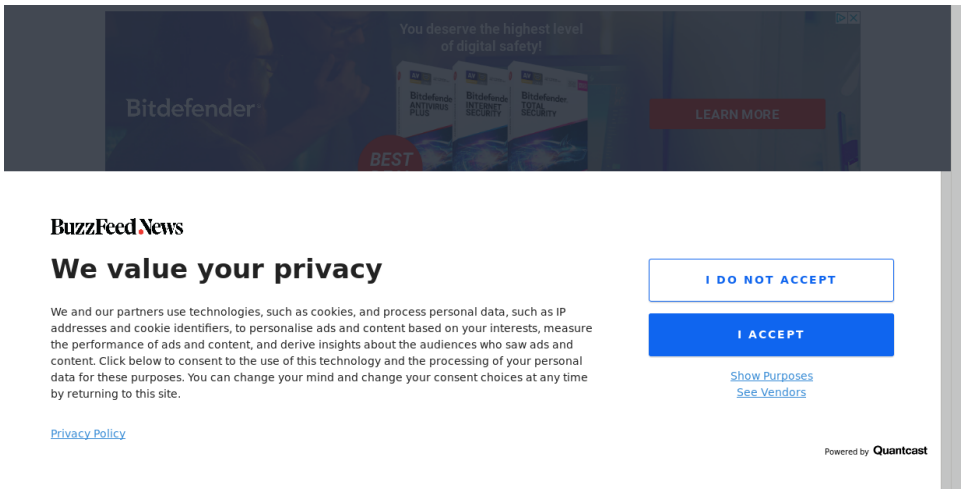
Has decline button	Has options button	Amount and Frequency
False	False	1376 (36.61%)
False	True	1753 (45.95%)
True	False	429 (10.83%)
True	True	200 (5.21%)

Table 3: Amount and frequency of the existence or absence of decline and options button **in presence of an accept button**

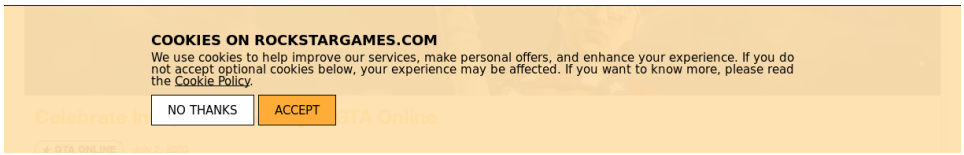
more visible. Next, in Figure 2c from the Mental floss website is a banner with an options and accept button but no way to immediately decline the cookie usage, and once again the accept button is the most visible.

The banner from the Insider website in Figure 2d has only an accept button and nothing else (beside links to the privacy policy). Lastly, the OpenDNS banner in Figure 2e only provides an option button ('change your preferences').

Figure 2: Examples of different configurations of existence/non existence of accept, decline and options buttons in cookie banners



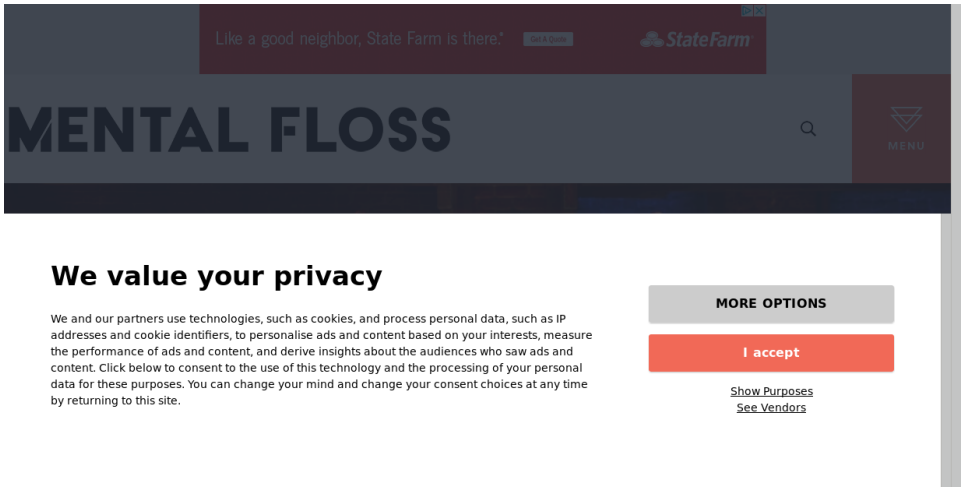
(a) Banner from Buzzfeednews website, containing an accept, decline and options button ('purposes').



