UI Dark Patterns and Where to Find Them: A Study on Mobile Applications and User Perception

Linda Di Geronimo
University of Zurich
Zurich, Switzerland
digeronimo@ifi.uzh.ch

Larissa Braz
University of Zurich
Zurich, Switzerland
larissa@ifi.uzh.ch

Enrico Fregnan
University of Zurich
Zurich, Switzerland
fregnan@ifi.uzh.ch

Fabio Palomba
University of Salerno
Fisciano, Italy
fpalomba@unisa.it

Alberto Bacchelli
University of Zurich
Zurich, Switzerland
bacchelli@ifi.uzh.ch

ABSTRACT

A Dark Pattern (DP) is an interface maliciously crafted to deceive users into performing actions they did not mean to do. Although design experts have reported on DPs extensively, little effort has been made to study how pervasive they are, especially in mobile applications. In this work, we analyze DPs in 240 popular apps and conduct an online study with 589 users on how they perceive DPs. The results of the analysis showed that 95% of apps contain one or more forms of DPs and, on average, popular applications include at least seven different types of deceiving UIs. The online study shows that most users do not recognize DPs, and they would change their behavior on app usage once informed about them. We discuss the impact of our work and what measures could be applied to alleviate malicious design issues.

Author Keywords

Dark Patterns; Ethical Design; User Experiments

CCS Concepts

•**Human-centered computing** → **HCI theory, concepts and models;**

INTRODUCTION

Over the last decade, the CHI research community has seen an increasing interest in investigating critical aspects of UX practice, not only related to the impact of UX on the society [26, 29, 60, 79], but also from the perspective of designers and the way they apply responsible changes [53, 55, 89]. One of the outcomes of such interest is the definition of *Dark Patterns* (DPs)—user interfaces that trick the users into doing something they did not mean to do [31]. For example, DPs include sneaking unwanted items into the basket, adding users to costly subscriptions, and misleading with double negatives (e.g., ☒ *Uncheck here not to download the add-on*). DPs can also lead users to over-share personal information [38, 95], thus potentially leading to privacy breaches. Users might involuntarily accept to share personal data or give more permission than intended.

Researchers have been studying Dark Patterns under different lenses. For instance, Moser et al. [62], analyzed 200 top e-commerce websites and found multiple UI elements that trigger buying in most websites. Similarly, Mathur et al. [61] found that 11% of 11k e-commerce top web applications use some forms of DPs in their designs. Moreover, substantial effort has been spent on the elicitation of taxonomies to categorize different types of DPs [33, 42]. One of the most recent studies has been presented by Gray et al. [42], who proposed five different types of DPs covering various aspects like redirection from a task to another or UI malicious interferences.

In this work, we continue the academic discourse on Dark Patterns by exploring two new angles: (1) how **prominent** Dark Patterns are in popular mobile apps and (2) whether **users are aware or can recognize the presence of DPs**. In fact, while previous studies aimed at presenting the existence of Dark Patterns or at classifying their different categories, there is still a noticeable lack of knowledge on how prominently they appear in popular mobile apps and on the users' perceptions. The case of mobile apps is critical because of their extreme pervasiveness and role in social life [65].

In particular, we analyzed 240 apps (30 for each of the 8 main categories of applications on the Google Play Store [10]) to identify the instances of DPs they contain, classifying them into the taxonomy proposed by Grey et al. [42]. Unlike all previous works in the field [42, 61, 62]—which classified dark patterns by analyzing screenshots of segments of pages—we applied an active process in which two researchers jointly used each app, performing a series of common tasks to reach certain goals (e.g., creating an account, visiting the setting page), similarly to cognitive walkthrough techniques [67].

CHI'20, April 25–30, 2020, Honolulu, HI, USA

ISBN 123-4567-24-567/08/06…$15.00

DOI: http://dx.doi.org/10.475/123_4

From this study, we found that mobile apps have, on average, more than seven instances of DPs.

Subsequently, we conducted an online experiment using some of the DPs found during the classification phase and studied whether users could perceive them (*DP-blindness*). We found that users often cannot identify the presence of some malicious UI interactions, underlining the need for proper mechanisms to make users aware of malicious UIs and their potential threats.

We make the following contributions: (i) an analysis of DP prevalence in popular mobile apps; (ii) a dataset containing the recording of each app and its DPs classification;[1] and (iii) the results and discussion of an online evaluation on the users' perception on DPs.

BACKGROUND

A popular instance of DP is Homer's Trojan horse [57]. The wooden trap, disguised as a gift, was used by the Greek soldiers to confuse the Trojan adversaries. Nowadays, it seems that we come across similar tricks on a recurrent basis: Many web/mobile applications give the impression of tricking users, for example hiding relevant data or options. This led researchers to debate how ethical modern UX design is and investigate what DPs are used in the digital world we are living.

Ethical UX Design

Improving user interfaces and their usability is one of the main focuses of human-computer interaction. Frameworks, guidelines and various techniques have been proposed to improve the user experience of applications [51, 68, 69, 78]. The ten heuristics by Nielsen, established in 1994, have been the foundation for further improvements of user interfaces [66]. With the advent of mobile devices, additional guidelines and rules have been proposed to tackle new challenges [46, 64, 75, 96].

