# A Systematic Approach for A Reliable Detection of Deceptive Design Patterns Through Measurable HCI Features

EMRE KOCYIGIT, University of Luxembourg, Luxembourg
ARIANNA ROSSI, Sant'Anna School of Advanced Studies, Italy
GABRIELE LENZINI, University of Luxembourg, Luxembourg

Dark patterns are deceptive design elements of digital choice architectures that are implemented to drive users' actions towards decisions that are not necessarily in their best interest, such as accepting privacy-invasive practices. Most dark patterns are considered unlawful, but their description is rather informal. Thus, detecting dark patterns among the various existing design patterns and discerning what is an illegitimate design practice may depend on the subjective interpretation of expert users (such as regulators, civil society organizations, and academic researchers) who may not fully agree. The need to ground any evaluation on evidence calls for a reliable approach that is based on descriptions relying on observable, measurable features. Taking cookie consent as a use case, where dark patterns are ubiquitous and intensively under scrutiny, we propose a systematic approach to describe the characteristics of deceptive design patterns that are intended to reconcile the interpretations of expert users. In particular: i) we identify use case-specific dark pattern types using the ontology drafted by Gray *et al.* (2024); ii) we clarify the relationships between those types and the dark patterns' attributes proposed by Mathur *et al.* (2021); iii) we propose a list of observable and measurable user-interaction features of dark patterns covering visual, process, and language design aspects, iv) we describe the attributes based on our measurable features to lower the subjectivity of users' interpretation. Finally, we discuss our proposal's cross-domain applicability and the potential for future work, including how to improve the descriptions of the attributes via semiformal languages, to generate an objective and usable framework to assess the presence of deceptive design patterns in digital interfaces.

CCS Concepts: • **Security and privacy** → **Usability in security and privacy**; • **Human-centered computing** → *HCI design and evaluation methods.*

Additional Key Words and Phrases: deceptive design, dark patterns, detection, measurable features, cookie consent, web tracking, legal design

**ACM Reference Format:**
Emre KOCYIGIT, Arianna ROSSI, and Gabriele LENZINI. 2024. A Systematic Approach for A Reliable Detection of Deceptive Design Patterns Through Measurable HCI Features. In *The 2024 European Symposium on Usable Security (EuroUSEC 2024), September 30-October 1, 2024, Karlstad, Sweden.* ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3688459.3688475

## 1 INTRODUCTION

Users' privacy and data protection requirements are frequently violated in the user experience design of websites, social media platforms, mobile applications, and video games. Online deceptive designs, *i.e.*, dark patterns, often circumvent privacy-protecting mechanisms by forcing, misleading or manipulating users to make them take potentially harmful decisions that lower, or even nullify, the protection of their personal data and weaken their rights. For example, a restrictive design pattern can be implemented in cookie consent processes to coerce users to disclose their personal data with third parties for profiling and advertisement purposes. Such a practice is explicitly against the international privacy engineering principle standardized in ISO/IEC 29100 [25] concerning "consent and choice" which clearly states that users should have the ability to control the collection, use and disclosure of their personal information. In the EU, this also contravenes the requirements about the freedom of consent enshrined by the General Data Protection Regulation (GDPR - Regulation 2016/679) and the ePrivacy Directive (Directive 2002/58).

Detecting and reporting the presence of dark patterns in digital services holds significant importance for underpinning the scrutiny of potential violations of the law by independent authorities or administrative courts, for supporting organizations in their data protection compliance efforts, and for assisting privacy consultants in their advising activities. Nevertheless, several challenges hinder the identification of dark patterns. First, dark patterns


*EuroUSEC 2024, September 30-October 1, 2024, Karlstad, Sweden*

ACM ISBN 979-8-4007-1796-3/24/09.
https://doi.org/10.1145/3688459.3688475

are present across diverse environments, encompassing mobile applications, large online platforms, Virtual Reality environments, and all sorts of websites and digital services. This necessitates a large scale examination which is broader than an analysis directed at a specific environment, as well as a comprehensive understanding of the technologies used to design deceptive interactions in various environments. Second, dark patterns can be employed in various interfaces where users perform a variety of tasks such as registration pages, cookie consent banners, payment screens, content sharing on social media, as well as flight and hotel reservations. Each scenario differs in terms of design choices: while a deceptive design pattern such as a "bad default" [18] may steer users' decision toward a privacy-invasive option to obtain their personal information in a particular scenario, in a different scenario the same pattern may sneakily add an item to their shopping cart to increase the e-commerce vendors' profit . This situation calls for multiple strategies to detect all types of dark patterns, while developing a detection tool requires use case-specific proxies which help to recognize the dark patterns. Lastly, dark patterns can adopt distinct strategies of implementation characterized by different Human Computer Interaction (HCI) design elements, such as reducing the visibility of an option, diverting user attention, imposing restrictions by eliminating the affordances for certain user actions, or manipulating language to exploit users' emotions. This emphasizes the necessity of considering multiple data types such as images, text, etc. when devising a detection solution.

Various studies have been instrumental in defining the phenomenon of deceptive design practices and in developing taxonomies to categorize and recognize them, such as [5, 18, 19, 39]. However, a precise and objective definition of dark patterns based on measurable elements that directly impact the identification process, remains elusive. To address this gap with concrete tools and quantifiable elements, we selected one of the most observed dark pattern use case (*i.e.*, the cookie consent process) with the intention of producing objective descriptions of dark pattern characteristics. With our proposal, we aim to provide a process that relies on measurable features and that can be employed by anyone to reach reliable, reproducible conclusions on the detection of dark patterns.

While devising objective instruments to assess the presence of dark patterns, we focused on established attributes (such as asymmetric, restrictive, and deceptive) which have been defined by Mathur *et al.* [39]. We believe that the definitions of the attributes are high-level abstract concepts that can be implemented in various manners. For instance, the definition of asymmetry concentrates on the unequal burdens placed on available choices [40], but what exactly is this unequal load and how it can be measured remains uncertain. Quantifiable features that are observable and measurable (*e.g.*, the presence of buttons that provide the option to grant or deny consent to all available processing purposes) are critical for describing the dark pattern attributes. To prevent ambiguity and operationalize these abstract attributes, we investigated measurable features of cookie consent processes that can objectively describe the attributes in applied settings.

The deceptive design pattern types are mostly context-specific: for example, the dark pattern type "Hidden Costs" [7] can occur in check-out or payment web pages while it does not occur in other use cases. Moreover, each pattern type is not characterized by all attributes: for instance, Mathur *et al.* presented "Forced Action", *i.e.*, a deceptive design pattern type, with the only required attribute "restriction" among the six existing ones [40]. This led us to identify the deceptive design pattern types that are present in the target use case of cookie consent banners and to propose the identification of the defining attributes of relevant types.

**Research questions.** We formulated the research questions below to meet the research objective of developing a measurable and objective process for the detection of dark patterns:

**RQ1** Which deceptive design pattern types are commonly employed in cookie consent processes?
**RQ2** Which deceptive design pattern attributes define the deceptive design types that are present in cookie consent processes?
**RQ3** What are the measurable human-computer interaction design features (*i.e.*, visual, process-based and linguistic features) in cookie consent processes that can be operationalized to define the deceptive design pattern attributes?
**RQ4** How can the dark pattern attributes be objectively described through the measurable features?

**Contributions.** With respect to the research questions, our contributions are listed below in corresponding order:

- we identify the potential privacy-impacting dark pattern types in cookie consent processes.
- we elicit the relationships between the dark pattern attributes and the dark pattern types of cookie consent processes by presenting required, optional and non-required characteristics of each type.

- we propose measurable features that consider three main design angles (namely, visual, process and language) to recognize when a cookie consent process contains a dark pattern attribute, *e.g.*, asymmetry.
- we describe the attributes based on the measurable features that we propose aiming to provide objective descriptions which can be utilized in the dark pattern detection.

The rest of the paper is structured as follows. Section 2 is a compilation of the related work about dark patterns detection. Section 3 explains the background and Section 4 clarifies the methods used to answer the research questions, while Section 5 describes our findings. Crucial aspects of the detection process such as evidence-based methods, challenges, limitations, etc. are discussed in Section 6 in addition to the results.

## 2 RELATED WORK

Dark patterns are a widespread problem in various digital applications and services such as online video games [64], mobile applications [14], social networks [43], finance applications [50], home robots [33], IoT devices [50], web pages of travel agency websites [27], Virtual Reality [30] and more. Given the plethora of classification systems that has arisen in the last few years both in academic and regulatory sources, in 2024 Gray *et al.* [18] published a taxonomy that seeks to systematize such knowledge into categories and subcategories of dark patterns. Mathur *et al.* also tried to expand the vocabulary that is available to reliably discuss the topic by identifying 5 distinct attributes [39] that can help distinguish deceptive design patterns from non-deceptive ones. Subsequently, these attributes were expanded to 6 and further enhanced [40].

Different forms of dark patterns are embedded in user interfaces to mislead or force users to take potentially harmful decisions for their privacy [8] and to collect more personal information than strictly needed, often without informing or giving choice to the user, thereby countering privacy by design principles [17], such as transparency, consent and choice, data minimization, purpose limitation and individual participation and access [61]. In particular, the act of consenting or refusing cookies and other web tracking technologies on websites is an online process that is invaded by deceptive designs [24] and has been extensively studied, also because it is a very visible and easily inspectable instance of digital manipulation. Digital nudges in cookie consent experiences sway users towards privacy-unfriendly options [21] that make them give away their personal data for profiling and ad tracking purposes. This is why the detection of these malicious designs is paramount to protect the privacy and the rights of website visitors.

*Dark patterns in cookie banners.* In the intentions of the legislator, providing notice about cookies and requiring users' consent for certain processing purposes, such as advertisement and profiling, is meant to enhance the transparency of personal data collection on websites and strengthen the decisional autonomy of individuals. However, cookie banners are often poorly implemented in web design, and often contain manipulative elements that lower, or even remove, such transparency and autonomy, as many research studies have pointed out. Such studies have sought to analyse the influence of dark pattern designs on user's choices and to check their compliance with applicable laws, such as the GDPR and the ePrivacy Directive. For instance, Nouwens *et al.* demonstrated that only 11.8% of the designs of the most common consent management platforms on UK websites fulfill the minimal legal requirements of EU data protection law [44]. Another study examined data collection consent notices' compliance with the GDPR and indicated that dark patterns (*e.g.*,"nagging" that forces user to change their consent decision) are widespread [56]. Utz *et al.* [60] showed that certain design choices (such as the position of the banner, the granularity of options and pre-selections) exert an effect on whether and how users interact with the cookie consent notices. Without the pretence of being exhaustive, other scholars incontrovertibly exhibited that cookie consent processes are full of dark patterns [2, 3, 20, 22, 49, 58]. While some focused on a specific language, *e.g.*, German [31], or a country, *e.g.*, USA [35] or UK and Greece [26], what stands out from this copious body of literature is the fact that cookie consent processes are one of the main use cases that need to be cleared of dark patterns. Indeed, the non-compliance of the implementation of manipulative practices in consent decision-making has been sanctioned by data protection authorities, who have fined companies for using obstructive patterns that make it overly difficult or even impossible to reject advertising trackers, for designing asymmetric options between granting consent and refusing it, and for adopting a wrong language in the privacy policy directed to their users and for repeatedly prompting users with unsolicited content, among the others [34, 54].