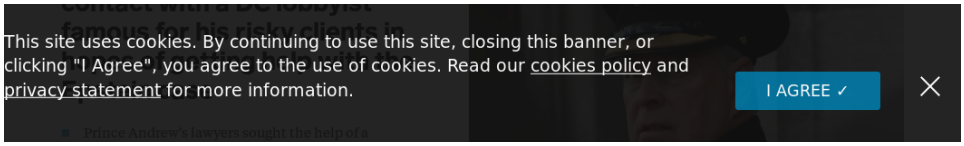
(b) Banner from Rockstargames website, containing and accept and decline button but no options button

9.2 Buttons visibility

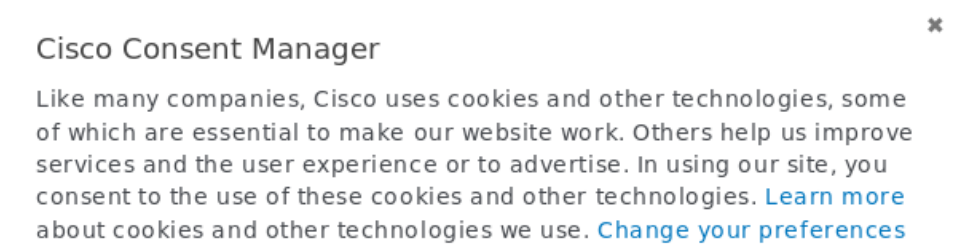
We present next the results in the differences in visibility that we computed from buttons attributes such as their sizes, fonts, or colors. This is done on decline and options buttons against accept buttons, we decided that comparing decline and options buttons visibility was not enough of interest.



(c) Banner from Mentalfloss website, containing an accept and options button but no decline button



(d) Banner from insider website, containing only and accept button



(e) Banner from OpenDNS website, containing only an option button ('change your preferences')

In the following, we focus on banners that have an accept button (61.14% of the dataset).

When a banner has a decline button, it is less visible than the accept button in 544 banners (86.49% over 629 banners). For options buttons, it happens in 1541 banners (78.9% over 1717 banners).

Table 4 showcases the difference in visibility of the options and decline buttons compared to accept buttons.

Table 4: Comparison in visibility between accept buttons and decline/options buttons

decline button less visible than accept button	Amount and Frequency
False	85 (13.51%)
True	**544 (86.49%)**

(a) Amount and frequency of decline buttons being less visible than accept buttons in banners that have both.

options button less visible than accept button	Amount and Frequency
False	176 (9.01%)
True	**1541 (78.9%)**

(b) Amount and frequency of options buttons being less visible than accept buttons in banners that have both.

These number are lower bounds since we can't focus on every possible aspect that makes a button less visible than another. In addition, the 'best' way to make one button more visible than an other is to only show one, which happens often as described above.

9.3 Dark patterns

Using our prior definitions of Dark patterns, we can observe that among these 6.146 banners, 5509 (89.63%) of them have a Obstruction Dark pattern. This obstruction indeed happens when there is no way to refute consent (i.e. no decline button).

The Interface interference Dark pattern can be seen when buttons have differences in visibility between the accept button and the decline button or the option button. To be precise, we say that there is an interface interference in a banner when it has an accept button and one of the following:

- a decline button less visible than the accept button
- an options button less visible than the accept button
- both of the above

This happens in 2085 (33.92%) of banners.

In the end we observe at least one of the two Dark patterns in 5.953 out of 6.146 (96.86%) banners in our dataset.

10 Discussion

We presented the state of the art and multiple definition of Dark patterns, Nudges and Sludges and discussed the complexity to use a single definition of Dark patterns.

