Investigating Deceptive Design in GDPR's Legitimate Interest

Lin Kyi
Max Planck Institute for Security and Privacy
Bochum, Germany
lin.kyi@mpi-sp.org

Sushil Ammanaghatta Shivakumar
Max Planck Institute for Security and Privacy
Bochum, Germany
sushil.shivakumar@mpi-sp.org

Franziska Roesner
University of Washington
Seattle, United States
franzi@cs.washington.edu

Cristiana Santos
Utrecht University
Utrecht, The Netherlands
c.teixeirasantos@uu.nl

Frederike Zufall*
Max Planck Institute for Research on Collective Goods
Bonn, Germany
zufall@coll.mpg.de

Asia J. Biega
Max Planck Institute for Security and Privacy
Bochum, Germany
asia.biega@mpi-sp.org

ABSTRACT

Legitimate interest is one of the six grounds for processing data under the European Union's General Data Protection Regulation (GDPR). The flexibility and ambiguity of the term "legitimate interests" can be problematic; coupled with the lack of enforcement from legal authorities and different interpretations from the various data protection authorities, legitimate interests can be taken advantage of as a loophole to collect more user data.

Drawing insights from multiple disciplines, we ran two studies to empirically investigate the deceptive designs being used when legitimate interests are applied in privacy notices, and how user perceptions line up with these practices. We identified six deceptive designs, and found that the ways legitimate interest is applied in practice does not match user expectations.

CCS CONCEPTS

• **Security and privacy** → **Human and societal aspects of security and privacy**; • **Human-centered computing** → **User studies**; • **Applied computing** → *Law*;

KEYWORDS

Deceptive Design, Dark Patterns, GDPR, Consent, Privacy Notice, Legitimate Interest, Human-Computer Interaction

ACM Reference Format:
Lin Kyi, Sushil Ammanaghatta Shivakumar, Franziska Roesner, Cristiana Santos, Frederike Zufall, and Asia J. Biega. 2023. Investigating Deceptive Design in GDPR's Legitimate Interest. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3544548.3580637

*Co-affiliated with the Waseda Institute of Advanced Study at Waseda University, Tokyo.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3580637

1 INTRODUCTION

The European Union's General Data Protection Regulation (GDPR) outlines legal grounds for the processing of personal data to be lawful (Article 6), including the legal basis of consent and legitimate interest, amongst other grounds. *Legitimate interest* is defined in Article 6(1)(f) of the GDPR as the processing that is *"necessary for the purposes of the legitimate interests pursued by the controller or by a third party* [25]. Pursuant to this paper, *data controller* refers to a website or company that is processing personal data, such as service providers, advertisers, and consent management platforms (CMPs) [66].

While the practical implementations of the legal basis of consent have already been studied in a variety of contexts [41, 54, 63, 66], the usage of legitimate interests as a legal basis for data processing by websites remains relatively unexplored. Yet, as highlighted by legal scholars, out of all legal grounds under the GDPR, legitimate interest is the most ambiguous one because it allows for broad interpretations of different processing purposes [28, 47]. In fact, the legal uncertainty of which data processing purposes should fall under the legal basis of legitimate interest is currently the subject of great attention at the EU level by the European Commission [22], Data Protection Authorities (DPAs) [16, 25], the EU Court of Justice (ECJ) [15], as well as other national courts. Moreover, the lack of legal enforcement from regulators and courts may allow for this legal basis to be exploited by data controllers as a loophole for dubious data practices [28, 47]. Hence, there is a need for further investigation into the use and applicability of legitimate interests in practice.

Even though legitimate interests allow data controllers to process data without explicit permission from the user, they have the right to *object* to legitimate interests (Art. 21(1) GDPR). As a result, legitimate interests appear in website privacy notices [1] more and more commonly (see Figure 1 for an example notice design). The potential for legitimate interests to be exploited to collect more user data raises the question of whether its practical implementations use any deceptive designs. In the context of consent, it has been shown that

[1]The notices we focus on this paper are also often called "consent notices", "cookie banners", or "cookie pop-ups". We use the term "privacy notices" (not to be confused with "privacy policies") here, since we refer to notices that do not just ask for consent but also inform users of data processing based on legitimate interests.

data controllers—including advertisers, service providers, websites, and consent management platforms (CMPs)—use deceptive designs in their privacy notices [41, 54, 63, 67].

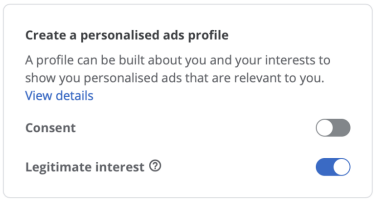

Figure 1: An example of a consent-legitimate interest toggle. This is a legally dubious design element because data collection purposes can only rely on either consent or legitimate interest, not both [43]. Additionally, it may be unclear if data is being collected even if "Consent" is switched off.

Deceptive designs, or dark patterns, are defined as user interfaces which lead users into making decisions that benefit the online service [40, 55, 62]. [2] In the context of privacy and data protection, we are interested in deceptive designs that deceive users into making *poor privacy decisions*, such as making it difficult to object to data collection or using obscure or technical language [5, 41, 63]. Exactly how often legitimate interests appear in privacy notices, what deceptive designs arise in this context, and whether the amalgamation of consent and legitimate interests settings into one notice makes decision making harder for users, remain open questions, which we investigate in this paper.

Beyond implementation practices and deceptive designs, it is also important to understand user perceptions of legitimate interests since they, as data subjects, are directly affected by an interference with their right to the protection of personal data from Art. 8 of the EU Charter of Fundamental Rights of the European Union [27].

Despite the many issues that legal scholars have noticed regarding the use of legitimate interests [28, 47], there is a lack of research investigating how this legal basis is applied in practice by data controllers and perceived by users. To the best of our knowledge, this paper is one of the first to provide the relevant empirical evidence. Drawing insights from multiple disciplines, we conduct a two-part investigation into legitimate interests. First, we examined the practices and deceptive designs in the usage of legitimate interests by data controllers, as evidenced by website privacy notices. Second, we study user understanding and perceptions of legitimate interests.

The contributions of this paper are:

(1) We examined how the legal basis of "legitimate interests" is being used in practice by data controllers (publishers and CMPs) in privacy notices on their websites and discuss the legal implications of these practices;
(2) We identified deceptive designs used by data controllers when it comes to implementing the legitimate interest legal basis for processing (Art. 6(1)(f) GDPR); and
(3) Building on our findings from the first study, we investigated how end-users perceived legitimate interest data collection purposes to provide a user context to these practices.

Study one identified deceptive designs used for the legal ground of legitimate interest. Out of the 10,000 sites we crawled, 474 (4.74%) included "legitimate interest(s)" in their privacy notices, and we found that various UI and linguistic deceptive designs were present in most of these notices. Legitimate interests, when disclosed, are often difficult to object to, are designed in ways that might confuse users, therefore leading to lower user engagement rates. Moreover, our results demonstrate that IAB Europe's Transparency and Consent Framework (TCF) [21] has a major role in how legitimate interest is applied in practice.

Based on the practices identified in the first study, *study two* surveyed 400 users to understand how users perceived these practices. We found that users were wary of the data collection practices that are commonly used in practice, and that a data collection purpose's *user acceptance*, meaning whether users find a purpose essential or are comfortable with it, is most impacted by who it is believed to benefit: users are less likely to feel comfortable sharing their data if they believe a purpose benefits data controllers more than other stakeholders. Overall, we found that the ways legitimate interests are used in practice are not in line with user beliefs about how their data should be used, indicating that user preferences should be taken into account when creating and revising data protection laws and defining industry standards.

2 BACKGROUND

2.1 Legal Background

Legal requirements for processing personal data.

Legal basis. As the processing of personal data constitutes an interference with the right to protection of personal data in Art. 8 EU Charter [27], it requires a justification based on Art. 8(2) of the EU Charter. This justification may be provided by the consent of the person concerned, or through "some other legitimate basis laid down by law". This requirement is substantiated by Art. 6(1) (a-f) of the GDPR that lists potential legal grounds for the processing of personal data to be legitimate. Personal data shall only be processed if at least one of six legal grounds listed in that Article apply: (a) consent the data subject, (b) performance of a contract with the data subject, (c) compliance with a legal obligation imposed on the controller, (d) protection of the vital interests of the data subject, (e) performance of a task carried out in the public interest, or (f) legitimate interests pursued by the controller, subject to an additional balancing test against the data subject's rights and interests.

Applicability of the GDPR and ePrivacy Directive. While the GDPR applies to the processing of personal data, the ePrivacy Directive provides rules on the confidentiality of data on the user's device. Whenever cookies and other tracking technologies are stored and read from the user's device, the ePrivacy Directive [19] requires controllers to request consent for the storage of such cookies in Art. 5(3) for certain purposes for processing data (such as advertising). Based on a recent ruling by the European Court of Justice (ECJ) [10], the GDPR and the ePrivacy Directive thus apply in parallel. A follow-up question is then whether in case of cookies, where Art. 5(3) ePrivacy Directive requires consent for the processing of personal

[2] We generally use the term "deceptive design" to refer to this kind of design pattern in this paper, but acknowledge that other terms, such as "misleading" or "manipulative", may be more precise in certain cases. The term "dark pattern" has been criticized https://medium.com/@carolinesinders/whats-in-a-name-unpacking-dark-patterns-versus-deceptive-design-e96068627ec4.

data, a justification based on legitimate interests (Art. 6(1)(f) GDPR) may still apply.

Legal requirements for the application of the legal basis of legitimate interest.

Definition of the legitimate interest legal basis. This legal basis (Art. 6(1)(f)) is elaborated by Article 29 Working Party (29WP) in its Opinion 217 on the notion of legitimate interests of the data controller [25]. [3] Accordingly, processing personal data will be lawful when it is necessary for the purposes of the legitimate interests of a controller or by a third party to whom data was disclosed to. The general provision on legitimate interest is open-ended, with a broad and unspecific scope [25], meaning that it can be relied upon a wide range of purposes and it is not purpose-specific, as long as its requirements are satisfied. This flexible definition also carries heightened obligations on controllers to balance its own interests with the rights and interests of users.