In order to detect dark patterns, they should first be recognized with concrete tools. There are qualitative and quantitative studies conducted in this context. For example, Bhoot *et al.* tried to define dark patterns with a quantitative study, *i.e.*, exploratory factor analysis, based on five different elements, such as frequency of occurrence, level of frustration etc., from the end-user perspective [36], while Maximilian addressed dark patterns

with focus groups and interviews from the same perspective [37]. Another example is a socio-technical approach which employed the Semiotic Framework to inspect interfaces with dark patterns [1].

*Measurable Elements in Cookie Consent Processes.* The endeavours that seek to discern what makes a design pattern manipulative, and even unlawful, has resorted more or less explicitly to the identification of measurable elements. On the user interface, measurable elements that can reliably indicate the presence of dark patterns include the text and the graphical elements of the notice, even though there is not a finite list of such elements - on the contrary there is much variation across approaches. For instance, Nouwens et al. [44] relied on buttons, number of clicks, and textual content (such as the list of processing purposes and vendors), while the "properties of consent notices" considered in Utz et al. [60] encompass the size, position, blocking behaviour, the design of choices, the text, the colors and pre-selections, the formatting and the links to additional information. Even though a great research focus has been placed on the UI design elements, the manipulative potential of language has also been exposed. For example, Santos et al. [55] carried out a linguistic analysis and showed that positive or negative framing was used in a third of their sample to persuade users to accept cookies, confirmed by Kampanos and Shahandashti [26] who proved that positive sentiment was present in around 80% of their sample of UK and Greece banner texts. Some automated detection approaches relied on both textual and graphical elements, such as position, html, text, width, height, fontsize, language, clickable elements, number of clicks, etc. for instance for assessing the compliance of consent notices [23] and for detecting dark patterns [57], or only on textual data for a machine learning classifier based on a decision tree model [6].

However, textual and graphical UI elements can only reveal certain manipulative practices, while others are hidden behind the front-end. For instance, while investigating the compliance of cookie banners with the GDPR and the ePrivacy Directive, Santos *et al.* used web tracking technologies to inspect the back-end [53]. Another study unveiled that sometimes consent is shared with third parties via the communication between the browser engine and the web servers regardless of the preference that the user expressed on the user interface [41] (*i.e.*, on the banner). Hence, it is necessary to check the back-end of the browsers, *e.g.*, searching the cookie data in the browser storage or the requests and responses managed by the browser after each user action. Structural elements of a web page such as HTML and CSS files are also easily accessible and their elements are quantifiable in cookie consent banner detection, as shown in [51]. [28] provided a list of measurable features of cookie banners by examining the interaction between the user, browser and web server, such as the total number of user actions at first visit, and the transferred file size after giving consent.

*Automated dark pattern detection.* The detection of dark patterns is becoming a serious concern and there are some empirical studies that employ manual detection [42]. However, developing semi/full automated detection is required since accurate, continuous manual assessments of the ever-changing plethora of applications and websites where dark patterns are present is simply impossible. There are studies that focus on automated dark pattern detection by utilizing user interface-related features [38]. However, for some dark patterns, developing a fully automated flow is quite challenging due to the variation of their implementation [13]. Some scholars have developed artificial intelligence-supported models to detect certain types of dark patterns, but they also highlighted that automated detection is challenging because dark patterns differ in terms of data type, *e.g.*, image, text, and design [57]. For instance, Gundelach and Hermann proposed an automated tool, *i.e.*,"Cookiescanner", that used a transformer architecture-based model to detect forced action by checking the option for refusing cookies, which reinforced a previous study's arguments about the many challenges of dark pattern detection [23]. Dark pattern detection often goes hand-in-hand with data protection compliance assessment. For example, Matte *et al.* [41] crawled thousands of European websites and showed that nearly half of the test set contained at least one potential violation of the data protection obligations on consent.

Artificial intelligence-based tools that are built on traditional machine learning models, computer vision and natural language processing techniques are commonly used in the detection models and enhance their performance scores [10]. For example, Yada *et al.* used well-known transformer architecture-based models such as BERT, RoBERTa etc. to detect dark patterns in e-commerce websites [63] after enhancing the dataset that was built by Mathur *et al.* in [39]. The automated detection of dark patterns is an ongoing research field that needs more comprehensive solutions. For example, while one study focused on the detection of dark patterns considering blind screen-reader users [29], another one developed the browser extension "Dark Pattern Detector" focusing on three widespread dark pattern types which are "Hidden Costs", "Disguised Ads", and "Sneark into Basket" [62].

# 3 TERMINOLOGY

Before examining the methods and results of our study, it is necessary to clearly define some of the terms that we use frequently in the article, such as *type*, *attribute* and *feature* (composed of *entity* and *metric*), to avoid any confusion.

## 3.1 Type

Although dark patterns have the similar objective of influencing users' behaviors online, they are implemented in a variety of forms that can trick individuals, *e.g.*, by hiding controls and information from them, forcing them to take predefined decisions or steering their actions towards desired outcomes in a predictable manner. The term "type" refers to a specific category of dark pattern within a system of classification that considers the tactics they employ and the scenarios where they can be found. For instance, there are 16 types in [7], such as "confirmshaming", "disguised ads", "forced action", "hidden costs", etc. However, there are many regulatory reports and academic papers that list and categorize dark pattern types. For instance, the European Data Protection Board's latest guidelines on dark patterns [15] listed 6 family types, *e.g.*, "overloading", "skipping", etc., and 16 subcategories such as "continuous prompting", "privacy maze", "dead end", "emotional steering", etc. . In order to discuss the types in our study, we used a recent comprehensive taxonomy published by Gray *et al.* [18] as our main source (refer to Section 4.1.1 for the reasons behind this decision).

## 3.2 Attribute

This term refers to the characteristics of dark patterns that are not dark pattern type-specific, but rather represent general properties that characterize online deceptive designs. One dark pattern type can be described through one or more attributes. Mathur et. al. first specified 5 attributes [39], and then added a sixth one in a following study [40]. We based our study on these attributes, namely *Asymmetric*, *Restrictive*, *Covert*, *Deceptive*, *Information hidden* and *Disparate treatment* (the definitions are given in Table 1). In the following sections, the attributes will be written in *italics*.

Table 1. Dark pattern attributes with their descriptions from Mathur et al. [40]

| **Attribute** | **Description** |
|---:|---|
| Asymmetry | Unequal burdens on the choices available to the user |
| Restriction | Eliminate certain choices that should be available to users |
| Information hidden | Obscure or delay the presentation of necessary information to users |
| Covert | Hiding the influence mechanism from users |
| Deception | Induce false beliefs in users either through affirmative misstatements, misleading statements, or omissions |
| Disparate Treatment | Disadvantage and treat one group of users differently from another |

## 3.3 Feature

To define features, entity and metric need to be described first. **Entity**: An entity refers to the specific element within the web design that is measurable. For example, 'full consent grant button' indicates whether there is a button that, once clicked, signifies user's consent to all processing purposes. This is one of the measurable elements on the cookie consent banner and can be detected through CSS components of a web page as was done in [51]. Another example of an entity is 'full consent grant process' which identifies the path that the user needs to follow to provide consent to all processing purposes. This entity can be observed manually or via semi-full automated tools, crawlers, etc. as performed in the large scale analysis of cookie notices in [6].

**Metric**: A metric indicates and measures a quantitative aspect of the entity. For instance, the entity "full consent grant button" can be evaluated via different metrics such as 'size', 'color contrast', 'position' etc. The first one simply calculates the height and width of the entity in pixels. The second one focuses on the color values, *e.g.*, RGB, of the entity and its surrounding, while the last one can measure the euclidean distance between the position of the entity and the center of the cookie consent banner. In other words, a metric can be conceptualized as a function where the entity serves as the input. The output of this function should be the same across different evaluators, thereby providing an objective assessment of the entity.

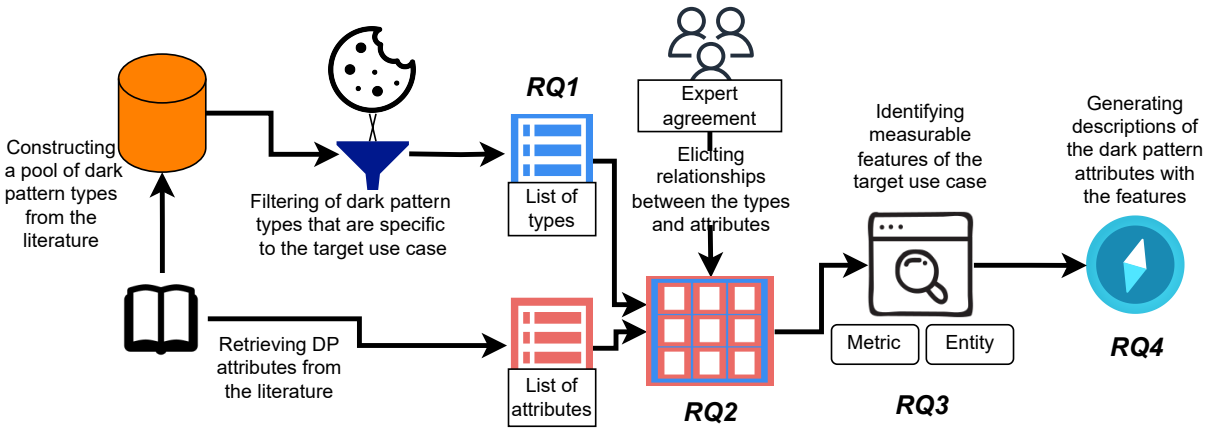


Fig. 1. Methodology workflow by exhibiting the steps and the relevant research questions.

**Feature**: An entity measured on a predefined metric is a feature. Features are helpful to describe the dark pattern attributes. For example, 'full consent grant button' is an entity and 'size' is a metric. 'The size of the full consent grant button' is a feature which is quantifiable and measurable (and can be used to identify the presence of an attribute, as we will show later). Features refer to various aspects of the human-computer interaction (HCI), including visual/graphical, process-based and linguistic/text-based aspects. For example, while the 'color contrast of the full consent grant button' represents a visual feature, the language used in the button can be featured as 'readability of the full consent grant language' as a linguistic feature, and the user's interaction with the system can be featured as 'number of clicks for full consent grant process' as a process-based feature. Methods and instruments to quantify the features differ for each of them. An example of a feature is shown in Figure 2 by annotating entities and utilized metrics.

## 4 METHODOLOGY

We follow a simple workflow composed of four steps as shown in Figure 1.

### 4.1 Identifying dark patterns in a cookie consent process

Before understanding the characteristics of deceptive design patterns in the cookie consent process, the identification of the types that are specific to that use case is crucial because (i) the detection of deceptive design patterns requires identifying measurable elements that are related to the specific use case and (ii) each use case is characterized by its own design elements. This is why searching for a "Countdown Timer" in the cookie consent processes is pointless. Therefore, we identified the specific deceptive design pattern types that can be present in our target use case as a first step, to answer **RQ1**.