It seems that in its current state and even with EU laws such as the GDPR or cookie laws, the privacy of users is still not as respected as stated in these laws. Several teams proved that cookie banners are often not lawful and multiple studies showed the presence of dark patterns in cookie banners.

This work contributes to the privacy research by presenting a tool that automatically detects cookie banners on websites that do not necessarily depend on a CMP and add to existing work by detecting Dark patterns on 6.416 cookie banners from 100k websites by looking at button types and shape.

We discovered at least one Dark pattern among two of them (Obstruction and Interface interference) in 96.86% of banners we observed. This entails that cookie banner often do not provide a way to decline consent for data collection or nudge users into accepting consent by making consent the default and easiest choice with asymmetry of visibility in buttons, or both.

This work has some downsides. We discovered late into the analysis that the segment score threshold was too low and thus we let through too much false positives, we only focus on the first layer of banners and only work with English banners. Furthermore, the final dataset is a bit small. However, all of these issues can be fixed relatively easily in future work, by raising the score threshold, adding other languages to the cookie banner vocabularies and generate more data.

This work spans 6 month from February to July 2020 and the results are dependant on this time period, and hopefully cookie banners will improve in quality in future years.

It is clear from related work and this one that cookie banners are not only deceptive but are deceptive *on purpose*. Most banners do not follow EU directives and a lot of them make use of cognitive biases of their websites visitors to lure them into accepting cookies.

This is worrisome, especially in cases that are harder to detect, and shows that the EU does not enforce their own regulations either from a lack of clear evidence (website owner can always pretend that they did not implement Dark patterns on purpose), or simply due to the sheer number of websites available on the internet.

A lot of internet users are not trained researchers nor technology enthusiasts and some of them already have a hard time navigating simple web pages. This makes cookie banners all the more effective, being an addition to an already complex and hard to master internet.

Most of the regulation on the internet is hard to enforce, partly due to its size

and lack 'borders' between legislative states. However, most websites **do** display cookie banners, which is already a step in the right direction.

11 Further work

There are several points to expand on from this project.

11.1 Beyond the first layer

We need an analysis of further layers of banners. This work focuses only on the first layer but most of the Dark patterns are found in a second and sometimes a third layer. In these layers, we can often find confusing wording, unclear interfaces, and complex ways to opt out of cookie collection. Even if most users don't go as far as to check further layers, and even with training in information technology, it can be a real pain to go through the interfaces of some banners.

Layers beyond the first one are even more complex and organization dependent, hence the difficulty to make an automatic analysis for them. These layers often include walls of text, on/off sliders that are sometimes unclear on whether they activate or deactivate cookies, confusing wording, multiple (and sometimes way too much) options to pick from, no way to decline all cookies etc...

11.2 NLP and sentiment analysis

An in depth analysis of the text of banners is needed to extract some Dark patterns. Some banners use specific wording to either confuse users or nudge them towards accepting cookies. For example, some of these texts use language that guilt users into accepting cookies with phrasing such as 'Cookies help us run this website' or 'EU laws force us to ask for your consent to collect cookies'.

It would be beneficial to have a full sentiment analysis of cookie banners text, and at all layers. Having a way to detect phrasing designed to be confusing and in general wording that is not neutral would be a great addition to this work.

This is in part what is done manually by some part of the team at PRIVATICS, and implementing an automatic way to make this analysis could drastically improve the scale of similar work.

11.3 Larger scale

Even though we collected 6.416 different banners from 100k websites, this dataset could be much bigger. The generation of this dataset took about 10 days on a

machine designed for heavy computation, with 8 cores and access to competitive internet connection in a cluster.

To construct a bigger dataset, a researcher would need more machines, more time and/or improve the complexity of the banner detection.

Our work depends on OpenWPM, Polyglot and multiple libraries and is very dependent on their inherent performance making it difficult to really improve the computation time

11.4 Add more languages

As stated at several points in this paper, we focused on english written websites and cookie banners. However, the way we designed our segment score assignment is independant of the language.

Anecdotally, we manually tested on several french websites to detect banners with a french vocabulary, and results were similar in term of banner detection quality to this work. We believe that with a sufficiently good cookie banner vocabulary designed for more (or most) languages spoken in the EU, this work could be augmented to target any website in Europe.