Role of legitimate interest. The 29WP recognizes the role and usefulness of this legal basis provided that its requirements are fulfilled [25]. Legitimate interest gives data controllers the ability to innovate and provide better services, while ideally keeping them accountable for their actions [30]. It aims at a balanced approach, which ensures the necessary flexibility for data controllers for situations where there is no undue impact on data subjects, while at the same time providing sufficient legal certainty and guarantees to data subjects that this open-ended provision will not be misused [25].

Requirements. The open-ended nature of this provision raises important questions regarding its exact scope and application [25]. It is mandatory that such processing is necessary for the purposes of a given interest. Given the required justification of an interference with the right to the protection of personal data by Art. 8 EU Charter [27], these legitimate interests may justify data collection if they override the data subject's interests and rights, such as the right to privacy [9, 14]. Accordingly, some obligations impend over controllers:

(1) they are required to perform a *balancing decision* in every single context as to whether this requirement is met;
(2) they are obliged to inform users about the specific legitimate interests pursued, and about the right to object to those (Arts. 13(1)(d), 14(1)(b), 21(1) GDPR);
(3) they must pursue legitimate interests that are "lawful","sufficiently clearly articulated" (i.e., transparent) and "represent a real and present interest" [25, p. 25,52]; and
(4) they must use such interests under processing purposes that are specific (sufficiently detailed) and explicit (unambiguous, not hidden), and understood in the same way by everyone involved.

Relationship between the legitimate interest legal basis and other legal bases. Legitimate interest can be seen as a complementary to other legal bases. It may help to prevent consent fatigue [44] and over relying on other legal grounds (e.g., consent or contract) [25]. Legitimate interest should not be used as a "catch-all", "open door" or "the weakest link" ground to fill in gaps for rare and unexpected situations where the other legal basis are not applicable [25]. Legitimate interest should neither be seen as a preferred legal basis, should not be automatically chosen, or its use unduly extended on the basis of a perception that it is less constraining than the other grounds legitimizing data processing [25].

Design requirements. Certain data protection rules apply to deceptive design. In particular, *default settings* must be designed with data protection in mind, and hence, data controllers must "implement appropriate technical and organizational measures which are designed to implement data protection principles" according to Art. 25 GDPR. They must ensure that, "by default, only personal data which are necessary for each specific purpose of the processing are processed". Promoting the idea of *"data protection by design"* through this provision, the GDPR does not get more specific in regulating deceptive designs.

Consent Management Platforms. Consent Management Platforms (CMPs) provide privacy notices that can be embedded in websites to enable streamlined compliance with the legal requirements for consent mandated by the ePrivacy Directive and the GDPR. Hils et al. found that several CMPs used legitimate interests as a way of collecting data [41]. Matte et al. found that 19% of advertisers collect data under legitimate interests grounds [56] where consent should be used instead. Santos et al. found that CMPs offer default consent banners featuring deceptive designs [66].

The GDPR and stakeholder guidelines provide a set of requirements for the formulation of legitimate interests. Due to the prevalence of legitimate interests being used to collect user data, it is important to identify the purposes that data controllers are using as legitimate interest purposes, and how users react to them.

2.2 Deceptive Design

Why deceptive design is commonly used. Deceptive design is ubiquitous in online privacy notices [63] because data controllers are incentivized to collect as much user data as possible for their own gains [31, 77]. As a result, users are faced with ambiguity about what is happening to their data [31, 51]. User data is a very valuable commodity in this current online economy [77]. Data can be used to benefit both the user and the data controllers by creating better personalized services, provide user and market insights, and gaining revenue from advertisers [8, 77].

Deceptive design vs. poor design vs. nudging. There is a fine line between poor design, nudging, and deceptive design [13, 62]. We cannot ascertain how intentional data controllers' deceptive design practices are, but research has shown that these design elements confuse and deceive users, as they can nudge most users into making poor decisions by taking advantage of user psychology [33, 34, 40, 46, 55, 62, 75] and can harm them [36]. Paternalism is the idea that UI designs should nudge, or influence, users into making decisions that are better for them [4, 69]. With deceptive design, we see the opposite; users are nudged into making decisions that are not in their best interests [31]. Regardless of the data controller's intentions, whether accidental or intentional, deceptive design is very effective because it can nudge most users into making poor decisions [33, 34, 40, 46, 55, 62, 75].

[3] The Article 29 Working Party (29WP) was the former European Data Protection Board, a European Union body whose purpose is to ensure consistent application of the GDPR and to promote cooperation among the EU's data protection authorities. While their opinions or guidelines are not formally binding, they hold much authority in member states and provide comprehensive guidelines for data controllers as to how they should apply the concept of personal data in practice.

Deceptive design in privacy notice consent settings. Only 11.8% of privacy notices fit legal standards [63, 67] since most are riddled with deceptive designs [35, 63]. It is common to see default selections that highlight the "Select all cookies" option, and making it harder to reject them, hiding privacy-preserving choices, and overloading users with too many options and layers in a privacy notice [5, 31, 37]. In mobile applications, the situation with deceptive designs is even more severe due to the lack of GDPR enforcement in this space [48]; only 9.9% of apps included some form of privacy notice [49].

Previous work demonstrated the impact of deceptive designs in privacy notices: users are more likely to make consent decisions that benefit the service more than themselves [41, 63, 73]. Santos et al. stated that most of the research on the legality of privacy notices tends to focus on the user interface level, and less on the text within privacy notices [67]. They found that 89% of privacy notices violated the law with just their wording and framing of purposes; 50% were too vague, 30% used positive framing (i.e., making it sound like data collection only benefits the user), and 42.41% misled users by using false statements, which violates GDPR standards [67].

The majority of research has focused on the informed consent aspects of privacy notices, but not the legitimate interest aspects of privacy notices. Legitimate interest does not require user consent to collect data, but is still included in many privacy notices. Therefore, we are interested in investigating how legitimate interest is being used in privacy notices to expand on this research.

Deceptive design regulation and enforcement. Most recently, several data protection authorities (DPAs) and court decisions were issued and forbade certain deceptive design practices: preselection of choices [10]; difficulty to refuse consent as easily as to accept (e.g., several clicks are necessary to refuse data collection consent purposes) [11, 12]; misinforming users on the purposes of processing data and how to reject them [12, 53].

The future Digital Single Act (DSA) [72] will finally ban deceptive designs explicitly. It prohibits "online platforms to design, organize or operate their interfaces in a way that deceives, manipulates or otherwise materially distorts or impairs users' free and informed decisions" (Article 23a(1)).

Policy-making efforts attempt to address deceptive practices [23], but regulation within EU data protection law lacks operational guidelines for implementation by practitioners. As a result, there is a lot of freedom for interpretation when it comes to deceptive design within privacy notices. This is especially problematic because enforcement is still too slow compared to the speed and ubiquity of deceptive practices.

2.3 User Perceptions of Personal Data Collection

The majority of research on user perceptions of data collection practices have been related to personalization of services and advertisements. We expand on this literature by investigating user perceptions of legitimate interest practices. Previous work has shown that users are generally aware that AI and personalization services are reliant on user data to function [50], and have a general understanding of online tracking [8]. As a result, many users have concerns about their online privacy, and feel they have little control over their data [3].

Kozyreva et al. found in their survey of users that there is an *acceptability gap* in the use of personal data for personalization. Users are more likely to use and accept personalized services, but are less accepting of giving up the personal data needed for these services [50], which highlights the tradeoffs users need to consider in exchange for convenience online. Chanchary and Chiasson's study showed that users generally were accepting of online behavioral advertising (OBA), but had preferences for the kind of information they were willing to share online [8]. Overall, users were more concerned about third-party tracking on online banking sites compared to other websites (e-commerce, search engine, and social networking sites). When it came to sharing personal information, users were more willing to share location information, and demographic and computer information compared to personal identification and financial information [8].

3 STUDY 1: PRIVACY NOTICE WEB CRAWL

The purpose of our first study was to understand how legitimate interest is being (mis)used in practice, and identify deceptive designs that are prevalent when applying the legal basis of legitimate interest. As explained in Section 1 and 2.1, users have the right to object to data processing based on legitimate interests. Hence, our focus in this study was on privacy notices because they appear to more frequently allow users to express their legitimate interest preferences. Research questions we investigated in this study were:

(1) What are the data collection purposes used under legitimate interest grounds?
(2) What kind of deceptive designs are used when it comes to legitimate interest in privacy notices?
(3) Are the practices surrounding legitimate interest used in privacy notices legal?

3.1 Web Crawl Methodology

We created a web crawler to analyze 10,000 websites' privacy notices from the Tranco top sites list [4] which provides rankings that are oriented for research, and aims to be reproducible [65]. We created a script using Selenium for Python [5] which went through the top websites from March 12, 2022. We ran the script in May 2022 from a French IP address. Due to the GDPR's requirement that websites need to inform their users when their data is being collected, we expected every website to contain a privacy notice for us to analyze. However, we knew from our own observations that not every website mentioned their legitimate interests in their privacy notices, therefore we used a web crawler to investigate the prevalence and uses of legitimate interests from a larger dataset.

The Selenium library uses a driver which initiates an automated test software that opens up Google Chrome in another tab (using Chrome driver). The driver then opened websites individually from Tranco's top websites list, collected that webpage's data source, and checked if "legitimate interest" or "legitimate interests" appeared. If "legitimate interest(s)" was not found in the first page of the privacy

[4] https://tranco-list.eu/
[5] https://selenium-python.readthedocs.io/

notice, our crawler went to the next layer of the privacy notice, and searched for it there.

As CMPs use different wording in their privacy notices, such as "Show purposes" or "Select choices", for buttons that users can click on to modify their preferences, we created an additional script that identified these phrases used in our list of 10,000 websites. Overall, we found 63 such terms used on English websites, including "More options," "Customize", and "Learn more." Using this list of phrases, our crawler clicked the buttons containing these phrases if "legitimate interest(s)" was not found in the first layer.