*4.1.1 Construction of a pool of deceptive design pattern types.* The dark pattern literature is rife with taxonomies and categorizations. Such an abundance causes problems such as different labels associated with the same design pattern and varying levels of granularity in the categorization. For instance, the pattern based on constantly interrupting users' tasks through prompts is called "Nagging" in the European Commission's report [16], while it is named "Continuous prompting" in the European Data Protection Board's guidelines [15]. Recently, Gray *et al.*, who are well-known scholars of the deceptive design research community, published an ontology that maps deceptive design pattern types contained in the academic literature and authoritative sources' reports into a hierarchical table [18]. We used this study as the main source for the construction of the deceptive design types pool because it is the most up-to-date and comprehensive on the matter (*e.g.*, encompassing dark patterns relevant to EU data protection law [15] and US consumer protection law [11]), while the mapping table solves the conflict of different naming conventions. The table has 64 patterns in total .

It is paramount to consider granularity for assessing the presence of deceptive design patterns because the detection task requires descriptions that are as specific as possible, as opposed to abstract ones, as Mathur *et al.* highlighted in their cornerstone study [40]. Moreover, the attributes that characterize the low-level (*i.e.*, the most granular) patterns do not necessarily correspond to the attributes that characterize the related higher level (*i.e.*, the

less granular) patterns. To enable the detection of specific instances, it is necessary to accurately characterize the attributes that help recognizing why a certain design pattern is deceptive - or not. In light of these considerations, we decided to employ the most granular available types to construct a pool of deceptive design pattern types.

*4.1.2 Filtering of deceptive design pattern types that are specific to the target use case.* Only a selection of dark patterns among the 43 types applies to the cookie consent process. Independently of each other, three experts, who have composite interdisciplinary backgrounds covering computer science (software development and artificial intelligence), data protection law, human aspects of privacy and security, and linguistics, and who have a research track on privacy-invasive design patterns, selected the dark patterns that are relevant for cookie consent processes and filtered out the irrelevant ones. Each expert was asked to perform a binary classification on the 43 types, where they selected the patterns they considered fitting and excluded the others. For each deceptive design pattern type, its definition from the main source [18], as well as an example from reports [11, 12, 15, 16, 45] or papers [5, 33] were given to the experts, together with criteria for exclusion and inclusion. To exclude a type, it had to be solely related to (a) purchasing decisions, (b) ways of spending time and attention, (c) implemented in social robots or (d) proper of the privacy domain but not applicable to a cookie consent decision. If the expert assessed that the type under consideration did not meet any of the exclusion criteria, the expert evaluated whether the relevant type was a plausible fit for the target use case, *i.e.*,"cookie consent process." The voting results of the three experts were combined to compose the final list that consists of the types that achieved full consensus. Agreement through internal discussions was sought for the three cases where there was no full consensus.

## 4.2 The attributes of deceptive design patterns in target use case

Contrary to the prosperous work concerning the establishment of types and taxonomies, there has been a limited focus on defining overarching deceptive design pattern attributes. In this two-round step, first, three experts independently evaluated whether each filtered dark pattern type could be described with one or more of the 6 selected attributes, to answer **RQ2**.

Two of the three experts also performed the previous filtering of types (see Section 4.1), with the third expert additionally providing a domain knowledge in formal methods and logic, in addition to human aspects of privacy and security. During this assessment, the relationship between the type and the attribute was expressed as - `Attribute is CODE in TYPE` -, where `CODE` could take one of three values: "*required*," "*optional*," or "*not required*". Definitions of the types and the attributes, as well as one example to validate their understanding, were shared with the experts. A total of 112 evaluations were conducted in the first round, covering 17 types and 6 attributes by each expert. When combined, each evaluation can result in a label representing one of four categories:

- "`Consensus`": all experts agreed on one of three CODE;
- "`Majority decision without conflict`": agreement only between two experts on "*required*" or "*not required*", with the third expert indicating "*optional*";
- "`Majority decision with conflict`": agreement only between two experts on a "*not required*" code, with the third indicating "*required*", or viceversa;
- "`High uncertainty`" each expert provides a different evaluation (*i.e.*, "*required*," "*not required*", "*optional*").

The second round was aimed at solving the conflicts classified as "`High uncertainty`" and "`Majority decision with conflict`" and reach consensus. Three sessions for a total of 10 hours enabled the mutual sharing of ideas, the discussion of definitions, and the provision of arguments and counterarguments, after which the three experts reviewed their votes and made their final decisions to solve any conflict or high uncertainty cases. Given that each evaluation is categorized into three discrete values — "*required*," "*not required*", "*optional*" - and the participation of three experts, the permutations amount to 3 over 3, resulting in 27 possible combinations. Considering 112 evaluations, the total number of possible scenarios is 3024. Since consensus can be provided in 3 ways (*i.e.*,"required-required-required", "optional-optional-optional", or "not required-not required-not required"), the probability of reaching consensus for one evaluation is 11,11%, while the probability for full consensus is $3/27^{112}$, which is close to zero. For this reason, experts aimed to solve conflicts during the second round.

## 4.3 Identifying measurable features of cookie consent processes

This stage of the study answers **RQ3**. For doing so, we identified the measurable `feature` concept by clarifying its components, *i.e.*, `entity` and `metric`, which were explained in Section 3. The necessity of this notation is

better explained through an example. Habib *et al.* presented “readability of the notice” as a design parameter that can be built by different sub-elements such as “fonts”, “colors”, “contrast” etc. [24]. This parameter can be named “Readability of Privacy Notice Language” and classified as a `feature` according to our conceptualization. The privacy notice text, *i.e.*, the language, is a measurable element that can be classified as an `entity`, and its readability is a (linguistic) `metric` which reflects whether the text is in plain language. “Readability of Privacy Notice Banner” can be classified as another `feature` that utilized the banner as an `entity`. Moreover, the readability here is a different `metric` and concerns the contrast, which is a visual design element. As explained in these examples, the notation we propose intends to reduce potential ambiguity and help the formalization of deceptive design patterns in the next step. As a consequence, the conceptual framework expressed by the equation below will be employed in defining measurable features of the cookie consent processes.

$$\texttt{Feature} = \texttt{Metric(Entity)}$$

We extracted a set of features after reviewing [24, 45, 59, 54, 56, 9, 51], with specific attention to the measurable features of cookie banners provided in [29], that we validated and enhanced through the authors’ collaborative analysis of some examples of cookie banners. We thus decided to group the entities (*i.e.*, the measurable elements) into three categories: **visual features** (*e.g.*, “full consent grant button”), **linguistic features** (*e.g.*, “cookie policy text”) and **process-based features** (*e.g.*, “full consent grant user path”). An example of the method for extracting and defining the measurable elements of a cookie consent process is shown in Figure 2. As stated in Section 3, buttons are evident measurable design elements and the "agree and close" button on the cookie banner in Figure 2 can be identified as an `entity`. This `entity` can be measured via different `metrics`, *e.g.*, “background color”, “size”, “position” etc., and this metric-entity pairs describe visual features such as “background color of full consent grant button”. “Readability of cookie consent banner language” is a linguistic feature that is composed of the text on the banner, *i.e.*, an `entity`, and a `metric` such as “readability”. The “Learn more” button on the first layer of the cookie banner can direct the user to the second layer as shown in Figure 2, and the user path, *i.e.*, the process, can be extracted as an `entity`. “Number of clicks”, “time” etc. can be `metric` for this `entity`, and together they identify a process-based feature. With this approach, we extracted a finite list of linguistic, visual, process-based features in Section 5 that can be expanded at will as new examples and new features are analyzed.

### 4.4 Describing the deceptive design pattern attributes based on the features

In order to define the attributes of deceptive design patterns with observable and measurable features, they should be expressed as within a certain logic (*e.g.*, as a rule), taking into account the relationship between the features. For instance, the features "size of the full consent grant button" and "size of the full consent refusal button" do not reveal the presence of dark patterns. However, a logical expression can elicit the relationship between them, such as "if the size of the full consent grant button is not equal to the size of the full consent refusal button”, indicates an asymmetric pattern, which is one of the attributes of dark patterns. In this regards, we defined rules using the features determined by the methods described in the previous section to minimize subjective interpretations. Eventually, the outputs of this step answer the last research question, *i.e.*, **RQ4**.

## 5 IDENTIFYING, MAPPING AND DESCRIBING DARK PATTERNS

### 5.1 Identified Dark Pattern Types in the Cookie Consent Process

Following the process described in Section 4.1, we identified 17 deceptive design pattern types that can be found in cookie consent processes out of the 43 types within the pool of deceptive design pattern types. The types and their definitions can be seen in Table 2, which is not a definitive list since other, or even novel, types of dark patterns can influence consent decision-making as new technologies and scenarios arise. While "Privacy Zuckering” was included with the majority voting, three experts unanimously selected the other 16 types. The "Forced Continuity” and "Personalization” types were selected by only one expert, and have been excluded after discussion. The remaining 24 types were not included in the list with full consensus. The filtering process’ inter-rater agreement score is nearly 0.926 according to Cohen’s Kappa metric. While there are 43 types in total in the pool, more than one third (39.5%) are relevant for cookie consent processes.

### 5.2 Identifying the Attributes Characterizing Deceptive Design Patterns Types

We followed the methodology detailed in Section 4.2 to associate the attributes with the 17 types that were selected in the previous step. The detailed results of the mapping between the two is reported in Figure 3.

Fig. 2. Visual, linguistic and process-based features extraction example from a cookie consent process