List of Figures

List of Tables

References

[1] Hendrik P Van Dalen and Kène Henkens. "Comparing the effects of defaults in organ donation systems". In: *Social science & medicine* 106 (2014), pp. 137–142.

[2] Shai Davidai, Thomas Gilovich, and Lee D Ross. "The meaning of default options for potential organ donors". In: *Proceedings of the National Academy of Sciences* 109.38 (2012), pp. 15201–15205.

[3] Eric J Johnson and Daniel G Goldstein. "Defaults and donation decisions". In: *Transplantation* 78.12 (2004), pp. 1713–1716.

[4] Isaac Dinner et al. "Partitioning default effects: why people choose not to choose." In: *Journal of Experimental Psychology: Applied* 17.4 (2011), p. 332.

[5] Douglas MacKay and Alexandra Robinson. "The ethics of organ donor registration policies: Nudges and respect for autonomy". In: *The American Journal of Bioethics* 16.11 (2016), pp. 3–12.

[6] Richard H Thaler and Cass R Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009.

[7] Richard H. Thaler and Cass R. Sunstein. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press, 2008.

[8] Harry Brignull. *Dark Patterns*. https://90percentofeverything.com/2010/07/08/dark-patterns-dirty-tricks-designers-use-to-make-people-do-stuff/. 2010.

[9] Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[10] Dilip Soman et al. "Seeing Sludge: Towards a Dashboard to Help Organizations Recognize Impedance to End-User Decisions and Action". In: (2019).

[11] Arunesh Mathur et al. "Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites". In: *Proceedings of the ACM Human-Computer Interaction* 3 (2019).

[12] Harry Brignull. *Dark Patterns*. https://www.darkpatterns.org. 2018.

[13] *Reddit: Hard to opt-out of a New York Times subscription*. https://www.reddit.com/r/assholedesign/comments/hy70xi/after_going_through_five_pages_to_cancel_your_4/.

[14] Carol Moser, Sarita Y Schoenebeck, and Paul Resnick. "Impulse Buying: Design Practices and Consumer Needs". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–15.

[15] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32016R0679. 2016.

[16] *Directive 2009/136/EC of the European Parliament and of the Council*. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:337:0011:0036:En:PDF. 2009.

[17] Iskander Sánchez-Rola et al. "Can I Opt Out Yet?: GDPR and the Global Illusion of Cookie Control". In: *Proceedings of the ACM Asia Conference Computer and Communications Security*. 2019, pp. 340–351.

[18] Christine Utz et al. "(Un)informed Consent: Studying GDPR Consent Notices in the Field". In: *Conference on Computer and Communications Security*. 2019.

[19] Cristiana Santos, Nataliia Bielova, and Célestin Matte. *Are cookie banners indeed compliant with the law? Deciphering EU legal requirements on consent and technical means to verify compliance of cookie banners*. Available at https://arxiv.org/abs/1912.07144. 2019.

[20] Célestin Matte, Cristiana Santos, and Nataliia Bielova. "Purposes in IAB Europe's TCF: which legal basis and how are they used by advertisers?" In: *Annual Privacy Forum, APF*. Lecture Notes in Computer Science. Accepted for publication. https://hal.inria.fr/hal-02566891. 2020.

[21] *IAB Transparency and Consent Framework*. https://iabeurope.eu/transparency-consent-framework/.

[22] Sandra C Matz et al. "Psychological targeting as an effective approach to digital mass persuasion". In: *Proceedings of the national academy of sciences* 114.48 (2017), pp. 12714–12719.

[23] Muhammad Ali et al. "Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging". In: *arXiv preprint arXiv:1912.04255* (2019).

[24] Alessandro Acquisti et al. "Nudges for privacy and security: Understanding and assisting users' choices online". In: *ACM Computing Surveys (CSUR)* 50.3 (2017), pp. 1–41.

[25] Eric J Johnson, Steven Bellman, and Gerald L Lohse. “Defaults, framing and privacy: Why opting in-opting out”. In: *Marketing Letters* 13.1 (2002), pp. 5–15.