In the second layer of the privacy notice, the crawler repeated the same steps; it looked at the page source, and searched for "legitimate interest(s)". If these words were not found, the crawler would mark it as not being present in the privacy notice. If the words were present, the crawler made note of which page(s) of the privacy notice it was in, and the number of clicks needed to access legitimate interest(s) in the privacy notice.

The crawler flagged websites that contained the phrase "legitimate interest" or "legitimate interests" in the HTML of the web page (which is almost always in the privacy notice), and took screenshots of the privacy notices that mentioned these phrases for the qualitative stage of data analysis. We took screenshots of the whole privacy notice, including ones that were long and required scrolling relevant pages containing "legitimate interest(s)". Refer to Section 1 of our Supplementary Materials for examples of the privacy notices the web crawler flagged and screenshots it took. Additionally, we captured the text relevant to legitimate interest(s) using the Beautiful Soup Python library to supplement the screenshots [6]. This library scraped the text data by parsing over the HTML of the webpages.

We only went up to the second layer because the most relevant information for users tends to be on the first two layers; investigating deeper than these two layers would require very complex actions from our web crawler, most likely customized for individual privacy notices. Additionally, most users only tend to look at the first layer of a privacy notice [41].

Dataset. Our web crawler analyzed the top 10,000 websites from the Tranco list, and flagged 643 websites as containing "legitimate interest(s)" in the consent banner. Out of these sites, 474 were valid for our study. We excluded cases where i) the privacy notice was in a language other than English, ii) the website was changed or removed when we conducted our analyses, and iii) the website only contained the phrase "legitimate interest" in the HTML but not the privacy notice.

3.2 Data Analysis

Quantitative analysis: Using the data from the web crawl, we conducted statistical tests to understand the relationship between implementation practices around the use of legitimate interest in privacy notices.

Qualitative analysis: We conducted a content analysis of our privacy notice screenshots taken by the web crawler. We analyzed the page(s) of the privacy notice "legitimate interest(s)" was found in, and focused on details relevant to legitimate interest in the annotating process. The annotating was performed by two annotators, including the first author of this paper and a co-author. The annotators initially used an inductive coding approach, which turned into a deductive approach once they knew what to look for. The annotators first met and coded 15 randomly selected privacy notices together to determine what is important for the analysis. They decided that in addition to focusing on UI design and linguistic elements relevant to legitimate interest, it would be important to keep track of:

- Whether the privacy notice mentions that legitimate interest can be objected to;
- Whether users can object to legitimate interest in the privacy notice (either by interacting with the privacy notice buttons or being re-directed to another page/website to object);
- Whether the privacy notice listed the purposes for which the data controllers were using legitimate interest for; and
- Whether there are elements showing a data controller is relying on both consent and legitimate interest as legal grounds (such as toggles, see Figure 1),

After coding the initial 15 notices together, the annotators coded the same set of 40 randomly selected notices independently. They had a few more meetings and maintained online communications to adjust and calibrate their codes. For the 55 notices which they coded together, the agreement rate was 84.8%, with a Cohen's kappa of $\kappa = 0.70$, which meant there was substantial agreement [60]. Since the inter-rater reliability rate was high, and privacy notices are often very similar to each other due to the standardized use of CMPs [41], the annotators halved the remaining set of notices and coded them independently. According to McDonald et al., it is acceptable to divide the qualitative analysis if there is a large dataset [59], such as ours with over 400 privacy notices.

3.3 Results

In our qualitative analysis of the 474 privacy notices that contained "Legitimate Interest(s)", we identified 87 codes, and 8 themes. See Section 3 of the Supplementary Materials for our codebook. We will be using terminology from recent qualitative CHI papers to discuss frequencies for our qualitative data in this section (see Figure 2) [18, 38].

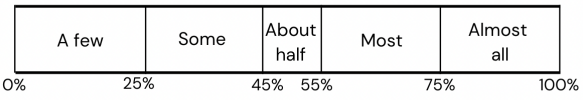

Figure 2: The terminology used to represent the frequency of themes.

3.3.1 Prevalence and Usage of Legitimate Interests. We found that 474 websites (4.74%) of the 10,000 sites we crawled included their legitimate interests in their privacy notices. Out of these websites, 273 (57.59%) websites mentioned their legitimate interests on the first page of the privacy notice, and 201 (42.41%) mentioned it after the user performs one click (e.g., the "next" or "show purposes" button).

[6] https://beautiful-soup-4.readthedocs.io/en/latest/

Common descriptions of legitimate interest used by most websites were "legitimate business interest" and "processing without user consent"

In our analysis, every website that mentioned "legitimate interest" was using IAB Europe's TCF [21], and therefore reusing this framework's legitimate interest purposes in their privacy notices. This signifies the central role that the TCF has in the consent landscape [41]. These purposes are described in Section 2 of our Supplementary Materials.

When privacy notices described whose legitimate interests these purposes were meant for, almost all mentioned these purposes were for third party vendors, such as advertising partners, TCF and IAB vendors, social media partners, and analytics partners. Only a few mentioned that the service provider themselves were collecting data for their own legitimate interest purposes.

3.3.2 Deceptive Design Elements. Our analysis found several deceptive designs used when legitimate interest is included in privacy notices. We organized the legitimate interest-related deceptive design elements into three categories: i) general UI elements, ii) complex choice architectures, and iii) linguistic elements. Table 1 summarizes the UI, complex choice architecture, and linguistic deceptive elements we identified. We fully describe the deceptive designs we identified in Section 6.1.

3.3.3 General UI deceptive design elements. We noticed that i) many privacy notices lacked an option to object to all legitimate interest-based purposes, and instead prompted users to individually deselect them (often at purpose and/or third party vendor level, see Figure 3 for an example); ii) accessing legitimate interest purposes often requires users to click through several pages, and iii) legitimate interests may not be mentioned on the first page of privacy notices (seen in 42.41% of our privacy notices).

3.3.4 Complex choice architectures. Complex choice architectures refer to UI deceptive design elements that complicate the user's interactions with a website's UI in the context of decision making. Such elements were observed in our analysis (see Figure 1) and included: i) toggles that rely on both consent and legitimate interest as legal bases for collecting data for the same purposes, a practice which has been declared illegal [43]; ii) making it difficult to object to legitimate interests by directing users to the privacy policy or the third party vendor's website to object to processing; and iii) requiring users to expand each legitimate interest purpose in the privacy notice to manually opt out of each preselected purpose.

3.3.5 Linguistic deceptive design. Given that users may have varying commonsense interpretations of the term "legitimate interest", we considered a privacy notice to use linguistic deceptive design when it had poor, or missing explanations of what legitimate interest meant. Our analysis yielded several observations. Firstly, we found placebic explanations [17] of legitimate interest, with websites providing mostly tautological definitions of what the concept means, e.g., *"How does legitimate interest work? Some vendors are not asking for your consent, but are using your personal data on the basis of their legitimate interest"*. Secondly, some websites mentioned "legitimate interest" without providing any definitions. Thirdly, most privacy notices did not mention whose legitimate interests the data processing would benefit. Lastly, only 86.3% of privacy notices listed purposes for which data was collected on the basis of legitimate interest.

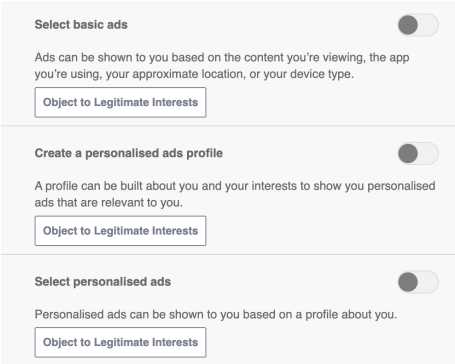

Figure 3: Users are often required to individually object to legitimate interest at the purpose and/or vendor level.

3.3.6 Positive Framing. Most privacy notices used positive framing when describing why users should accept a data collection purpose, such as mentioning that user data provides users with free services, that more data provides better results, and that accepting cookies ensures proper site functioning. This result aligns with prior findings on the framing of privacy notice purposes [67]. Highlighting the positive aspects of processing makes users pay less attention to other (negative) aspects (e.g., targeted advertising) [6] which are important for a meaningful and informed decision.

3.3.7 Advertising. Using legitimate interest for advertising purposes plausibly violates the lawful principle (Article 5(1)(a)GDPR), hence rendering the practice unlawful [2, 43, 57]. Yet, many websites disclosed that they were processing data for advertising purposes on the basis of their legitimate interests. We found that in those cases the privacy notices implemented the TCF, which lists several advertising purposes as legitimate interests.

4 STUDY 2: USER REACTIONS TO LEGITIMATE INTEREST AND CONSENT PURPOSES

Based on the results of our web crawl study, we conducted a survey to better understand how users react to various data collection purposes. The full survey is available in Section 4 of our Supplementary Materials.

The research questions for this study were:

(1) Does the website category impact user opinions of which data collection purposes are acceptable to be used?
(2) What is people's understanding of the term "legitimate interest"?
(3) What do people expect or think is reasonable for companies to collect under legitimate interest?
(4) What do people think the harms associated with data collection under legitimate interest purposes are?

Table 1: The number of sites with different deceptive designs in the usage of Legitimate Interest (LI).

	Privacy notice mentions that users can object to LI	Users can object to LI in privacy notice	Complex choice architecture: LI and consent toggle present	Privacy notice lists LI purposes
Yes	473 (99.8%)	464 (97.9%)	147 (31%)	409 (86.3%)
No	1 (0.2%)	10 (2.1%)	327 (69%)	65 (13.7%)

4.1 Methodology

4.1.1 Survey design. We distributed our survey using Qualtrics. The median time needed to complete the survey was around 10 minutes. After briefly explaining the purpose of our survey, we asked general background questions such as whether participants have ever come across a privacy notice (Q1), whether they have ever adjusted their privacy preferences (Q2), how concerned they are about their data privacy (Q3), and how much control they feel over their online privacy (Q4).