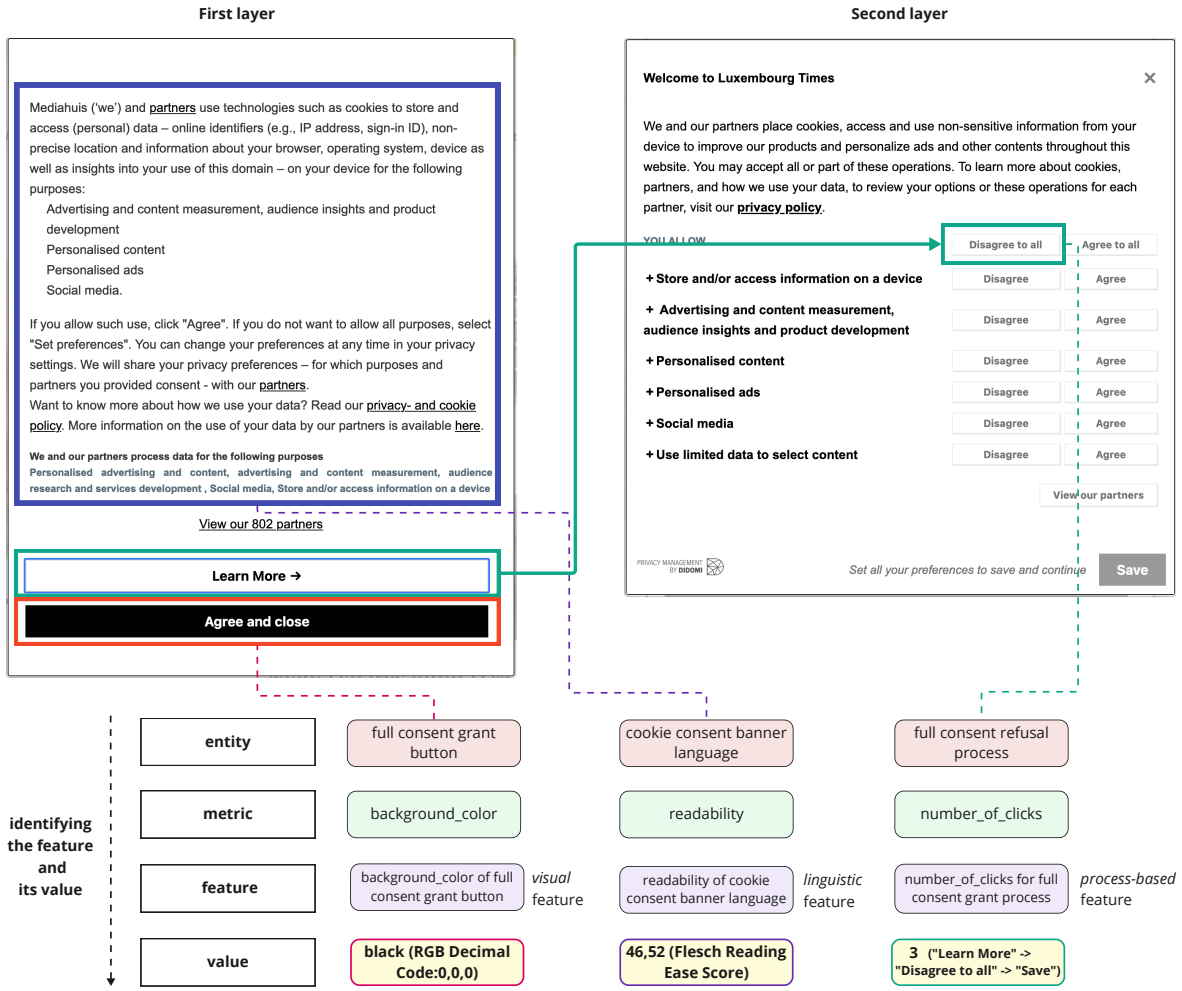


Table 2. Dark Pattern Types Used in Cookie Consent Processes (the Definitions are Given in Table 3, in Appendix)

| | |
|---|---|
| Dead End | Privacy Maze |
| Nagging | Wrong Language |
| Complex Language | Confirmshaming |
| Information without Context | Hidden Information |
| Positive-Negative Framing | Conflicting Information |
| False Hierarchy | Bad Defaults |
| Visual Prominence | Feedforward Ambiguity |
| Trick Questions | Choice Overload |
| Privacy Zuckering | |

Our first finding was the difficulty of the process of eliciting the relationships between the types and the attributes that we will discuss in Section 6. This can be seen on the results of Round 1, which are shown in Figure 4, in appendix. There was `Consensus` on 53 out of 119 evaluations (44.54%), 25 (21%) conflicts (encompassing `Majority decisions with conflict` and `High uncertainty`), and 41 (34.45%) `Majority decisions without conflicts`. Expert users pointed out that the definitions of the attributes are so broad that utilizing them for a concrete evaluation is hard. This challenge was also observed in the executive discussions held after the

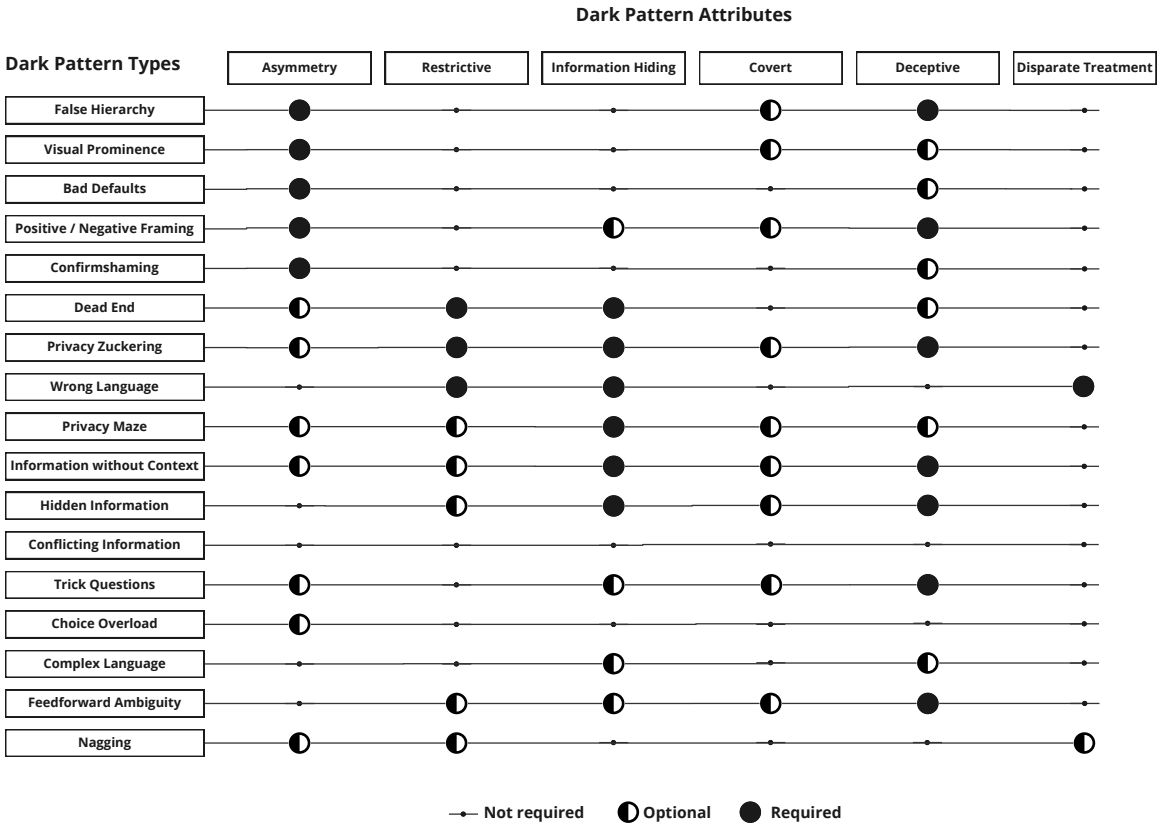


Fig. 3. Association of dark pattern attributes with dark pattern types that are present in cookie consent processes

first round, as it took 10 hours of intense dialogical exchanges to eliminate all conflicts. Eventually, we concluded the experiment with `Consensus` on 89 evaluations (74,79%) and 30 (25.21%) `Majority decisions without conflicts`.

The overall findings of this expert analysis are:

(1) No single attribute is consistently present in all dark pattern types, even though *Deceptive* is a frequently recurring attribute across the types.
(2) No dark pattern type includes all attributes: each type has its own characteristics.
(3) *Disparate treatment* is *required* only for "Wrong Language" and *optional* in "Nagging", *i.e.*, it is not a common attribute in cookie consent processes.
(4) Wherever there is the *Covert* attribute (always *optional*), the *Deceptive* attribute is also present.
(5) "Conflicting Information" is not associated with any of the existing attributes. "Complex Language" and "Choice Overload" also lack the required attributes. We believe this indicates that the current list of the attributes is not sufficient to represent the characteristics of all dark patterns (for instance, *Complexity* could be added, *i.e.*, unnecessary cognitive load for users). We discussed this issue in detail in Section 6.

### 5.3 Defining Measurable Features of Dark Patterns

An attribute can appear in more than one form. For instance, asymmetric design can be based on graphical elements (*e.g.*, full consent grant (FCG) and full consent refusal (FCR) buttons are not equal in terms of size), process-based elements (*e.g.*, FCG and FCR processes are not equal in terms of number of clicks), or linguistic elements (*e.g.*, FCG and FCR options' language is not equal in terms of clarity). Following the methodology explained in Section 4.3, we defined measurable `features` for assessing whether a design pattern is potentially dark. Some examples of `features` such as "size of the FCG button", "readability of the cookie consent info text",

and "availability of the FCG process" are given with their `entities`, `metrics` and descriptions in Table 4 in appendix, considering each category: graphical, linguistic and process-based, respectively.

These `features` can be used alone or in combination to define dark pattern attributes. For instance, a common dark pattern attribute in cookie consent processes, *i.e.*, *Asymmetry*, which is a *required* attribute in five different types and *optional* in seven others, can be measured by employing the "FCG button" and "FCR button" `entities`. The comparison between two `entities` can be performed with different visual `metrics` such as "size", "color", "contrast". That is, "size of the FCG button" and "size of the FCR button" can be defined as `features` and if their values are not equal, there is an *Asymmetry* on the cookie consent banner. The *Asymmetry* can also be detected via `linguistic` or `process-based features`. For instance, for the process-based entities such as "FCG process" and "FCR process", the `metrics` can be "number of clicks", "time" etc., the values of which can be compared to detect the *Asymmetry*. The `features` can be quantitative (*e.g.*, size, number of clicks, etc.) or binary (*e.g.*, availability), according to their `metrics`. For instance, the *Restrictive* attribute can be evaluated by checking the availability of certain `entities` (*e.g.*, choices). If the "availability of FCR process" is false, this is a clear indication of a *Restrictive* pattern.

### 5.4 Measurable Feature-based Descriptions of the Dark Pattern Attributes

Without intending to be exhaustive, we enumerated a set of potential descriptions for each attribute with an easily applicable rule-based approach, as reported below. The features are shown in italic.

***Asymmetry*** can be described as:

A1 If *minimum number of clicks for the FCG process* and *minimum number of clicks for the FCR process* are not equal.
A2 If *availability of the FCG process* is True and *availability of the FCR process* is False.
A3 If *number of paths for the FCG process* and *number of paths for the FCR process* are not equal.
A4 If *size of the FCG button* and *size of the FCR button* are not equal.
A5 If *clarity of the FCG text* and *clarity of the FCR text* are not equal.

As seen in the Algorithm 1 that shows an example for A1, these descriptions are easily convertible to programmable rules, supporting automated detection and classification.

**Algorithm 1** An Example for Asymmetry Detection with Measurable Features

**Ensure:** *asymmetry* (Boolean indicating presence of asymmetry)
**Step 1: Detect all possible FCG and FCR processes**
*processesFCG* ← detectProcesses(FCG)
*processesFCR* ← detectProcesses(FCR)
**Step 2: Calculate minimum number of clicks for each process**
*minClicksFCG* ← min(clicks required for each path in *processesFCG*) ▷ Feature
*minClicksFCR* ← min(clicks required for each path in *processesFCR*) ▷ Feature
**Step 3: Initialize asymmetry variable**
*asymmetry* ← *False*
**Step 4: Check for asymmetry**
**if** *minClicksFCG* ≠ *minClicksFCR* **then**
    *asymmetry* ← *True*
**end if**

***Restriction*** can be described as:

R1 After *action status of the Consent Request (CR)* is FCR or FCG, if *availability of the CR process* is False.
R2 If *availability of the FCR process* is False.
R3 If *language of the CC information text* and *language of country the website serves* are not equal.

***Information Hidden*** can be described as:

I1 If *completeness of the CC information text* is False.
I2 If *semantic of the cookie policy heading* and *semantic of the CC information text* are similar.

***Covert*** can be described as:

C1 If *availability of the FCG icon* is True and *availability of the FCR icon* is False.