[26] Cass R Sunstein. “Sludge and ordeals”. In: *Duke LJ* 68 (2018), p. 1843.

[27] Cass R Sunstein. “Sludge audits”. In: *Behavioural Public Policy* (2020), pp. 1–20.

[28] Jamie B. Luguri and Lior Jacob Strahilevitz. “Shining a Light on Dark Patterns”. In: 2019.

[29] Norwegian Consumer Council. “Deceived by design, How tech companies use dark patterns to discourage us from exercising our rights to privacy”. In: *Norwegian Consumer Council Report* (2018).

[30] Gregory Conti and Edward Sobiesk. “Malicious interface design: exploiting the user”. In: *Proceedings of the 19th international conference on World wide web*. 2010, pp. 271–280.

[31] Shruthi Sai Chivukula et al. “" Nothing Comes Before Profit" Asshole Design In the Wild”. In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–6.

[32] Fiona Westin and Sonia Chiasson. “Opt out of privacy or" go home" understanding reluctant privacy behaviours through the FoMO-centric design paradigm”. In: *Proceedings of the New Security Paradigms Workshop*. 2019, pp. 57–67.

[33] René Van Bavel and Nuria Rodriguez-Priego. *Testing the Effect of the Cookie Banners on Behaviour*. `https://ec.europa.eu/jrc/en/publication/testing-effect-cookie-banners-behaviour`. 2016.

[34] Martin Degeling et al. “We value your privacy... now take some cookies: Measuring the GDPR’s impact on web privacy”. In: *arXiv preprint arXiv:1808.05096* (2018).

[35] Than Htut Soe et al. “Circumvention by design–dark patterns in cookie consents for online news outlets”. In: *arXiv preprint arXiv:2006.13985* (2020).

[36] Midas Nouwens et al. “Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence”. In: *CHI*. 2020.

[37] Célestin Matte, Cristiana Santos, and Nataliia Bielova. “Do Cookie Banners Respect my Choice? Measuring Legal Compliance of Banners from IAB Europe’s Transparency and Consent Framework”. In: *IEEE Symposium on Security and Privacy (IEEE S&P 2020)*. 2020.

[38] *Alexa Top sites list.* https://www.alexa.com/topsites.

[39] *Webshrinker categorization.* https://www.webshrinker.com/.

[40] *Polyglot natural language processor.* http://polyglot.readthedocs.org/.

[41] *HDBSCAN clustering algorithm.* https://hdbscan.readthedocs.io/en/latest/index.html.

[42] Steven Englehardt et al. "OpenWPM: An automated platform for web privacy measurement". In: *Manuscript, mar* (2015).

[43] *Selenium browser automation tool.* https://www.selenium.dev/.

[44] *Alexa Top 1M free list.* s3.amazonaws.com/alexa-static/top-1m.csv.zip.

[45] *Tranco list G2WK.* https://tranco-list.eu/list/G2WK/1000000. 2020.

[46] *Cisco Top 1M free list.* https://s3-us-west-1.amazonaws.com/umbrella-static/top-1m.csv.zip.

[47] *Majestic Top 1M free list.* http://downloads.majestic.com/majestic_million.csv.

[48] *Quantcast Top 1M free list.* https://ak.quantcast.com/quantcast-top-sites.zip.

[49] *Fortiguard categorization.* https://fortiguard.com/updates/webfiltering.

[50] *Scrapy web crawling and scraping framework.* https://scrapy.org/.

[51] *Scrapy Splash JS rendering tool.* https://github.com/scrapy-plugins/scrapy-splash.

[52] Colin M. Gray et al. "The Dark (Patterns) Side of UX Design". In: *Proceedings of the CHI Conference Human Factors in Computing Systems*. 2018, p. 534.