A central part of our survey was a vignette section where participants imagined themselves using one of eight randomly assigned website types (news, e-commerce, search engine, social media, government, non-profit, entertainment, or a banking site). We used a between-subjects design, where everyone evaluated only one type of website category, and we had an equal number of participants evaluating each website category. Each vignette (i.e., website category) included a screenshot to mentally situate the participant in the environment of the website. These screenshots were non-branded, meaning that they did not represent real websites, so that participants would not be primed by their opinions of particular services. The website categories were compiled based on previous research by Habib et al. [38], Chanchary and Chiasson [8], and Amos et al. [1]. Participants were presented with one randomly selected vignette, but all survey questions and the evaluated data collection purposes were the same, regardless of the website category a participant was assigned to.

To gauge at whether certain consent-based purposes (*Functional and strictly necessary, UX improvements*, and *Sharing with third parties*) could potentially be future legitimate interests, we asked participants how essential they deemed this purpose for the functioning and service offering of that particular kind of website (Q5, 7, 9). We used a 5-point likert-scale, ranging from "1 - Completely disagree" to "5 - Completely agree". We also asked participants to rate how much they think each purpose benefited different stakeholders (*the user, service provider, third party vendors, other users,* and *society*) (Q6, 8, 10) on a 4-point likert-scale, ranging from "1 - Not at all" to "4 - A lot".

From an operational point of view, the application of legitimate interest means that, for certain purposes, personal data is collected without the user's explicit permission. We investigated how our participants felt about this practice with respect to the purposes commonly used under legitimate interest (*Personalizing and measuring content, Personalizing and measuring ads, Analytics, Develop and improve products, Future innovations, Archiving, Security and debugging,* and *Fraud and law enforcement*), asking how comfortable they were with websites collecting data for these purposes without asking for user permission (Q11, 13, 15, 17, 19, 21, 23, 25). Additionally, we asked participants to rate on a 4-point likert-scale how much they think each purpose benefits different stakeholders (*the user, service provider, third party vendors, other users,* and *society*) (Q12, 14, 16, 18, 20, 22, 24, 26). In this part of the survey, we did not use the term "legitimate interest" because we anticipated that some participants might not understand it. Instead, we paraphrased it using an operational framing: "purposes that use data without your permission".

The next section of our survey investigated participants' general understanding of the concept of legitimate interest and its practical implications. Questions included: asking participants whose legitimate interests they thought websites were referring to when collecting data for a "legitimate interest" purpose (Q27), presenting screenshots of the four possible consent/legitimate interest toggle configurations and asking participants in which configuration(s) they thought data was being collected (Q28), as well as asking an open-ended question about the perceived harms of using legitimate interest for data collection (Q29).

The last section of the survey consisted of demographic questions. We asked about participants' technical and privacy knowledge using the web skills use survey (Q30-38) [39], their age (Q39), gender (Q40), how long they had been living in the EU (Q41), where they lived in the EU (Q42), and the language they primarily used the Internet in (Q43).

4.1.2 Survey Validation. We initially piloted the survey with three participants; one participant had a security and privacy background, and two had non-computational backgrounds. Based on the findings from our pilot, we revised the wording and presentation of the survey, and released a pre-test with 30 participants, which is a recommended sample size for survey validation [64]. We distributed the pre-test on Prolific [7], using the same pre-screening criteria as in the final survey. We did not use the pre-test data for our final analysis. The pre-test with 43 survey items had a Cronbach's alpha of $\alpha = 0.92$, which indicates that our survey had good internal reliability [70]. We therefore did not need to further change our questions.

To maximize the survey validity, we reused or adapted questions from previous studies where applicable. Some questions and multiple choice responses were adopted from previous work by Kozyreva et al. [50] (for the question, *How concerned are you about your data privacy when using the Internet?*). Habib et al.'s codebook was used to inform some multiple choice responses [38]. In the demographics section, we used the web skills use survey by Hargittai and Hseih [39]. We were also careful to ensure that there was a correspondence between the survey questions and our research questions. For example, for *RQ3: What do people expect or think is reasonable for companies to collect under legitimate interest?* we

[7] https://www.prolific.co/

asked, *Sometimes, websites might collect data for the following purposes without asking for permission. For [insert legitimate interest purpose], how comfortable are you with this?*

4.1.3 Taxonomy of purposes. There is no centralized nor standard list of legitimate interest purposes, as this legal basis is determined on a case-by-case basis with the legitimate interests balancing test [32, 44]. Accordingly, pursuant to our survey, we used the purposes from the Cookiepedia Database [8], which provides an extensive database and categorization of cookies, and is often used in empirical studies on cookies [42, 66]. Additionally, we looked to guidelines from the European Data Protection Board (EDPB) [25], Center for Information Policy Leadership (CIPL) [7], and the TCF [21] to find which purposes can be based on legitimate interest.

After several discussions, we agreed to survey participants on 11 purposes for section two of our survey - eight of which were based on legitimate interest (therefore not requiring consent), and three of which are subject to consent, but were interested in seeing if users might deem them essential, therefore potential legitimate interests. We were broadly interested in the legitimate interest purposes we identified from the first study, but also purposes related to the repurposing of data, such as *Future innovations, Archiving,* and *Product development* because of their relevance to critiques of data minimization [29, 76] and their general importance to scientific, social, and product development. For the three consent-based purposes, we included them because they are commonly used purposes [41, 66], therefore we were interested to see if we could reduce user fatigue by including these as potential legitimate interests in the future.

For purposes that were very similar, such as TCF-based *Create a personalized ad profile, Select personalized ads, Create a personalized content profile,* and *Select personalized content* we amalgamated them into general purposes called *Personalized content delivery and measurement* and *Personalized ad delivery and measurement* to reduce repetition for participants. Below, we list the purposes and definitions we presented to participants. The table in Section 6 of our Supplementary Materials describes the reasoning behind each purpose and where we sourced it from.

- *Functional, strictly necessary purposes:* enables you to move around the website and use its features
- *User experience (UX) improvements*: collect and process information about your use of the website to provide you with personalized enhanced features, like to remember the choices that you made
- *Sharing data with third parties:* sharing your information with third-parties beyond the website you are visiting
- *Personalizing and measuring content:* create and display personalized content that is relevant to you, content you interact with are measured for performance and effectiveness
- *Personalizing and measuring ads:* deliver, personalize ads, select and measure the effectiveness of these ads. Advertising and marketing material can be shown to you based on the content you're viewing, the app you're using, your approximate location, or your device type. Ads you interact with are measured for performance and effectiveness
- *Analytics, statistics, and audience insights:* measure, improve and report on your engagement with the website service, like the number of unique visits to a website, how long users stay in the site, what parts and pages of the website are browsed, main searched keywords, etc. Apply market research to learn more about audiences who visit sites/apps and view ads
- *Developing and improving products:* Your data can be used to improve existing systems and software, and to develop new products and functionalities
- *Future innovations:* your data can be used for future innovations unrelated to the service the website currently provides
- *Archiving data for scientific or historical research, public interest, or statistical purposes:* your data can be used for future archiving purposes in the public interest, scientific or historical research purposes or statistical purposes
- *Security and debugging:* your data can be used to ensure systems are working properly and securely
- *Fraud detection and law enforcement:* your data can be used to monitor for and prevent fraudulent activity, and indicating possible criminal acts and threats to public safety.

4.1.4 Participants. For both the pre-test and final survey, we surveyed internet users who speak English and have been living in the EU for at least one year, as we wanted participants who have been exposed to the GDPR and cookie policies. We used Prolific and recruited participants with a minimum approval rate of 90% on the platform to ensure high-quality answers. Our institutional review board declared that our study was exempt from federal human subjects regulation.

We conducted a power analysis to determine the sample size for our survey with 8 website category conditions. To achieve high (0.8) statistical power, we needed approximately 400 participants. Since participants were based in the EU, we expected all of them to have seen privacy notices. Yet, one participant said they had never encountered a privacy notice before, we therefore excluded them from the analysis. Participants were compensated 2,70€ in exchange for approximately 15 minutes of their time. From our observations during the pilot studies and median completion time from our pre-test survey, it was determined that 15 minutes was likely to be more than enough time to complete the survey.

In total, we analyzed 399 responses. We had 250 male, 145 female, and 4 non-binary participants. As is the case with most research using online crowdsourcing platforms, our participants were mostly young, with 76% being between 18 and 34 years old [45, 68]. Almost all (96.24%) had lived in the EU for over four years, and over half (54.14%) primarily used the internet in English.

[8] https://cookiepedia.co.uk/classify-cookies

Table 2: When a website tells you they are collecting data for "legitimate interest" purposes, whose legitimate interests do you think they mean? [Select all that apply]

User Group	Count
You (the user)	64
The company offering the service (service provider)	367
3rd party vendors (e.g., advertisers)	255
Society	46
Other users of the service	40
Unsure	26

4.2 Results

Table 3: "Under which scenarios do you think your data would be collected? [Select all that apply]"

Toggle Options	Count
Both toggles selected	355
Consent only selected	249
Legitimate interest only selected	280
No toggles selected	56

4.2.1 User understandings of legitimate interest. Our survey indicated that the majority of participants believe that, when data is collected for legitimate interest purposes, those are the legitimate interests of the service provider and third party vendors. However, a non-negligible number of respondents incorrectly believed that data was collected in the legitimate interest of themselves, other users, or society (see Table 2).

When it comes to the consent/legitimate interest toggles present in privacy notices, most participants also tended to believe that data was being collected unless no toggles were selected, as seen in Table 3. Participants generally have a correct understanding of whom legitimate interests benefit the most and what happens in practice when consent and legitimated interest decision making is amalgamated, but there are some users whose beliefs are incorrect. This result shows that some users are likely making ill-informed decisions with regard to their data processing preferences.