***Deception*** can be described as:

D1 If the *action status of the CC process* is Refusal and *availability of unnecessary cookies (UN)* is True.
D2 After *expiration date of the cookies* exceeded, if *availability of the CR process* is False.

***Disparate Treatment*** can be described as:

T1 After *action status of CC process* is FCG for user1 and FCR for user2, if *availability of the CRR process* is False for user1 and *availability of the CRR process* is True for user2.
T2 If *frequency of the CRR process* for user1 and *frequency of the CRR process* for user2 are not equal.

### 5.5 Implementation example

The cookie consent process example in Figure 2 was assessed to check if it has dark patterns. Firstly, *Asymmetry* was detected by comparing the "number of clicks for the FCG process" and "number of clicks for the FCR process" `features`, *i.e.*, A1 defined in the previous section. Users can perform the FCG with one click on the "Agree and close" button while they need to perform 3 clicks for the FCR through 'Learn More', 'Disagree to all' and 'Save' buttons. Therefore, the two patterns are asymmetric and the second one corresponds to the dark pattern known as "Privacy Maze". Another assessment was performed by employing the visual features such as 'size of the FCG button' and 'size of the FCR button'. Considering only the second layer, these `features` are equal, because the 'Disagree to all' and 'Agree to all' buttons' size is equal. Therefore the design is symmetric on the second layer. *Asymmetry* was also detected through visual features related to the color of the available buttons, i.e., "background color of full consent grant buttons" is different than the "background color of the consent setting button". Making a certain choice more visible than another may be labelled as "False Hierarchy". Lastly, the linguistic feature "readability of cookie consent banner language" was used to compute the clarity of the text: the Flesch Reading Ease Score equals 46,52 which corresponds to the College grade educational level needed to understand the text, a symptom of *Complexity*. Therefore it is a "Complex Language" dark pattern. The cookie consent management process in Figure 2 may also contain other dark patterns, but hereby, we merely aim at showing the concrete application of our approach to a real example.

## 6 DISCUSSION

### 6.1 Dark pattern types present in cookie consent processes

Not all dark pattern types can be found in all use cases. Different types can be combined but they are not necessarily all present at the same moment. In the cookie consent process, we have identified 17 different types of deceptive design patterns that can potentially be present, presumably related to the fact that there are great economic interests in luring users to accept web tracking for profiling and advertising purposes.

### 6.2 Attributes characterizing dark patterns

In our study, certain attributes seem to constitute the basic characteristics of certain dark pattern types. For example, *Asymmetry* is required in five deceptive design types - therefore, the assessment of those dark patterns can be carried out by solely or predominantly focusing on the asymmetric nature of certain features. Yet, only a handful of types can be detected through a single attribute. Most times, dark patterns can (and should!) be analyzed under various perspectives that go beyond the superficial level offered by the graphical user interface (reflected in the various `features` that can be examined) and in different contexts they may expose different characteristics, as the need for *optional attributes* shows. For instance, *Restrictive* and *Information hiding* are essential attributes of the "Dead End", while they are optional in "Feedforward Ambiguity". From this follows that certain dark pattern types may be easier to detect through their essential attribute(s), whereas others have a multifaceted nature that requires more complex descriptions.

***Disparate treatment*** is only required in "Wrong language" and occasionally found in "Nagging" in the case under examination. This may mean that *Disparate treatment* is not a frequent quality of the deceptive design patterns implemented on cookie banners, whereas it could be in other cases (*e.g.*, online purchases). As a methodological choice, we decided to interpret *Disparate treatment* not in terms of its outcomes (*i.e.*, whether it has a disproportionate outcome on certain users or group of users) but in terms of its implementation (*i.e.*, whether the dark pattern is implemented differently for different users or group of users). For instance, continuous prompts (*i.e.*, nagging) may be used to exasperate those users who did not consent to advertising and make their navigation more cumbersome until they agree, as a recent sanction by the Italian DPA shows [48, 54]. This

is a case of disparate treatment in the implementation. However, certain dark patterns may have more severe implications on certain users rather than others (*i.e.*, disparate treatment in terms of outcomes). For instance, "Wrong language" may disproportionately impact the people that do not master that language, while "Privacy maze" may more severely affect users with poorer digital skills. We decided to focus on the first interpretation as it is more objective to establish, even though a growing body of literature seeks to prove how certain people or groups of people may suffer from deceptive design patterns' influence and harms more severely than others [46, 52].

**Eliciting the relationships between the types and the attributes** was a complex task, one reason being that the definitions of the attributes are broad and abstract, thus using them for a concrete evaluation is challenging. Furthermore, we reached different conclusions than Mathur *et al.* [39] for certain patterns. For instance, we deliberated that "Trick question" is always *Deceptive* due to its misleading nature, while it is not labelled as such in their study. Unlike the previous study, we also argue that it can optionally be *Asymmetric*, since it may place unequal burdens on two or more options, and *Hiding information*, since it may obscure necessary information, but not necessarily. This evidence suggests that defining measurable features is a necessary work and that future discussions among experts should elaborate more specific definitions of the attributes.

**The *Covert*** attribute needs to be discussed on its own, since we couldn't find that there is no case such that a pattern is *Covert* but not *Deceptive* in our analysis: since the influence mechanism is hidden from users, covert dark patterns are always misleading (*i.e.*, deceptive). Therefore, it seems that *Covert* patterns can be categorized as a subset of *Deceptive* ones, even though further work should determine whether this conclusion applies to a broader range of use cases. The application of this attribute was lively discussed among the experts in this step as the consensus rate was generally low, likely reflecting the low usability of such an attribute for providing an objective criterion of detection of dark patterns. This mostly depends from the interpretation of "influence mechanism": does this notion refer to the influence strategy (*e.g.*, for positive/negative framing: focusing on the positive consequences of a user decision, while purposefully withholding the negative ones) or does it rather refer to the effect (*e.g.*, the swayed decision)?

***Complexity.*** Moreover, some dark pattern types cannot be characterized by any attribute that can be found in the literature. For instance, "Choice Overload" increases the number of options and thus the cognitive burden placed on users to complete the process. However, none of the attributes by [40] can account for this behavior. Therefore, we found it necessary to add *Complexity* as an attribute, which describes the additional unnecessary burden for users in terms of time, actions, cognitive effort, etc. Similarly, the European Commission proposed a behavioural taxonomy on design practices infringing consumer protection provisions [16] that considers complexity as the essential attribute of the dark patterns mentioned in the report. Cognitive load has been recognized as a harm [40] since it causes individuals to waste time, energy and attention, especially when one considers the cumulative effects of deceptive design practices. Another example that indicates the need for adding *Complexity* is that even though "Complex Language" can include *Information hiding* or *Deceptive* attributes, there can be scenarios where none of them are present. For example, all necessary information can be present in the cookie banner without any omission, but the language may be too complicated for users to enable an informed decision. Thus "information hidden" or "deception" are not constitutive elements of this deceptive design type, but complexity is. This insight leads to the hypothesis that the list of attributes may be expanded, also to account for other types of interfaces: for example, voice interfaces may embed dark patterns that play on linearity and volume of voice [47].

**Context and expertise.** Moreover, we have observed that including context and domain expertise is paramount for the assessment process. For example, a design pattern may be *Restrictive*, as it forces the user to undertake a certain course of action. However, sometimes restrictive designs are necessary, one reason being that legal obligations impose that user consent is asked before collecting their personal data for certain purposes. A design pattern may be *Restrictive* but mandated by law, thus not necessarily illegitimate, even though it is open to discussion determining whether a lawful design choice that does not serve the user interest (*e.g.*, "Choice Overload" in specific consent regimes) amounts to a dark pattern. A too narrow focus on GUI elements should also be avoided, as many dark patterns are hidden in the interaction between user and system. All these considerations further reinforce the necessity of elaborating interdisciplinary approaches for evaluating whether a design pattern is dark or not in a reliable manner. A future development of this work should seek to map these attributes to legally relevant attributes, such as the definition of abusive commercial practices or the violation of equal treatment in specific jurisdictions. For example, it should be determined what kind of asymmetry in the visibility of certain options is illegal, since a simple difference cannot be automatically labelled as dark.

### 6.3 Measurable features

The measure of certain features can indicate whether a design pattern is potentially dark, especially when multiple features hint at the (co-)presence of one or more attributes. However, certain attributes are not black or white properties, and rather have degrees. For example, *Complexity* can be measured through the number of required clicks, but there is not necessarily a threshold for such a number. Moreover, even though measurable features are critical components of the dark patterns' detection process, they vary according to the use case. The list of process-based, visual-based and linguistic features we identified may be expanded when the number and typology of examined websites is broadened. Furthermore, these features are use case-specific, thus in other use cases (*e.g.*,"e-shopping"), the features should be updated to detect the attributes in that use case.

Further, the `metrics` of the feature must be clearly defined before the feature is built to avoid measuring in different units. For example, for "readability of cookie consent (CC) info text", which is a linguistic feature, the "readability" `metrics` can be defined via different instruments such as the Flesch Reading Ease Score, the Flesch-Kincaid Grade Level, the Gunning Fog Index, etc. Another example is the "size" `metric`, which can commonly be measured in pixels with the order of width and height. Therefore, the description, scale and units of the `metrics` should be clearly defined. In addition, customized `metrics` can be defined for particular use cases. For instance, "completion status" can be a binary `metric` which can take the values of "True" and "False" for the "cookie consent process" `entity`. A crucial aspect of defining and using `metrics` is that they must generate consistent and objective results. This ensures that every implementation of the `metrics` equates to the same conclusion. Moreover, it should be considered that selection metric is crucial because different metrics can vary in their effectiveness depending on the context. For instance, when measuring the "clarity" metric of the text, various readability metrics can be employed, such as the Gunning Fog Index or the Flesch Reading Scale.

### 6.4 Relevance

While our focus has been on the cookie consent process, the method employed in this study can serve as a blueprint for investigating dark patterns in other use cases. For example it can be employed in any consent management UI (beyond cookies), in privacy settings of digital services as well in service registration processes or e-commerce check-outs (where asymmetries and restrictions are common). It can also be applied to uncover information hidden dark patterns in e.g., legal documents such as privacy policies and consumer contracts. Even though this work has focused on GUIs, further work that defines the features of e.g., voice interactions could leverage our method to determine with higher certainty if those interactions are manipulative. The potential applications are many and go beyond these few examples.

Studies like ours will reinforce the world of bright design patterns in privacy [59], because they ultimately aim at establishing a set of requirements to avoid dark pattern design attributes. If, for instance, it is established that *Symmetry* is a required attribute for design patterns that counter *Asymmetric* dark patterns, it becomes possible to establish clear, tangible requirements for designers and developers, thereby going beyond predominant current research on dark patterns that exclusively "tell[s] a designer what not to do, when a designer is usually seeking advice on what to do" [9, p.1]. However, it is important to lay down requirements that offer lawful guidance but do not deviate from what empirical research demonstrates about the effect of certain design choices on users. The two do not always coincide, as the gaps identified by Bielova *et al.* [4] show.