A Cookie banner vocabulary

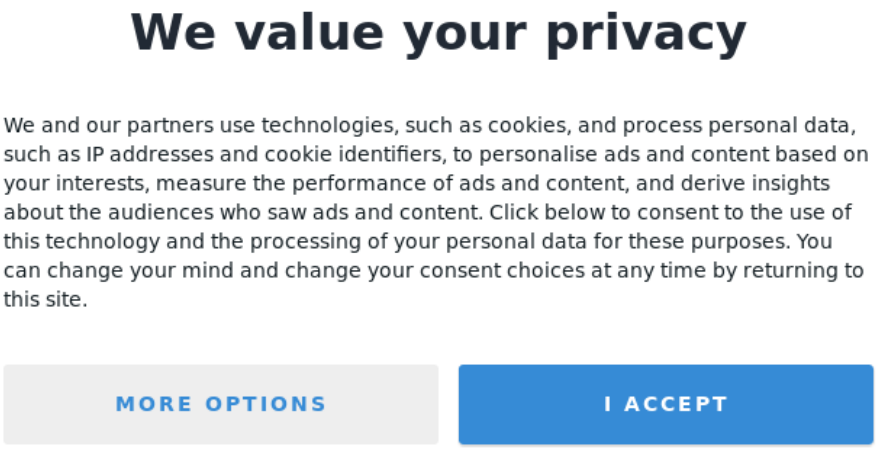
Below is the set of words used to compute segments score for cookie banners, as described in the methodology section. Note that we don't use stemming during the analysis.

{'cookie', 'cookies', 'privacy', 'analytic', 'analytics', 'personalisation', 'personalization', 'personalize', 'personalise', 'improve', 'tailor', 'tailored', 'ad', 'ads', 'agree',

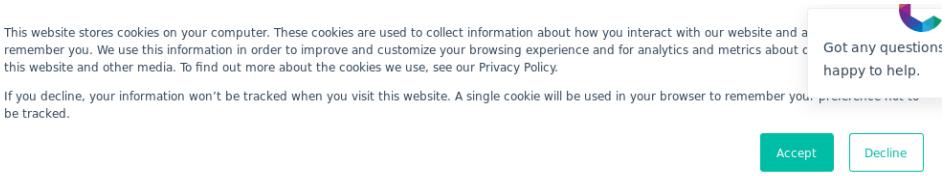
'policy', 'measure', 'allow', 'collect', 'consent', 'control', 'data', 'gdpr', 'opt', 'advertisement', 'advertisers', 'advertiser', 'security', 'settings', 'require', 'required', 'reminder', 'accept', 'ok', 'decline', 'party', 'parties'}

B Cookie banner examples

This section contains several examples of cookie banner encountered on the internet and referenced to in this work. There exists many different types of banners and we tried to showcase diverse versions of them.