Table 4: User Evaluations of Legitimate Interest Purposes

Legitimate Interest Purpose	Mean	Median
Personalized content and measurement	2.07	2
Personalized ads and measurement	1.80	2
Analytics	2.42	2
Develop and improve products	2.85	3
Future innovations	2.47	2
Archiving data	2.80	3
Security and debugging	3.17	3
Fraud and law enforcement	3.03	3

4.2.2 User evaluations of legitimate interest data collection purposes (out of 5). For the eight legitimate interest purposes we tested, we aggregated our data across all website categories (see Table 4 for mean and median scores). Herein we found that *security and debugging*, and *fraud and law enforcement* were the most likely to be purposes for which users felt comfortable sharing data for without their permission. Wilcoxon signed-rank tests found that *security and debugging* and *fraud and law enforcement* were statistically significant with every single purpose except for each other ($Z = 84603, p = 0.15$). This confirms that users were more comfortable with these purposes compared to other purposes, but users' comfort with *security and debugging* and *fraud and law enforcement* did not differ significantly from each other.

Personalized ad delivery and measurement received the lowest levels of user comfort scores. Wilcoxon signed-rank tests confirmed that the *Personalized ad delivery and measurement* purpose was statistically significant with every legitimate interest purpose, therefore confirming that users felt the least comfortable with this purpose. Refer to Section 7 of the Supplementary Materials for detailed Wilcoxon test results. We found that users are not very comfortable with *develop and improve products*, *archiving data*, *future innovations*, *analytics*, and *personalized content and measurement*, but they are not viewed as negatively compared to *personalized ad delivery and measurement*. Figure 4 illustrates the user ratings for the legitimate interest purposes included in our study.

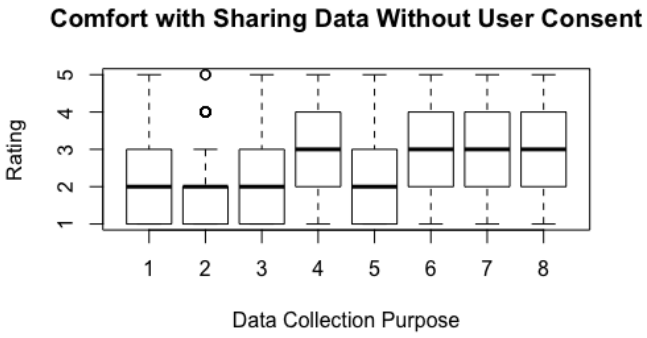

Figure 4: User ratings of how comfortable they are when their data is collected without permission for the following purposes: 1) Personalizing and measuring content, 2) Personalizing and measuring ads, 3) Analytics, 4) Develop and improve products, 5) Future innovations, 6) Archiving, 7) Security and debugging, 8) Fraud and law enforcement. The box displays the median, first, and third quartiles, the dots represent outliers.

4.2.3 Purposes deemed essential. For the three consent-based purposes included in our survey, we found that participants found the *sharing with third parties* and *UX improvement* purposes the least essential, and *functional, strictly necessary* the most essential (see Table 5 for mean and median scores). Figure 5 illustrates user ratings for these purposes in our study. From this, we see that *functional, strictly necessary* purposes could potentially be included as a future

Table 5: User Evaluations of Potential Essential Purposes

Purpose	Mean Score	Median Score
Functional, strictly necessary	3.9	4
UX improvement	3.02	3
Sharing with third parties	1.7	1

Table 6: Regression Table of the Factors Impacting User Acceptance of a Data Collection Purpose

Variable	β	t	p
Purpose benefits the user (themself)	0.43	7.30	< 0.001**
Purpose benefits society	0.17	2.10	0.04*
Purpose benefits other users	0.21	3.56	< 0.001**
Purpose benefits service providers	-0.16	-2.30	0.02*
Purpose benefits third party vendors	-0.19	-3.32	< 0.001**
Participant's knowledge of privacy settings	0.04	0.80	0.43
Participant's age	0.01	0.18	0.86
Participant's gender	0.01	0.14	0.89

legitimate interest purpose. We do not suggest *UX Improvements* or *Sharing with third parties* be legitimate interests due to low ratings.

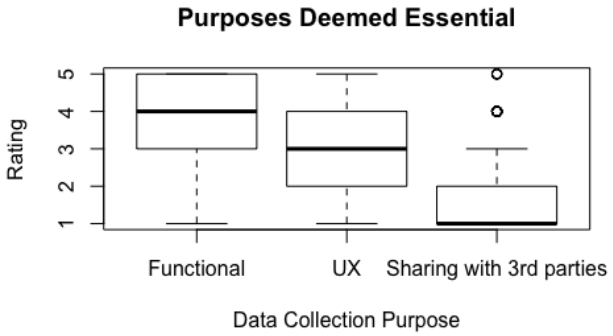

Figure 5: User ratings of how much they deem the following purposes essential for the functioning of their given website category. The box displays the median, first, and third quartiles, the dots represent outliers.

4.2.4 Impacts on user acceptance of a data collection purpose. We ran a multiple regression to see what impacts the user acceptance of a data collection purpose the most (see Table 6 for results). Here, we define *user acceptance* as how comfortable users are with sharing data for a specific purpose without their consent or how essential a purpose is judged to be. We found that user acceptance of a data collection purpose is impacted most by whom the data collection purpose is believed to benefit, $F(58, 341) = 5.99, p < 0.001$, $R^2 = 0.50, R^2_{adjusted} = 0.42$. If users believed a purpose benefits themselves, society, or other users more, they are more likely to accept this purpose, regardless of their demographics or knowledge of privacy settings. However, if a purpose is believed to benefit service providers or third party vendors more, users are less likely to be accepting of this purpose.

4.2.5 Some website categories impact how some data collection purposes are evaluated. Due to our data being likert-scale, we ran one-way Kruskal-Wallis tests for the different data collection purposes to check whether website category impacts user acceptance of the 11 purposes.

We found significant results for *functional and strictly necessary* ($H(7) = 19.32, p = 0.007$) and *sharing data with third parties* purposes ($H(7) = 17.14, p = 0.02$). After identifying purposes where there were significant effects, we ran Dunn's post-hoc tests to identify which website categories were impacted.

The category a website belongs to impacts user opinions of how legitimate interest purposes are used to some extent, but only for news, search engine, and banking websites, and for certain purposes only. Participants found *functional and strictly necessary* purposes more essential for search engine ($p = 0.04$) and banking sites ($p = 0.01$) compared to news sites. Additionally, participants found *sharing data with third parties* to be significantly less essential for banking sites compared to search engine ($p = 0.02$) and news websites ($p = 0.01$). This is a similar finding to previous research which found that users were most concerned about third party tracking on banking sites compared to other kinds of sites [8].

4.2.6 Harms associated with legitimate interest data collection. We introduced an open-ended question in our survey asking participants what harms they believed were associated with data collection for legitimate interest purposes. With these responses, we conducted a qualitative analysis to identify the harms participants mentioned.

Table 7 lists these identified harms, and how often they came up; we explain the meaning of these harms, along with example quotes from participants. In total, we received 457 different responses because some participants listed down several harms. Since this is a qualitative analysis, it is possible that some harms may overlap.

4.3 Demographic Effects

In this section, we ran Spearman correlations between various demographic factors with user acceptance ratings for data collection purposes, and general privacy behaviors. To avoid spurious correlations, we used a Bonferroni correction [52].

4.3.1 The effect of privacy settings knowledge. A participant's privacy settings knowledge was determined by what they rated themselves for their knowledge of privacy settings in the web skills use survey. We found that those with more privacy settings knowledge were more likely to adjust their privacy preferences in privacy notices ($r(397) = -0.18, p < 0.001$), and more likely to feel concerned about their data privacy ($r(397) = 0.17, p = 0.001$). We did not find any significant correlations between privacy settings knowledge and user acceptance of any data collection purposes.

Table 7: Potential Harms of Legitimate Interest According to Users

Harm	Prevalence	Participant Quote
Unwanted advertising: Users are concerned about their data being used for profiling, or just being collected to send them unwanted ads.	16.4% (out of 457 responses)	"Being profiled and getting targeted ads that uses personal data for benchmarking and displaying relevant ads that people don't want"
Data breach: Concerns that if companies that have their data are the victim of a security attack, their data will be compromised.	28.5%	"Companies could be hacked, my personal data could be leaked or easily accessed by someone else"
Privacy concerns: General beliefs about how data collection infringes on one's privacy.	20.6%	"The lack of privacy, as your habits and likes can be easily tracked"
Misuse of data: Beliefs that companies that have access of one's data may sell their data, or use it for nefarious reasons.	15.6%	"Companies could sell data to government or terrorist organizations"
Loss of control: There is a lack of transparency in what happens to user data once they give it away, therefore users are concerned about who has access, and what it is being used for.	11.5%	"I have no control over what companies are holding information on me and what it is being used for, parties are profiting from my information with no compensation to me"
Manipulation of users: Users have an understanding that data can be used for personalizing content and advertisements, and view it as something that can influence them.	6%	"Isolating users in thought bubbles and making different parts of society unable to coexist peacefully"
Are not worried	0.02%	"I see no serious damage, I have nothing to hide."
Unsure	1%	

4.3.2 Age. We found that age has no correlation with whether one modifies their privacy settings in privacy notices, how concerned they claimed to be about their data privacy, or how much control they feel over their data. However, we found that compared to older individuals, younger individuals are more comfortable with certain purposes being used for data collection. Younger individuals are more comfortable with *personalized content delivery and measurement* ($r(397) = -0.14, p = 0.007$), *personalized ad delivery and measurement* ($r(397) = -0.13, p = 0.008$), *developing and improving products* ($r(397) = -0.11, p = 0.02$), *future innovations* ($r(397) = -0.12, p = 0.02$), *security and debugging* ($r(397) = -0.12, p = 0.02$), and *fraud detection and law enforcement* purposes ($r(397) = -0.12, p = 0.02$).