We stress the necessity of developing an easy, reliable, (semi-)automatized detection of dark patterns which may enable fairer interactions in digital services. Our work, *i.e.*, both the precise descriptions of the extracted features and the rule-based descriptions of the dark pattern attributes, can be utilized in automated dark pattern detection tools such as artificial intelligence-based and rule-based models. For instance, our feature concept can be employed in the feature engineering step of machine learning pipelines to automatically detect dark patterns in cookie consent processes.

Describing dark patterns in a feature-based manner can facilitate the work of various expert users: academic researchers that study the effects of design elements on user behaviour and that develop automated approaches to dark pattern detection; designers and developers that need tangible "how-to" guidance in their everyday work, rather than abstract principles and no-gos; supervisory authorities that inspect interfaces to investigate the compliance of businesses and provide guidance on technology design by interpreting the law; as well as civil society organizations that scrutinize and denounce problematic design practices. Observable HCI features and objective descriptions can represent a common language for enabling strategic collaboration across these sectors and across various domains (computer science, law, and UX/UI design).

### 6.5 Future improvements.

Natural languages, *e.g.*, English, are powerful in terms of expressiveness, but their complexity and their inherent semantic ambiguity can generate unclear, ambivalent descriptions. Therefore, while describing the dark pattern attributes or defining rules that assess the presence of dark patterns, formal or semi-formal languages can be employed to bring precision and clarity. Considering the multidisciplinary nature of the dark pattern research field, a "Controlled Natural Language", which is placed somewhere between natural languages and formal languages (*e.g.*, First-Order Logic) [32], could be adapted to further define the descriptions of the dark pattern attributes, because it is close to human natural language and allows everyone to reach the same conclusion on the presence of dark patterns by eliminating ambiguity, instead of a formal language that is technically difficult to understand and thus requires expertise. Such languages can be used to define requirements which instead of specifying bad practices (*i.e.*, dark patterns), they specify good practices what should be implemented (*i.e.*, fair patterns or bright patterns) and then detect the deviation from such ground truth to identify dark patterns. In both cases, the same features that we defined in this study can be used. However, first, best practices should be established and widely accepted by the community, which is a matter of current discussions.

## 7 LIMITATIONS

In this work, we identified a set of useful GUI features without striving for completeness. Beyond the GUI related features that we defined (following [29], the ones related to human-computer interaction), future work should include machine-to-machine interaction features (e.g., size, typology, expiration date of cookies) and maybe other features that are related to the effect of certain designs on users (cognitive features?) that could be useful for detecting dark patterns. As Mathur *et al.* stated in the "What makes a dark pattern dark pattern" study [40], actionable dark pattern definitions require as much specificity as possible, and actual use cases should be taken into consideration while increasing the specificity. In this study, we carried out feature extraction and dark pattern attribute description focusing on the "Cookie Consent Process" use case, which could limit the aspects of examining dark patterns. Moreover, as we have found necessary to add “complexity” to the set of attributes, it may be that other attributes should be identified, especially when it comes to non-GUI dark patterns, such as the linearity in voice interactions [47].

## 8 CONCLUSION

As deceptive designs patterns routinely and widely violate users’ privacy, their detection becomes necessary and urgent. In order to detect dark patterns, we first identified the potential dark pattern types present in a target use case which was the cookie consent process in our systematic approach. Then we established the relationship between the types and the attributes, showing that only in certain cases they are essential to types, but more in general they characterize the types in combination. To enable objective dark pattern detection, we extracted measurable visual, process and linguistic features to describe them. We also proposed a concept which clarifies the components of the features such as the entity and the metric to provide precise measurable descriptions. Finally, we described the attributes through rule-based descriptions using the measurable features, aiming to minimize ambiguity. We discussed implications and limitations of this approach. Furthermore, semi-formal or formal languages can be employed to eliminate any subjective interpretations of the descriptions in future work. We believe that after applying these objective descriptions on a large scale and assessing their applicability and limits, they can be employed to build datasets-an ongoing effort of ours-to detect dark patterns. Furthermore, these descriptions and measurable features can be implemented into automated applications to reliably prove their presence.

## ACKNOWLEDGMENTS

This paper is published as a part of the project Decepticon (grant no. IS/14717072) supported by the Luxembourg National Research Fund (FNR). This work was also funded by the PNRR/NextGenerationEU project "Biorobotics Research and Innovation Engineering Facilities “IR0000036” – CUP J13C22000400007". We would like to thank Cristiana Teixeira Santos for her contribution to filtering deceptive design pattern types of the cookie consent use case. We are also grateful to the reviewers and the colleagues with whom we discussed earlier versions of this work.

## REFERENCES


[1] Luiz Adolpho Baroni, Alisson Andrey Puska, Luciana Cardoso de Castro Salgado, and Roberto Pereira. 2021. Dark Patterns: Towards a Socio-Technical Approach. In *Proceedings of the XX Brazilian Symposium on Human Factors in Computing Systems* (Virtual Event, Brazil) *(IHC '21)*. Association for Computing Machinery, New York, NY, USA, Article 15, 7 pages. https://doi.org/10.1145/3472301.3484336

[2] Benjamin Maximilian Berens, Mark Bohlender, Heike Dietmann, Chiara Krisam, Oksana Kulyk, and Melanie Volkamer. 2024. Cookie disclaimers: Dark patterns and lack of transparency. *Computers & Security* 136 (2024), 103507.

[3] Carlos Bermejo Fernandez, Dimitris Chatzopoulos, Dimitrios Papadopoulos, and Pan Hui. 2021. This website uses nudging: Mturk workers' behaviour on cookie consent notices. *Proceedings of the ACM on human-computer interaction* 5, CSCW2 (2021), 1–22.

[4] Nataliia Bielova, Cristiana Santos, and Colin M. Gray. 2024. Two worlds apart! Closing the gap between regulating EU consent and user studies. *Harvard Journal of Law & Technology* 37 (2024), 1295–1333. https://jolt.law.harvard.edu/assets/articlePDFs/v37/Symposium-12-Bielova-Santos-Gray-Two-Worlds-Apart-Closing-the-Gap-Between-Regulating-EU-Consent-and-User-Studies.pdf

[5] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 2016. Tales from the dark side: privacy dark strategies and privacy dark patterns. *Proc. Priv. Enhancing Technol.* 2016, 4 (2016), 237–254.

[6] Ahmed Bouhoula, Karel Kubicek, Amit Zac, Carlos Cotrini, and David Basin. 2023. Automated, Large-Scale Analysis of Cookie Notice Compliance. In *33rd USENIX Security Symposium (USENIX Security 24)*.

[7] Harry Brignull. 2010. Formerly darkpatterns.org. https://www.deceptive.design. Last accessed: 25 August 2024.

[8] Corina Cara et al. 2019. Dark patterns in the media: A systematic review. *Network Intelligence Studies* 7, 14 (2019), 105–113.

[9] Evan Caragay, Katherine Xiong, Jonathan Zong, and Daniel Jackson. 2024. Beyond Dark Patterns: A Concept-Based Framework for Ethical Software Design. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–16. https://doi.org/10.1145/3613904.3642781

[10] Jieshan Chen, Jiamou Sun, Sidong Feng, Zhenchang Xing, Qinghua Lu, Xiwei Xu, and Chunyang Chen. 2023. Unveiling the Tricks: Automated Detection of Dark Patterns in Mobile Applications. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–20. https://doi.org/10.1145/3586183.3606783

[11] Federal Trade Commission. 2022. Bringing Dark Patterns to Light. https://www.ftc.gov/system/files/ftc_gov/pdf/P214800%20Dark%20Patterns%20Report%209.14.2022%20-%20FINAL.pdf

[12] Competition and Markets Authority. 2022. Evidence review of Online Choice Architecture and consumer and competition harm. https://www.gov.uk/government/publications/online-choice-architecture-how-digital-design-can-harm-competition-and-consumers/evidence-review-of-online-choice-architecture-and-consumer-and-competition-harm#taxonomy-of-oca .

[13] Andrea Curley, Dympna O'Sullivan, Damian Gordon, Brendan Tierney, and Ioannis Stavrakakis. 2021. The design of a framework for the detection of web-based dark patterns. (2021).

[14] Linda Di Geronimo, Larissa Braz, Enrico Fregnan, Fabio Palomba, and Alberto Bacchelli. 2020. UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3313831.3376600

[15] EDPB. 2023. Guidelines 03/2022 on Deceptive design patterns in social media platform interfaces: how to recognise and avoid them - version 2.0. https://edpb.europa.eu/system/files/2023-02/edpb_03-2022_guidelines_on_deceptive_design_patterns_in_social_media_platform_interfaces_v2_en_0.pdf. [Accessed 10-11-2023].

[16] Directorate-General for Justice, Consumers (European Commission), Francisco Lupiáñez-Villanueva, Alba Boluda, Francesco Bogliacino, Giovanni Liva, Lucie Lechardoy, and Teresa Rodríguez de las Heras Ballell. 2022. *Behavioural study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation: final report.* Publications Office of the European Union, LU. https://data.europa.eu/doi/10.2838/859030

[17] Lothar Fritsch. 2017. Privacy dark patterns in identity management. In *Open Identity Summit (OID), 5-6 october 2017, Karlstad, Sweden*. Gesellschaft für Informatik, Karlstad, 93–104.

[18] Colin M Gray, Nataliia Bielova, Cristiana Santos, and Thomas Mildner. 2024. An Ontology of Dark Patterns: Foundations, Definitions, and a Structure for Transdisciplinary Action. In *CHI 2024 Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–22. https://doi.org/10.1145/3613904.3642436

[19] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–14. https://doi.org/10.1145/3173574.3174108

[20] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–18. https://doi.org/10.1145/3411764.3445779 arXiv: 2009.10194.

[21] Paul Graßl, Hanna Schraffenberger, Frederik Zuiderveen Borgesius, and Moniek Buijzen. 2021. Dark and Bright Patterns in Cookie Consent Requests. *Journal of Digital Social Research* 3, 1 (2021), 1–38.

[22] Johanna Gunawan, Cristiana Santos, and Irene Kamara. 2022. Redress for Dark Patterns Privacy Harms? A Case Study on Consent Interactions. In *Proceedings of the 2022 Symposium on Computer Science and Law*. ACM, Washington DC USA, 181–194. https://doi.org/10.1145/3511265.3550448

[23] Ralf Gundelach and Dominik Herrmann. 2023. Cookiescanner: An Automated Tool for Detecting and Evaluating GDPR Consent Notices on Websites. In *Proceedings of the 18th International Conference on Availability, Reliability and Security*. ACM, Benevento Italy, 1–8. https://doi.org/10.1145/3600160.3605000

[24] Hana Habib, Megan Li, Ellie Young, and Lorrie Cranor. 2022. "Okay, whatever": An Evaluation of Cookie Consent Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New Orleans LA USA, 1–27. https://doi.org/10.1145/3491102.3501985

[25] ISO. 2011. *Information technology - Security techniques - Privacy framework*. Standard. International Organization for Standardization, Geneva, CH.

[26] Georgios Kampanos and Siamak F. Shahandashti. 2021. Accept all: The landscape of cookie banners in Greece and the UK. In *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer, 213–227.

[27] Woo Gon Kim, Souji Gopalakrishna Pillai, Kavitha Haldorai, and Wasim Ahmad. 2021. Dark patterns used by online travel agency websites. *Annals of tourism research* 88 (2021), 1–6.