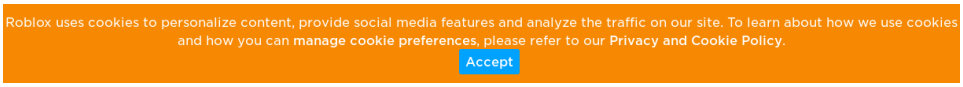


(a) Banner from the Know Your Meme website

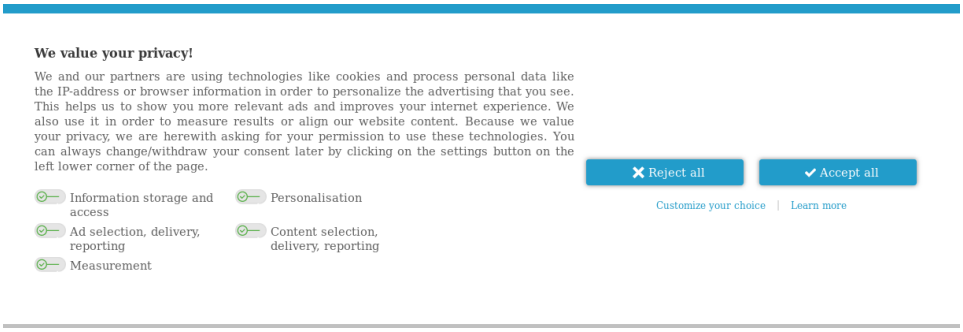


(b) Banner from the Linked website

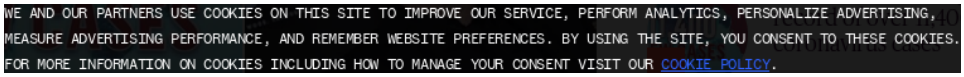
(c) Banner from the Moodle website



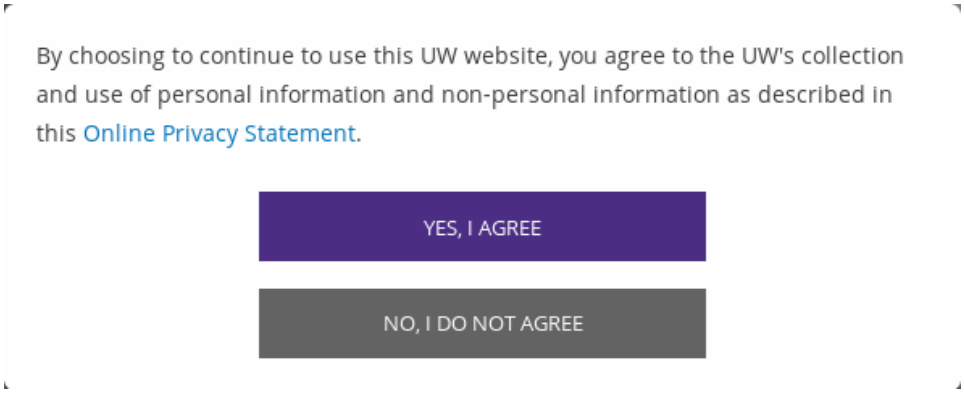
(d) Banner from the Roblox website



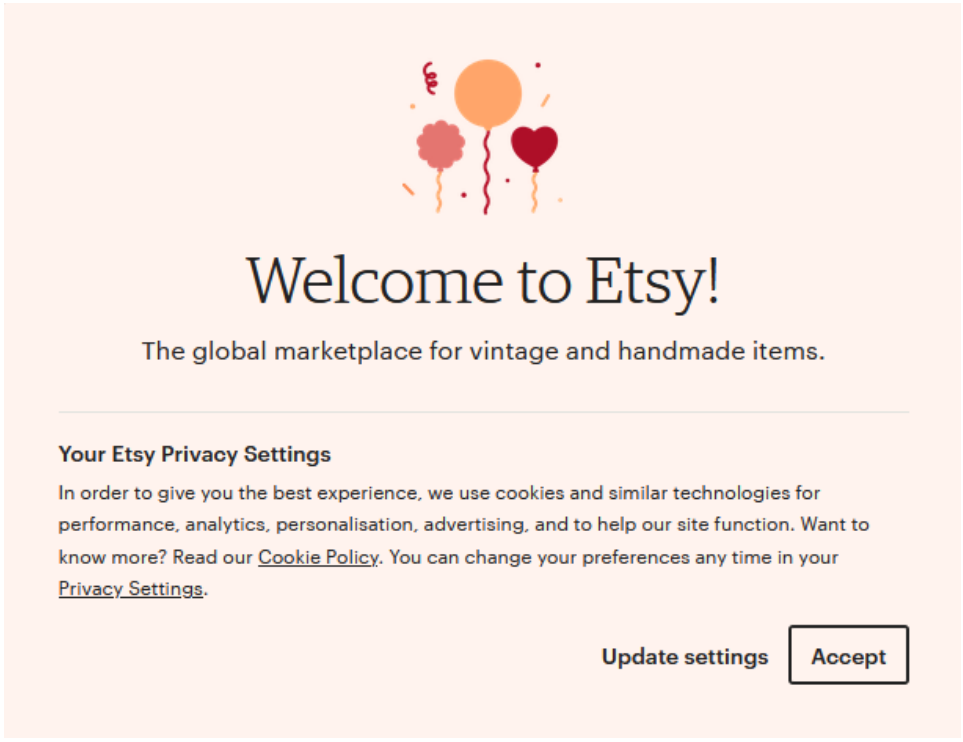
(e) Banner from Ronaldo7 website



(f) Banner from Today website



(g) Banner from Washington edu website



(h) Banner from the Etsy website, extracted on 06/06/2020

(i) Banner from the Tedibear website (in French), extracted on 29/07/2020