4.3.3 Gender. There were no correlations between gender and whether one changes their privacy settings, or their concern over their data privacy. However, we found that women were more likely than other genders to feel like they had less control over the data that is collected from them online ($r(397) = 0.14, p = 0.006$). Compared to other genders, men were generally more comfortable with certain purposes being used for data collection such as: *personalized content delivery and measurement* ($r(397) = -0.12, p = 0.02$), *analytics, statistics and audience insights* ($r(397) = -0.14, p = 0.005$), *future innovations* ($r(397) = -0.15, p = 0.004$), *archiving data* ($r(397) = -0.10, p = 0.04$), and *security and debugging* purposes ($r(397) = -0.14, p = 0.005$).

5 LIMITATIONS

In our web crawl, we removed websites that were not in English for our analysis, and only analyzed English privacy notices. As a result, our results may not be generalizable to all languages spoken in the EU. Additionally, we ran our crawler back in May 2022, and the internet is constantly updating and changing, therefore we may not capture all the recent updates made to privacy notices.

In study two, participants were generally younger, which is commonly seen in studies using online crowdsourcing platforms [45, 68], therefore our results may not be reflective of how the general population feels about various data collection purposes. To ensure that we got a sufficient number of responses to the open-ended harms question (outlined in Section 4.2.6, we made the question obligatory. Therefore, the proportion of users who claim to be unsure of the harms of legitimate interest may underrepresented in our results. Additionally, we note that the harms participants mentioned might not be specific to legitimate interest-based data collection, but rather to data collection more generally – as we found in Section 4.2.1, not all participants have an accurate understanding of legitimate interest.

It is also possible that because users are not always aware of the various data collection purposes used to collect their data, some purposes we presented could have been misunderstood by our participants. We tried mitigating this by including user-friendly texts for each purpose, but participants could have skimmed our

definitions and based their judgements on what they *think* they know about a given purpose.

6 DISCUSSION

Study one identified how legitimate interest is being used in practice, and *study two* investigated how users perceive these practices. Our findings indicate that we must not only work towards making data processing under legitimate interests more transparent for users, but also that we must improve the surrounding design practices so as to incentivize user engagement with legitimate interest preference mechanisms.

The use of legitimate interest is exploited through deceptive design practices both at UI and linguistic levels. From our findings, we noticed that the implementation of this legal ground has many UI and linguistic deceptive design elements so far unreported in the literature of deceptive designs. In particular, more attention should be paid to deceptive designs that go beyond the UI. Complex deceptive designs often combine multiple dubious practices in one interface [13], as our study found with linguistic and UI deceptive designs being used when presenting legitimate interest. Section 6.1 outlines the deceptive designs we identified in more detail.

Legitimate interest deceptive designs may be intentional. We believe that legitimate interests might be purposefully designed in a way to limit user interactions with these purposes so that websites and CMPs can collect as much user data as possible, and therefore receive more revenue. Several findings from Study 1 support this position. Firstly, only 4.74% of the top 10,000 websites mentioned "legitimate interest(s)" in their privacy notices; presumably, many more sites are using this legal basis without disclosing it to users. Secondly, it is difficult to object to legitimate interests, as 42.41% of sites only mention legitimate interest in the second layer of a privacy notice. Thirdly, 31% of sites included a legally dubious consent/legitimate interest toggle. Lastly, several sites required users to object individually to legitimate interest purposes and vendors. These design elements are problematic because they will likely lead to low user interaction rates. A study by Hils et al. showed that only 1.3% of user interactions with privacy notices involved adjusting toggles to consent to specific vendors/purposes, and most user interactions occur only in the first layer of a privacy notice [41, 58, 71, 73].

IAB Europe's TCF has a major impact on how legitimate interest is applied in practice. Notwithstanding several legitimate interest-related requirements and guidelines (see Section 2.1), we only saw the TCF's legitimate interests purposes being implemented in practice. Notably, websites continue to use the TCF even after this framework was declared to breach the GDPR by using unlawful practices and for collecting data for advertising purposes on the ground of legitimate interests in February 2022 [2, 74]. In September 2022, the TCF was brought to the highest court of the EU (European Court of Justice). [9]

This shows the impact of the TCF in how legitimate interest is implemented in practice and the lack of enforcement to ensure TCF adheres to the legal standard. We noticed that the way TCF-based CMPs implemented privacy notices differed in terms of their design: some CMPs use more deceptive designs than others, which has been reflected in previous research [63]. Since CMPs provide websites design templates for privacy notices at scale, it is important that this market segment is monitored [63, 66].

Users are not fully aware of which legal basis data is being collected on, who is collecting data, or whom it benefits. Users tended to believe that i) legitimate interests benefit service providers and/or third party vendors, and ii) their data was being collected when both or at least one (consent or legitimate interest) toggle was selected in the privacy notice. Even though our findings show that the majority of participants were correct in their assessments, they also highlight that a non-negligible number of users have ill-informed beliefs about legitimate interests. Thus, a question arises of whether all users can object to legitimate interests in an informed manner, especially since legitimate interests are often not disclosed to users in privacy notices or meaningfully explained. The way the legal basis is used might in particular not be transparent to users without a legal background who rely on commonsense understanding of the term "legitimate" and their own intuitions about whose "interests" are being considered.

Users are not fond of personalization nor advertising purposes, despite how often these are used in practice. Overall, the purposes of *personalized ad delivery and measurement* and *personalized content delivery and measurement* received low scores from users, which is at odds with how often personalization purposes are used in practice. However, *personalized content delivery and measurement* was more acceptable to users compared to advertising purposes. This indicates that users may not view personalization as neutral, but instead view it as harmful, as 6% of the identified harms had to do with how personalization could be used to manipulate users. Moreover, data collection purposes often group personalized ads and content together, but our study found that users evaluate personalized content and personalized ads differently, echoing previous findings [8, 50]. Many websites listed several advertising purposes as their legitimate interests, a practice which users tend to disapprove of according to our results, and is illegal [43].

User acceptance of a data purpose is impacted most by whom it is believed to benefit. We found that users tend to think most data collection purposes benefit service providers and third party vendors. There were some variations in user acceptance for some purposes, such as *fraud and law enforcement* being judged to benefit all user groups more equally. Who users believe a purpose is judged to benefit has the biggest impact on user acceptance of a data purpose, regardless of user demographics or privacy knowledge. If a purpose is judged to benefit the users themselves, other users, and/or society more than the service provider and/or third party vendor, then there is more user acceptance for data processing.

6.1 Deceptive Design Elements

We describe the legitimate interest-related deceptive designs identified in our study using existing privacy deceptive design taxonomies which were identified by Bosch et al. [5] and Brignull [40], as well as describing other deceptive designs not previously covered in the literature. Deceptive designs we identified that were previously described by Bosch et al. included *maximize, centralize,* and *obscure* [5].

[9] https://www.dataprotectionauthority.be/iab-europe-case-the-market-court-refers-preliminary-questions-to-the-court-of-justice-of-the-eu

Deceptive designs we identified that were previously described by Brignull included *roach motel* [40] (also called *false hierarchy*) [13]. We also additionally identified the deceptive designs of *complex choice architecture*, and *misleading terminology*.

Maximize means to collect as much personal data as possible—more than what is needed for the task [5]. This deceptive design was exemplified by the following: i) each purpose can only have one legal basis, but our results show that 31% of sites have purposes in privacy notices that rely on both legal bases (consent and legitimate interest); ii) legitimate interest-based purposes and/or vendors were often pre-selected as a default (in privacy notices that allow users to object); and iii) objecting to legitimate interests can be complicated, as it sometimes requires users to individually opt out by purpose and/or vendor, or go through multiple layers of a privacy notice to object.

Centralize refers to when personal data is "collected, stored, or processed at a central entity" [5]. With the presence of various adtech companies and CMPs embedding the TCF, we noticed that most of the same TCF-based CMPs collected data from users across multiple websites.

Obscure means to make it difficult for data subjects to know how their personal data is being collected and used by data controllers [5]. In our analysis, we found that "legitimate interest" is only mentioned in 474 (4.74%) out of the 10,000 websites we analyzed, therefore many users are probably unaware that legitimate interests are being used to collect data without their consent. Only 86.3% of sites disclosed their legitimate interest purposes, but not all sites allowed users to object to such processing easily.

Roach motel, also referred to as a *false hierarchy* [13], are deceptive designs where it is easy for users to enter into a situation, but it is difficult to get out of it [40]. While it is easy for users to accept all purposes, objecting to legitimate interests and managing one's choices after accepting is much more difficult.

The GDPR requires an *opt-in* consent approach, where users manually opt into purposes/vendors they want to consent to when they make their granular consent choices [10]. However, such a rule is not explicitly required for legitimate interest, which is often managed in an *opt-out* manner. Users then need to manually opt out of pre-selected purposes/vendors they do not want to share data with. Such an approach is not effective for reverting preferences for data collection because users tend to not adjust their privacy settings once the defaults have been selected [5, 54].

As mentioned in Section 2.1, practices related to unbalanced choices in privacy notices were deemed illegal and the companies implementing them were fined by DPAs. The same reasoning regarding balanced choices applies to legitimate interest settings (i.e., objecting to such purposes should be as easy as it is to accept or reject consent) and reflects the principle of data protection by design (Art. 25 GDPR).

In addition to the privacy deceptive design patterns identified using previous taxonomies, we also identified two other deceptive design patterns related to legitimate interests:

Complex choice architecture. This term describes a choice architecture that is complicated, therefore confusing to users and/or obscuring information, whether intentionally or accidentally. We found that legitimate interest elements in privacy notices contain many complex choice architectures. Examples include i) legally dubious toggles that rely on both consent and legitimate interests, therefore introducing complex decision making logic, ii) mentioning their respective legitimate interests in deeper layers of a privacy notice, therefore obscuring information, and iii) requiring users to manually opt out of individual vendors and/or purposes, therefore flooding users with multiple decision choices.