[28] Emre Kocyigit, Arianna Rossi, and Gabriele Lenzini. 2022. Towards Assessing Features of Dark Patterns in Cookie Consent Processes. In *IFIP International Summer School on Privacy and Identity Management*. Springer, Online, 165–183.

[29] Satwik Ram Kodandaram, Mohan Sunkara, Sampath Jayarathna, and Vikas Ashok. 2023. Detecting Deceptive Dark-Pattern Web Advertisements for Blind Screen-Reader Users. *Journal of Imaging* 9, 11 (2023). https://doi.org/10.3390/jimaging9110239

[30] Veronika Krauss, Pejman Saeghe, Alexander Boden, Mohamed Khamis, Mark McGill, Jan Gugenheimer, and Michael Nebeling. 2024. What makes XR dark? Examining emerging dark patterns in augmented and virtual reality through expert co-design. *ACM Transactions on Computer-Human Interaction* (2024).

[31] Chiara Krisam, Heike Dietmann, Melanie Volkamer, and Oksana Kulyk. 2021. Dark patterns in the wild: Review of cookie disclaimer designs on top 500 German websites. In *Proceedings of the 2021 European Symposium on Usable Security*. ACM, Karlsruhe Germany, 1–8.

[32] Tobias Kuhn. 2014. A survey and classification of controlled natural languages. *Computational linguistics* 40, 1 (2014), 121–170.

[33] Cherie Lacey and Catherine Caudwell. 2019. Cuteness as a 'dark pattern'in home robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, ACM, Daegu Republic of Korea, 374–381.

[34] Mark Leiser and Cristiana Santos. 2024. Dark Patterns, Enforcement, and the Emerging Digital Design Acquis: Manipulation beneath the Interface. *European Journal of Law and Technology* 15, 1 (2024).

[35] Danyang Li. 2022. The FTC and the CPRA's Regulation of Dark Patterns in Cookie Consent Notices. *The University of Chicago Business Law Review* 1, 1 (2022), 19.

[36] Aditi M. Bhoot, Mayuri A. Shinde, and Wricha P. Mishra. 2020. Towards the identification of dark patterns: An analysis based on end-user reactions. In *Proceedings of the 11th Indian Conference on Human-Computer Interaction*. ACM, Online India, 24–33.

[37] Maximilian Maier. 2019. *Dark patterns – An end user perspective*. Ph. D. Dissertation. https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-160937

[38] SM Hasan Mansur, Sabiha Salma, Damilola Awofisayo, and Kevin Moran. 2023. Aidui: Toward automated recognition of dark patterns in user interfaces. In *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*. IEEE, ACM, Melbourne, Australia, 1958–1970.

[39] Arunesh Mathur, Gunes Acar, Michael J Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–32.

[40] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. 2021. What makes a dark pattern... dark? design attributes, normative considerations, and measurement methods. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. ACM, Yokohama Japan, 1–18.

[41] Célestin Matte, Nataliia Bielova, and Cristiana Santos. 2020. Do cookie banners respect my choice?: Measuring legal compliance of banners from iab europe's transparency and consent framework. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, IEEE, Online, 791–809.

[42] Thomas Mildner, Merle Freye, Gian-Luca Savino, Philip R. Doyle, Benjamin R. Cowan, and Rainer Malaka. 2023. Defending Against the Dark Arts: Recognising Dark Patterns in Social Media. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, Philadelphia USA, 2362–2374. https://doi.org/10.1145/3563657.3595964

[43] Thomas Mildner and Gian-Luca Savino. 2021. Ethical user interfaces: Exploring the effects of dark patterns on facebook. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yojohama Japan, 1–7.

[44] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. 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*. ACM, Honolulu USA, 1–13.

[45] OECD. 2022. Dark Commercial Patterns. https://www.oecd-ilibrary.org/docserver/44f5e846-en.pdf?expires=1707456299&id=id&accname=guest&checksum=063A07EB53611E9EB1941F5DECEF38C3. No 336.

[46] OECD. 2023. *Consumer vulnerability in the digital age*. Number 355 in OECD Digital Economy Papers. Paris. https://www.oecd.org/publications/consumer-vulnerability-in-the-digital-age-4d013cc5-en.htm

[47] Kentrell Owens, Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Tadayoshi Kohno, and Franziska Roesner. 2022. Exploring Deceptive Design Patterns in Voice Interfaces. In *2022 European Symposium on Usable Security*. ACM, Karlsruhe Germany, 64–78. https://doi.org/10.1145/3549015.3554213

[48] Garante per la Protezione dei Dati Personali. 2023. Provvedimento prescrittivo e sanzionatorio nei confronti di Ediscom S.p.A. - 23 febbraio 2023 [9870014]. https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9870014

[49] Lorenzo Porcelli, Massimo Ficco, and Francesco Palmieri. 2023. Mitigating User Exposure to Dark Patterns in Cookie Banners Through Automated Consent. In *International Conference on Computational Science and Its Applications*. Springer, 145–159.

[50] Ivana Rakovic and Yavuz Inal. 2023. Dark finance: exploring deceptive design in investment apps. In *IFIP Conference on Human-Computer Interaction*. Springer, 339–348.

[51] Ali Rasaii, Shivani Singh, Devashish Gosain, and Oliver Gasser. 2023. Exploring the Cookieverse: A Multi-Perspective Analysis of Web Cookies. In *International Conference on Passive and Active Network Measurement*. Springer, 623–651.

[52] Arianna Rossi, Rachele Carli, Maria Wilhelmina Botes, Angelica Fernandez, Anastasia Sergeeva, and Lorena Sanchez Chamorro. In press. Who is vulnerable to deceptive design patterns? A transdisciplinary perspective on the multi-dimensional nature of digital vulnerability. *Computer Law & Security Review* (In press).

[53] Cristiana Santos, Nataliia Bielova, and Célestin Matte. 2020. Are cookie banners indeed compliant with the law? *Technology and Regulation* 2020 (2020), 91–135.

[54] Cristiana Santos and Arianna Rossi. 2023. The emergence of dark patterns as a legal concept in case law. *Internet Policy Review* (July 2023). https://policyreview.info/articles/news/emergence-of-dark-patterns-as-a-legal-concept

[55] Cristiana Santos, Arianna Rossi, Lorena Sanchez Chamorro, Kerstin Bongard-Blanchy, and Ruba Abu-Salma. 2021. Cookie Banners, What's the Purpose?: Analyzing Cookie Banner Text Through a Legal Lens. In *Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society*. ACM, Virtual Event Republic of Korea, 187–194. https://doi.org/10.1145/3463676.3485611

[56] Than Htut Soe, Oda Elise Nordberg, Frode Guribye, and Marija Slavkovik. 2020. Circumvention by design-dark patterns in cookie consent for online news outlets. In *Proceedings of the 11th nordic conference on human-computer interaction: Shaping experiences, shaping society*. 1–12.

[57] Than Htut Soe, Cristiana Teixeira Santos, and Marija Slavkovik. 2022. Automated detection of dark patterns in cookie banners: how to do it poorly and why it is hard to do it any other way. *arXiv preprint arXiv:2204.11836* (2022).

[58] Michael Toth, Nataliia Bielova, and Vincent Roca. 2022. On dark patterns and manipulation of website publishers by CMPs. In *PETS 2022-22nd Privacy Enhancing Technologies Symposium*.

[59] Hellen Truong and Axel Dalbard. 2022. *Bright Patterns as an Ethical Approach to Counteract Dark Patterns. A Closer Investigation of The Ethics of Persuasive Design.* Ph. D. Dissertation. Jonkoping University, Jonkoping. https://www.diva-portal.org/smash/get/diva2:1680425/FULLTEXT01.pdf

[60] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019. (Un)informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19)*. Association for Computing Machinery, London, United Kingdom, 973–990. https://doi.org/10.1145/3319535.3354212

[61] Philippe Valoggia, Anastasia Sergeeva, Arianna Rossi, and Wilhelmina Maria Botes. 2024. Learning from the Dark Side About How (not) to Engineer Privacy: Analysis of Dark Patterns Taxonomies from an ISO 29100 Perspective. In *Proceedings of the 10th International Conference on Information Systems Security and Privacy*. SCITEPRESS-Science and Technology Publications.

[62] Ryan Matthew Wood. 2023. *Understanding the Impact of Dark Pattern Detection on Online Users.* Ph. D. Dissertation. Virginia Tech.

[63] Yuki Yada, Jiaying Feng, Tsuneo Matsumoto, Nao Fukushima, Fuyuko Kido, and Hayato Yamana. 2022. Dark patterns in e-commerce: a dataset and its baseline evaluations. In *2022 IEEE International Conference on Big Data (Big Data)*. 3015–3022. https://doi.org/10.1109/BigData55660.2022.10020800

[64] José P Zagal, Staffan Björk, and Chris Lewis. 2013. Dark patterns in the design of games. In *Foundations of Digital Games 2013*.

# A APPENDIX

## A.1 Deceptive Design Pattern Types and Their Definitions

**Table 3.** Deceptive Design Pattern Types After Deriving From [18] and Applying Granular Specifications with Their Descriptions. Cookie consent process-based filtered types are bold.

| Deceptive Design Pattern Type | Description |
|---|---|
| Immortal Accounts | Create a Roach Motel and use Obstruction to make it difficult or impossible to delete a user account once it has been created. |
| **Dead End** | reate a Roach Motel and use Obstruction to prevent users from finding information through inactive links or redirections that limit or completely prevent the display of relevant information. |
| Intermediate Currency | Creates Barriers and use Obstruction to hide the true cost of a product or service by requiring the user to spend real money to purchase a virtual currency that is then used to purchase a product or service. |
| **Privacy Maze** | Adds Steps and use Obstruction to require a user to navigate through many pages to obtain relevant information or control without a comprehensive and exhaustive overview. |
| Price Comparison Prevention | Creates Barriers and uses Obstruction by excluding relevant information, limiting the ability of a user to copy/paste, or otherwise inhibiting a user from comparing prices across two or more vendors. |
| Disguised Ad | Bait and Switch and use Sneaking to style interface elements so they are not clearly marked as an advertisement or other biased source. |
| Sneak into Basket | Hides Information and uses Sneaking to add unwanted items to a user's shopping cart without their consent. |
| Hidden Costs | Hides Information and uses Sneaking to reveal new charges or costs, present only partial price components, or otherwise delay revealing the full price of a product or service through late or incomplete disclosure. |
| Reference Pricing | Hides Information and uses Sneaking to include a misleading or inaccurate price for a product or service that makes a discounted price appear more attractive. |
| **Conflicting Information** | Uses (De)contextualizing Cues and Sneaking to include two or more sources of information that conflict with each other. |
| **Information without Context** | Uses (De)contextualizing Cues and Sneaking to alter the relevant information or user controls to limit discoverability. |
| **False Hierarchy** | Manipulates the Visual Choice Architecture, using Interface Interference to give one or more options visual or interactive prominence over others, particularly where items should be in parallel rather than hierarchical. |
| **Visual Prominence** | Manipulates the Visual Choice Architecture, using Interface Interference to place an element relevant to user goals in visual competition with a more distracting and prominent element. |
| Bundling | Manipulates the Visual Choice Architecture, using Interface Interference to group two or more products or services in a single package at a special price. |
| Pressured Selling | Manipulates the Visual Choice Architecture, using Interface Interference to preselect or use visual prominence to focus user attention on more expensive product options. |
| **Bad Defaults** | Subverts the user's expectation that default settings will be in their best interest, instead requiring users to take active steps to change settings that may cause harm or unintentional disclosure of information. |
| Cuteness | Uses Emotional or Sensory Manipulation and Interface Interference to embed attractive cues in the design of a robot interface or form factor. |