Vague terminology. The term "legitimate interest" is used in a vague and ambiguous way in practice. This has created a huge margin of interpretation that is, as our studies show, exploited to the detriment of data subjects' rights. In Article 6(1)(f) GDPR, "legitimate interests" are referred to as the *"legitimate interests pursued by the controller or by a third party"*, yet none of the privacy notices we analyzed used this wording.

The term "legitimate interest" by itself is vague, and we anticipate that many users will not understand whose legitimate interests are being considered. Additionally, describing data processing purposes as "legitimate" implies that these purposes are authentic, or justified, which may not always be the case. This is especially problematic because our analysis indicated that most companies provided unhelpful explanations of legitimate interest that do not explain what it means, and only 86.3% of privacy notices listed the purposes that relied on legitimate interest. The fact that privacy notices do not mention what legitimate interest is could violate the requirement that the interests pursued by data controllers must be clearly articulated and transparent [25]. Also, if purposes are not listed, this might violate the GDPR transparency principle, and might constitute a violation of the legal requirement of specificity and explicitness that demand purposes to be detailed and precise and entail a shared common understanding, irrespective of any different cultural/linguistic backgrounds, level of understanding or special needs [25].

6.2 Implications

Our studies seem to indicate several areas of lack of enforcement around the legitimate interest legal basis. Here, we add legal and design implications drawn from our study to put our results into practice.

Neither participants nor the law approve of advertising purposes under the legitimate interests legal basis, despite its common usage. As observed in Section 2.1, the use of the open-textured term, "legitimate interests" allows controllers to base the processing of personal data on a balancing decision when indeed their own legitimate interests outweigh those of the data subjects. This is necessary because the law (Art. 6(1) GDPR) may not foresee all potential scenarios that may occur in practice.

"Legitimate interest" *is essential to ensure the processing of personal data in cases where these legitimate interests are likewise to be protected as fundamental rights.* However, as outlined above in Section 2.1, advertising does not constitute a legitimate interest. Online behavioral advertising (and consequent profiling of data subjects) is often used to finance online services and data collection for advertising is not "strictly necessary" from the point of view of a website user. Hence this purpose instead requires the legal basis of consent according to Art. 5(3) ePD, as mentioned in other studies [57, 61]. Several policy documents [24, 26, 43] and regulatory decisions [2]

postulated this stance that advertising purposes should only rely on the legal basis of consent and not under legitimate interest.

The results of our second study showed that participants' evaluations of legitimate interests are well in line with this legal assessment as they do not approve of ad-related purposes. In contrast, our first study found that a lot of websites based their data processing on legitimate interests for advertising purposes. This result clearly points to a lack of enforcement and a mismatch between user preferences and reality regarding this legal basis.

Lack of enforcement. The function of "legitimate interests" as a legal basis for cases that were not foreseen by the law has one important consequence: the decision on whether legitimate interests actually outweigh the protection of personal data, and thus are a valid legal basis, is reliant on the case-by-case assessment of executive decisions by DPAs and judicial findings of courts. This means that this huge margin of interpretation ultimately depends on assessing these cases in practice. Otherwise, as our studies show, it may be exploited to the detriment of data subjects' rights. With all this flexibility of the term comes the concern that too many non-compliant cases occur in practice. Our results suggest that this is the point we have now reached, which is that there is effectively a lack of enforcement of legitimate interests in privacy notices. The 29WP already asserted that the lack of a consistent approach of this legal basis may result in a lack of legal certainty and predictability, may weaken the position of data subjects, and may also impose unnecessary regulatory burdens on businesses [25]. Such inconsistencies have already led to litigation before the Court of Justice of the European Union [9, 14, 15].

Data Protection Authorities need to explicitly assert in their guidelines and decisions that balanced (or parity) choices (e.g., accept and object to all buttons, reducing the number of clicks, etc.) apply to legitimate interest-based settings as well, not only to consent settings. DPAs should also closely monitor the current practices of CMPs, since we find that these embed and disseminate, by default, deceptive practices in the privacy notices presented in thousands of websites. In particular, the adoption of guidelines, for example, by the EDPB that specify concrete requirements for the design in case of legitimate interests-based data processing, would be an important step. While these exist for the balancing test itself with Working Paper 217 [25], and on deceptive designs with the EDPB Guidelines on Dark Patterns 3/2022 [23], the cases where deceptive designs are used when legitimate interests are applied as a ground for data processing has not been addressed specifically. Our studies indicate that it is this combination that poses specific risks and that are in fact exploited by data controllers.

Privacy by design recommendations. Art. 25(1) of the GDPR requires data controllers to implement appropriate technical and organizational measures designed to implement data protection by design and by default. As it has been raised by the EDPB Guidelines on Dark Patterns, it would be important to strengthen and further specify ex-ante requirements for both the design of data processing mechanisms focusing on the UI and the language of privacy notices [23, p. 10-11]. Along with previous findings [41, 54, 56, 63, 67], we suggest that possible implementations of the principle of data protection by design and by default (Art. 25 GDPR) could include i) legitimate interest object-all buttons, ii) presenting fewer choices to prevent overwhelming users, and iii) adopting intelligible language within legitimate interest settings using neutral language.

Along these lines, Art. 10 of the Proposal of the ePrivacy Regulation (EU Commission version) [20] envisions the possibility for consent to be expressed at the browser level. This change, if accepted, would reduce the amount of privacy notices and deceptive design practices mentioned in this paper. The results of the trilogue negotiations—currently taking place between representatives of the three bodies involved in the EU legislative process, the EU Commission, the Parliament and the Council of Ministers—might dictate the future of consent expression, though we posit that mechanisms for automated consent have the potential to address deceptive design practices.

7 CONCLUSION

Legitimate interest is a broad legal ground for collecting data under the GDPR that has the potential to be misused. Through a web crawl of 10,000 top websites, our paper empirically investigates how legitimate interest is used in privacy notices, and investigates user perceptions of these practices. In our first study, we identified the various deceptive designs that are applied when implementing legitimate interest in practice. Our second study found that the way legitimate interest is used is not in line with user opinions of how they prefer their data be used, and that user acceptance of a data purpose is impacted by who it is believed to benefit.

Based on the results, we identified deceptive designs being applied when legitimate interests are mentioned in privacy notices, and discuss the legal and design implications of such practices. Our studies indicate a need for better enforcement of the GDPR, reconsidering the way legitimate interests should be used, and consulting with end-users to integrate user opinions about how their data should be collected and processed.

ACKNOWLEDGMENTS

We would like to thank Michèle Finck, Frederik J. Zuiderveen Borgesius, attendees of the Max Planck Law-Tech Symposium, and UC Berkeley's Comparing Effects of and Responses to the GDPR and CCPA/CPRA Symposium for the discussions and providing us with helpful feedback for our paper.

REFERENCES

[1] Ryan Amos, Gunes Acar, Eli Lucherini, Mihir Kshirsagar, Arvind Narayanan, and Jonathan Mayer. 2021. Privacy policies over time: Curation and analysis of a million-document dataset. In *Proceedings of the Web Conference 2021*. 2165–2176.

[2] APD2022 2022. Decision on the merits 21/2022 of 2 February 2022 Complaint relating to Transparency & Consent Framework. https://www.autoriteprotectiondonnees.be/publications/decision-quant-au-fond-n-21-2022-english.pdf

[3] Brooke Auxier, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar, and Erica Turner. 2019. Americans and privacy: Concerned, confused and feeling lack of control over their personal information. *Pew Research Center* (2019).

[4] Adrien Barton and Till Grüne-Yanoff. 2015. From libertarian paternalism to nudging—and beyond. *Review of Philosophy and Psychology* 6, 3 (2015), 341–359.

[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] David M. Boush, Marian Friestad, and Peter Wright. 2009. *Deception in the Marketplace: The Psychology of Deceptive Persuasion and Consumer Self-Protection*. Rouledge.

[7] Centre for Information Policy Leadership. 2017. Recommendations for Implementing Transparency, Consent and Legitimate Interest under the GDPR. https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_

recommendations_on_transparency_consent_and_legitimate_interest_under_the_gdpr_-19_may_2017-c.pdf
[8] Farah Chanchary and Sonia Chiasson. 2015. User perceptions of sharing, advertising, and tracking. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. 53–67.
[9] CJEU. 2014. Judgment in Case C-131/12 Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González. ECLI:EU:C:2014:317.
[10] CJEU. 2019. Judgment in Case C-673/17 Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband eV v Planet49 GmbH. http://curia.europa.eu/juris/documents.jsf?num=C-673/17.
[11] CJEU. 2021. Deliberation of the restricted committee No. SAN-2021-023 of 31 December 2021 concerning GOOGLE LLC and GOOGLE IRELAND LIMITED. https://www.cnil.fr/sites/default/files/atoms/files/deliberation_of_the_restricted_committee_no._san-2021-023_of_31_december_2021_concerning_google_llc_and_google_ireland_limited.pdf.
[12] CNIL. 2022. Deliberation of the restricted committee No. SAN-2021-024 of 31 December 2021 concerning FACEBOOK IRELAND LIMITED. https://www.cnil.fr/sites/default/files/atoms/files/deliberation_of_the_restricted_committee_no._san-2021-024_of_31_december_2021_concerning_facebook_ireland_limited.pdf.
[13] European Commission, Directorate-General for Justice, Consumers, F Lupiáñez-Villanueva, A Boluda, F Bogliacino, G Liva, L Lechardoy, and T 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. https://doi.org/doi/10.2838/859030
[14] Court of Justice of the European Union. 2019. Judgment in Case C-40/17 Fashion ID GmbH and Co.KG v Verbraucherzentrale NRW eV. ECLI:EU:C:2019:629.
[15] Court of Justice of the European Union. 2022. Request for a preliminary ruling Case C-17/22. https://curia.europa.eu/juris/showPdf.jsf?text=&docid=255645&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=6693245.
[16] Dutch DPA. 2019. Dutch DPA guidance on legitimate interest. https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/normuitleg_gerechtvaardigd_belang.pdf.
[17] Malin Eiband, Daniel Buschek, Alexander Kremer, and Heinrich Hussmann. 2019. The impact of placebic explanations on trust in intelligent systems. In *Extended abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.
[18] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. 2019. Exploring how privacy and security factor into IoT device purchase behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
[19] ePD-09 2009. Directive 2009/136/EC of the European Parliament and of the Council. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:337:0011:0036:En:PDF.
[20] ePR-17 2017. Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications). https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:337:0011:0036:En:PDF.
[21] IAB Europe. 2021. Transparency and Consent Framework. https://iabeurope.eu/tcf-2-0/.
[22] European Commission. 2022. Letter to the Dutch DPA on legitimate interests. https://static.nrc.nl/2022/pdf/letter-dutch-dpa-legitimate-interest.pdf.
[23] European Data Protection Board. 2022. Dark patterns in social media platform interfaces: How to recognise and avoid them. https://edpb.europa.eu/system/files/2022-03/edpb_03-2022_guidelines_on_dark_patterns_in_social_media_platform_interfaces_en.pdf
[24] European Data Protection Board (EDPB). 2010. Opinion 2/2010 on online behavioural advertising, 22 June 2010, (WP171). https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp171_en.pdf
[25] European Data Protection Board (EDPB). 2014. Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC (WP 217). https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf
[26] European Data Protection Board (EDPB). 2019. Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects. https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_en.pdf
[27] European Union: Council of the European Union. 2007. The Charter of Fundamental Rights of the European Union (2007/C 303/01), C 303/1.
[28] Federico Ferretti. 2014. Data protection and the legitimate interest of data controllers: Much ado about nothing or the winter of rights? *Common Market Law Review* 51, 3 (2014).
[29] Michèle Finck and Asia Biega. 2021. Reviving purpose limitation and data minimisation in personalisation, profiling and decision-making systems. *Technology and Regulation* 44, 61 (2021).
[30] Centre for Informational Policy Leadership. 2017. Recommendations for Implementing Transparency, Consent and Legitimate Interest under the GDPR. *Centre for Information Policy Leadership GDPR Implementation Project* (2017).
[31] Forbrukerradet. 2018. Deceived by Design: How tech companies use dark patterns to discourage us from exercising our rights to privacy. https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf. Accessed: 2022-03-30.
[32] Elena Gil González and Paul De Hert. 2019. Understanding the legal provisions that allow processing and profiling of personal data—an analysis of GDPR provisions and principles. In *Era Forum*, Vol. 19. Springer, 597–621.
[33] PAJ Graßl, HK Schraffenberger, FJ Zuiderveen Borgesius, and MA Buijzen. 2021. Dark and bright patterns in cookie consent requests. *Journal of Digital Social Research* (2021).
[34] 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*. 1–14.
[35] 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* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 172, 18 pages. https://doi.org/10.1145/3411764.3445779
[36] 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 (CSLAW '22)*. https://doi.org/10.1145/3511265.3550448
[37] Hana Habib, Megan Li, Ellie Young, and Lorrie Cranor. 2022. "Okay, whatever": An Evaluation of Cookie Consent Interfaces. In *CHI Conference on Human Factors in Computing Systems*. 1–27.
[38] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2020. "It's a scavenger hunt": Usability of websites' opt-out and data deletion choices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.
[39] Eszter Hargittai and Yuli Patrick Hsieh. 2012. Succinct survey measures of web-use skills. *Social Science Computer Review* 30, 1 (2012), 95–107.
[40] Harry Brignull. 2022. Deceptive Design. https://www.deceptive.design/
[41] Maximilian Hils, Daniel W Woods, and Rainer Böhme. 2020. Measuring the emergence of consent management on the web. In *Proceedings of the ACM Internet Measurement Conference*. 317–332.
[42] Xuehui Hu, Nishanth Sastry, and Mainack Mondal. 2021. CCCC: Corralling cookies into categories with CookieMonster. In *13th ACM Web Science Conference 2021*. 234–242.
[43] Information Commissioner's Office. 2019. Update report into adtech and real time bidding. https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906-dl191220.pdf
[44] Information Commissioner's Office. 2022. What is the 'legitimate interests' basis? https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/what-is-the-legitimate-interests-basis/
[45] Panagiotis G Ipeirotis. 2010. Demographics of Mechanical Turk. *NYU working paper no. CeDER-10-01* (2010).
[46] Eric J Johnson, Suzanne B Shu, Benedict GC Dellaert, Craig Fox, Daniel G Goldstein, Gerald Häubl, Richard P Larrick, John W Payne, Ellen Peters, David Schkade, et al. 2012. Beyond nudges: Tools of a choice architecture. *Marketing letters* 23, 2 (2012), 487–504.
[47] Irene Kamara and Paul De Hert. 2018. Understanding the balancing act behind the legitimate interest of the controller ground: A pragmatic approach. *Brussels Privacy Hub* 4, 12 (2018).
[48] Konrad Kollnig, Reuben Binns, Max Van Kleek, Ulrik Lyngs, Jun Zhao, Claudine Tinsman, and Nigel Shadbolt. 2021. Before and after GDPR: Tracking in mobile apps. *Internet Policy Review* 10, 4 (2021).
[49] Konrad Kollnig, Pierre Dewitte, Max Van Kleek, Ge Wang, Daniel Omeiza, Helena Webb, and Nigel Shadbolt. 2021. A Fait Accompli? An Empirical Study into the Absence of Consent to Third-Party Tracking in Android Apps. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*. 181–196.
[50] Anastasia Kozyreva, Philipp Lorenz-Spreen, Ralph Hertwig, Stephan Lewandowsky, and Stefan M Herzog. 2021. Public attitudes towards algorithmic personalization and use of personal data online: Evidence from Germany, Great Britain, and the United States. *Humanities and Social Sciences Communications* 8, 1 (2021), 1–11.
[51] Oksana Kulyk, Annika Hilt, Nina Gerber, and Melanie Volkamer. 2018. "This website uses cookies": Users' perceptions and reactions to the cookie disclaimer. In *European Workshop on Usable Security (EuroUSEC)*, Vol. 4.
[52] Sangseok Lee and Dong Kyu Lee. 2018. What is the proper way to apply the multiple comparison test? *Korean Journal of Anesthesiology* 71, 5 (2018), 353–360.
[53] Luxembourg DPA. 2021. Decision regarding Amazon Europe Core S.À RL. https://cnpd.public.lu/fr/actualites/international/2021/08/decision-amazon-2.html.

[54] Dominique Machuletz and Rainer Böhme. 2020. Multiple purposes, multiple problems: A user study of consent dialogs after GDPR. *Proceedings on Privacy Enhancing Technologies* (2020), 481–498.

[55] 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.

[56] 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, 791–809.

[57] Célestin Matte, Cristiana Santos, and Nataliia Bielova. 2020. Purposes in IAB Europe's TCF: Which legal basis and how are they used by advertisers?. In *Annual Privacy Forum*. Springer, 163–185.

[58] Aleecia M. McDonald and Lorrie Faith Cranor. 2009. The Cost of Reading Privacy Policies. *Journal of Law and Policy for the Information Society* (2009).

[59] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.

[60] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica* 22, 3 (2012), 276–282.

[61] Victor Morel, Cristiana Santos, Yvonne Lintao, and Soheil Human. 2022. Your Consent Is Worth 75 Euros A Year - Measurement and Lawfulness of Cookie Paywalls. In *Proceedings of the 21st Workshop on Privacy in the Electronic Society* (Los Angeles, CA, USA) *(WPES'22)*. Association for Computing Machinery, New York, NY, USA, 213–218. https://doi.org/10.1145/3559613.3563205

[62] Arvind Narayanan, Arunesh Mathur, Marshini Chetty, and Mihir Kshirsagar. 2020. Dark Patterns: Past, Present, and Future: The evolution of tricky user interfaces. *Queue* 18, 2 (2020), 67–92.

[63] 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*. 1–13.

[64] Thomas V Perneger, Delphine S Courvoisier, Patricia M Hudelson, and Angèle Gayet-Ageron. 2015. Sample size for pre-tests of questionnaires. *Quality of Life Research* 24, 1 (2015), 147–151.

[65] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. 2019. Tranco: A research-oriented top sites ranking hardened against manipulation. *Proceedings of the 26th Annual Network and Distributed System Security Symposium (NDSS)* (2019).

[66] Cristiana Santos, Midas Nouwens, Michael Toth, Nataliia Bielova, and Vincent Roca. 2021. Consent Management Platforms under the GDPR: processors and/or controllers?. In *Annual Privacy Forum*. Springer, 47–69.

[67] 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*. 187–194.

[68] Eunjin Seong and Seungjun Kim. 2020. Designing a crowdsourcing system for the elderly: A gamified approach to speech collection. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–9.

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

[70] Siny Tsang, Colin F Royse, Abdullah Sulieman Terkawi, et al. 2017. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi Journal of Anaesthesia* 11, 5 (2017), 80.

[71] Joseph Turow, Michael Hennessy, and Nora Draper. 2018. Persistent Misperceptions: Americans' Misplaced Confidence in Privacy Policies, 2003–2015. *Journal of Broadcasting & Electronic Media* 62, 3 (Jul 2018), 461–478. https://doi.org/10.1080/08838151.2018.1451867

[72] European Union. 2022. Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act).

[73] 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*. 973–990.

[74] Michael Veale, Midas Nouwens, and Cristiana Santos. 2022. Impossible Asks: Can the Transparency and Consent Framework ever authorise real-time bidding after the Belgian DPA decision? *Technology and Regulation* 2022 (Feb. 2022), 12–22. https://doi.org/10.26116/techreg.2022.002

[75] Ari Ezra Waldman. 2020. Cognitive biases, dark patterns, and the 'privacy paradox'. *Current Opinion in Psychology* 31 (2020), 105–109.

[76] Tal Z Zarsky. 2016. Incompatible: The GDPR in the age of big data. *Seton Hall L. Rev.* 47 (2016), 995.

[77] Shoshana Zuboff. 2019. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.