**Table 3.** (continued) Deceptive Design Pattern Types After Deriving From [18] and Applying Granular Specifications with Their Descriptions. Cookie consent process-based filtered types are bold.

| Dark Pattern Type | Description |
|---|---|
| **Positive or Negative Framing** | Uses Emotional or Sensory Manipulation and Interface Interference to visually obscure, distract, or persuade a user from important information they need to achieve their goal. |
| **Trick Questions** | Subvert the user's expectation that prompts will be written in a straightforward and intelligible manner, instead using confusing wording, double negatives, or otherwise leading language or interface cues to manipulate a user's choice. |
| **Choice Overload** | Subverts the user's expectation that the choices they make should be understandable and comparable, instead providing too many options to compare or encouraging users to overlook relevant information due to the volume of choices provided. |
| **Hidden Information** | Uses Emotional or Sensory Manipulation and Interface Interference to embed attractive cues in the design of a robot interface or form factor. |
| **Wrong Language** | Leverages Language Accessibility, using Interface Interference to provide important information in a different language than the official language of the country where users live. |
| **Complex Language** | Leverages Language Accessibility, using Interface Interference to make information difficult to understand by using obscure word choices and/or sentence structure. |
| **Feedforward Ambiguity** | Subverts the user's expectation that their choice will be likely to result in an action they can predict, instead providing a discrepancy between information and actions available to users that results in an outcome that is different from what the user expects. |
| **Nagging** | Subverts the user's expectation that they have rational control over the interaction they make with a system, instead distracting the user from a desired task the user is focusing on to induce an action or make a decision the user does not want to make by repeatedly interrupting the user during normal interaction. |
| Forced Continuity | Subverts the user's expectation that a subscription created in the past will not auto-renew or otherwise continue in the future, instead causing undesired charges, difficulty to cancel, or lack of awareness that a subscription is still active. |
| Forced Registration | Subverts the user's expectation that they can complete an action without registering or creating an account, instead tricking them into thinking that registration is required, often resulting in the sharing of unneeded personal data. |
| **Privacy Zuckering** | Uses Forced Communication or Disclosure as a type of Forced Action to trick users into sharing more information about themselves than they intend to or would agree to if fully informed. |
| Friend Spam | Uses Forced Communication or Disclosure as a type of Forced Action to collect information about other users through extractive means that results in unwanted contact from the service. |
| Address Book Leeching | Uses Forced Communication or Disclosure as a type of Forced Action to collect information about other users through extractive means, which are often hidden to the user and/or conducted under false pretenses. |
| Social Pyramid | Uses Forced Communication or Disclosure as a type of Forced Action to manipulate existing users into recruiting new users to use a service, often by tying this recruitment to additional functionality or other benefits. |

**Table 3.** (continued) Deceptive Design Pattern Types After Deriving From [18] and Applying Granular Specifications with Their Descriptions. Cookie consent process-based filtered types are bold.

| Dark Pattern Type | Description |
|---|---|
| Pay to play | Uses Gamification as a type of Forced Action to initially claim that aspects of a service or product are available via purchase or download, but then later charging users to actually obtain that functionality. |
| Grinding | Uses Gamification as a type of Forced Action to require repeated, often cumbersome and labor- intensive actions over time in order to obtain certain relevant functionality. |
| Auto-Play | Uses Attention Capture as a type of Forced Action to automatically play new video after an existing video has completed. |
| High Demand | Uses Scarcity and Popularity Claims as a type of Social Engineering to indicate that a product is in high-demand or likely to sell out soon, even though that claim is misleading or false. |
| Low Stock | Uses Social Proof as a type of Social Engineering to indicate that a product is limited in quantity,even though that claim is misleading or false. |
| Endorsements and Testimonials | Use Social Proof as a type of Social Engineering to indicate that a product or service has been endorsed by another consumer, even though the source of that endorsement or testimonial is biased, misleading, incomplete, or false. |
| Parasocial Pressure | Uses Social Proof as a type of Social Engineering to indicate that a product or service has been endorsed by a celebrity, influencer, or other entity that the user trusts, even though the source of that endorsement is biased, misleading, incomplete, or false. |
| Activity Messages | Use Urgency as a type of Social Engineering to describe other user activity on the site or service, even though the data presented about other users' purchases, views, visits, or contributions are misleading or false. |
| Countdown Timer | Uses Urgency as a type of Social Engineering to indicate that a deal or discount will expire by displaying a countdown clock or timer, even though the clock or timer is completely fake, disappears, or resets automatically. |
| Limited Time Message | Uses Urgency as a type of Social Engineering to indicate that a deal or discount will expire soon or be available only for a limited time, but without specifying a specific deadline. |
| **Confirmshaming** | Uses Personalization as a type of Social Engineering to frame a choice to opt-in or opt-out of a decision through emotional language or imagery that relies upon shame or guilt. |
| Personalization | Subverts the user's expectation that products or service features are offered to all users in similar ways, instead using personal data to shape elements of the user experience that manipulate the user's goals while hiding other alternatives. |

## A.2 Statistics of Association of Dark Pattern Attributes with Dark Pattern Types Process

| | | | | | Round 1 | | Round 2 | |
|---|---|---|---|---|---|---|---|---|
| | Expert 1 | Expert 2 | Expert 3 | Result | Count | Overall | Count | Overall |
| Consensus | Required | Required | Required | Required | 14 | 50 | 21 | 82 |
| | Optional | Optional | Optional | Optional | 6 | | 21 | |
| | Not required | Not required | Not required | Not required | 30 | | 40 | |
| Majortiy voting without conflict | Required | Required | Optional | Required | 11 | 41 | 11 | 30 |
| | Required | Optional | Required | | | | | |
| | Optional | Required | Required | | | | | |
| | Optional | Optional | Required | Optional | 6 | | 8 | |
| | Optional | Required | Optional | | | | | |
| | Required | Optional | Optional | | | | | |
| | Optional | Optional | Not required | Optional | 9 | | 7 | |
| | Optional | Not required | Optional | | | | | |
| | Not required | Optional | Optional | | | | | |
| | Not required | Not required | Optional | Not required | 15 | | 4 | |
| | Not required | Optional | Not required | | | | | |
| | Optional | Not required | Not required | | | | | |
| Majority voting with conflict | Required | Required | Not required | Required | 4 | 8 | 0 | 0 |
| | Required | Not required | Required | | | | | |
| | Not required | Required | Required | | | | | |
| | Not required | Not required | Required | Not required | 4 | | 0 | |
| | Not required | Required | Not required | | | | | |
| | Required | Not required | Not required | | | | | |
| High uncertainty | Required | Optional | Not required | - | 13 | 13 | 0 | 0 |
| | Required | Not required | Optional | - | | | | |
| | Optional | Required | Not required | - | | | | |
| | Optional | Not required | Required | - | | | | |
| | Not required | Required | Optional | - | | | | |
| | Not required | Optional | Required | - | | | | |
| | | | | Total | 112 | | | |

Fig. 4. Statistics of Step 2: the table compares the results from the first round and the second round of expert mapping between attributes and dark pattern types, in terms of required, optional and not required attributes for each type. During the first round, the three experts evaluated the attributes independently, while in the second round disagreements were solved collectively to avoid major conflicts across the voting.

## A.3 Description of Measurable Features and Their Design Categories

Table 4. Description of measurable features and their design categories

| ID | Category | Entity | Metric | Feature | Description |
|---|---|---|---|---|---|
| 1 | visual | FCG button | size | size of the FCG button | Size of a button such as "accept all", e.g. in pixels. |
| 2 | | FCR button | | size of the FCR button | Size of a button such as "reject all", e.g. in pixels. |
| 3 | | CC banner | | size of the CC banner | Size of the cookie consent banner, e.g. in pixels. |
| 4 | | FCG button | background color | background color of the FCG button | Color of a button such as "accept all", e.g. in RGB codes. |
| 5 | | FCR button | | background color of the FCR button | Color of a button such as "reject all", e.g. in RGB codes. |
| 6 | | consent setting button | | background color of the consent setting button | Color of a button such as "preferences", "options" etc., e.g. in RGB codes. |
| 7 | | FCG icon | availability | availability of the FCG icon | Binary feature. True if an icon is present on the FCG button. Otherwise, False. |
| 8 | | CC image | | availability of the CC image | Binary feature. True if an image is present on CC banner. Otherwise, False. |
| 9 | | CC image | relevance | relevance of the CC image | Semantic relevance of an CC image with the use case. |
| 10 | linguistic | CC information text | readability | readability of the CC information text | Readability score of the CC info text, e.g. Flesch Reading Ease score. |
| 11 | | CC information text | comprehensibility | comprehensibility of the CC information text | Assessment of how easily text is understood by user, e.g. percentage of correctly answered questions related to the CC information text. |
| 12 | | FCG text | clarity | clarity of the FCG text | Clarity of a text such as "accept all", e.g. based on a survey-based scoring from 1 to 5. |
| 13 | | FCR text | | clarity of the FCR text | Clarity of a text such as "reject all", e.g. based on a survey-based scoring from 1 to 5. |
| 14 | | FCG text | sentiment | sentiment of the FCG text | Sentiment of a text such as "allow cookies", e.g. positive. |
| 15 | | FCR text | | sentiment of the FCR text | Sentiment of a text such as "disagree", "reject", e.g. negative. |
| 16 | | CC information text | consistency | consistency of the CC information text | Assessment of how consistent terminology is used through the CC text. |
| 17 | | CC information text | language | language of the CC information text | Language of the CC information text, e.g. English, Italian, Turkish etc. |
| 18 | process | FCG process | availability | availability of the FCG process | Binary feature. True, if user has a process such as FCG process. Otherwise False. |
| 19 | | FCR process | | availability of the FCR process | Binary feature. True, if use has a process such as FCR process. Otherwise False. |
| 20 | | FCG process | number of clicks | minimum number of clicks for the FCG process | Shortest user path to perform FCG process in number of clicks. |
| 21 | | FCR process | | minimum number of clicks for the FCR process | Shortest user path to perform FCR process in number of clicks. |
| 22 | | FCG process | number of paths | number of paths for the FCG process | Total number of paths are available to the users for the FCG process. |
| 23 | | FCR process | | number of paths for the FCR process | Total number of paths are available to the users for the FCR process. |
| 24 | | CC setting process | time | time for CC setting process | Duration of the CC setting configuration in seconds. |
| 25 | | FCG process | | time for the FCG process | Duration of the FCG process in seconds. |
| 26 | | FCR process | | time for the FCR process | Duration of the FCR process in seconds. |
| 27 | | first interaction process | | time for the first interaction process | Duration between the moment CC is presented to the user and user's first action in seconds. |
| 28 | | CRR process | frequency | frequency of the CRR process | Frequency of the CRR after consent decision is given by user. |