Nevertheless, a usable application does not imply an ethical one. Although there is no widely established definition of 'Ethical UI', experts in the field have provided their take on it. For example, Karr stated on ethical design [49]:

> "... I like to think of ethical things as thoughts, words, behaviors, designs, systems, and customs that are cumulatively more beneficial to life than they are harmful."

Also, Latham, from the UX Collective [56], connects ethical design to personal freedom and discusses that subtle manipulations in adverts and digital media may condition our choices.

Gray et al. [42] emphasize the important ethical aspects of DPs. While design is—by definition—a persuasive act and has the potential to manipulate the user [70, 71, 80], there are occasions where designers may abuse this power. In this respect, the HCI community is working toward a design that is more ethical [40, 44, 76, 81, 82] also for future UIs such as home robots and proxemic interactions [43, 54].

[1]For double-blind and privacy reasons each video must be manually anonymized. Given the resources necessary to perform this anonymization, we are publishing the dataset iteratively over the next months. The current subset serves as an example [4]. Furthermore, given the large dimension of the videos directory, the complete dataset would be provided upon request.

In the context of ethical design, researchers often have criticized neuromarketing strategies [63, 83, 90, 94]. Neuromarketing is a new field that uses techniques such as fMRI [47], EEG [50], and gaze detection [52] to investigate the effects of marketing inputs [27, 50, 91, 93]. Among various discoveries, neuromarketing research found that the feeling of "loosing out" is particularly effective in influencing users [36, 37, 87]. Based on this finding, e-commerce websites use countdowns and limited offers to pressure customers [32, 45, 62]. Although neuromarketing "choices" may improve user engagement, as well as fasten certain interactions on the website [41, 59, 88], they may become unethical when employed to coerce users [62].

Researchers have classified the artificially created sense of urgency and scarcity (included in the design of many e-commerce websites) as a Dark Pattern [31, 33, 42]. However, DPs go beyond shopping activities. Games, social media, news applications, and more can all include malicious designs [28, 30, 31, 39, 95]. Since most teens today use mobile devices extensively (95% of adolescents have access to a smartphone and spend significant portions of their days consuming media on mobile phones [25]) and minors are more easily manipulable [34, 72, 92], the presence of DPs in mobile applications becomes urgently relevant. Despite this, there is still a lack of knowledge on the prominence and types of malicious designs in everyday mobile applications—a gap that we address with the first part of this work.

Luguri and Strahilevitz [58] discuss DPs from a legal perspective and employ an online survey to study the impact of more aggressive DPs on users. In particular, Luguri and Strahilevitz faked a subscription system, where users were asked to accept or decline the offer of a six months (not free) data protection plan. In the mild version of the DP, users could either 'Accept (recommended)' the program or click on 'Other options', where they could eventually refuse the plan. In the aggressive version of the pattern, upon decline, users were asked to read additional information about identify theft and then wait ten seconds. The authors found that 26% (mild option) and 42% (aggressive option) of the treated participants accepted the plan, in contrast to only 11% among the participants without DPs. With the second part of this work, we extend on the findings discussed by Luguri and Strahilevitz, by studying DP-blindness. We hypothesize that DPs may also work because users are not always aware of the presence of DPs, especially in mild cases. DP-blindness may be the new Ads on display-blindness [73, 35, 74, 48].

Taxonomies of Dark Patterns

The `darkpatterns.org` portal (established in 2010 by Brignull [31]) collects various examples of DPs on web and mobile applications, gathered through the reports of users via Twitter. Brignull's goal is to raise awareness on the topic while also proposing a classification of DPs into different categories. The examples tweeted with the hashtag `#darkpatterns` populate the 'Hall of Shame' of the portal.

Conti and Sobiesk [33] proposed a taxonomy with eleven classes of DPs with twenty subclasses. Among the various categories, authors included `Distraction` (e.g., colors or

blinking animations used to attract users) and `Forced Work` (e.g., force users to watch an Ad) as types of DPs. The most recent taxonomy of DPs was proposed by Gray et al. [42], who have re-defined Brignull's taxonomy, starting from a set of artifacts gathered from blogs, websites, and social media. The categorization of DPs, as delineated by Brignull, was made sharper and more general.

Gray et al. [42] proposed five different types of DPs:

`Nagging` is defined as a redirection from the current task that can happen one or more times.

`Obstruction` patterns block the task flow, making it harder to perform it. The `Obstruction` class includes three subclasses: `Intermediate Currency` (multiple currencies, such as game gems), `Price Comparison Prevention` (uncopiable product names), and `Roach Motel` (easy to open an account, yet hard to delete it).

`Sneaking` patterns try to disguise relevant information to the user. This category comprises four subclasses: `Bait and Switch` (a certain action seems to have a specific result; instead it causes another, unwanted outcome), `Hidden Costs` (an item costs *X* but in the basket its value increases), `Sneak into Basket` (unwanted items are added in the basket), and `Forced Continuity` (e.g., subscription is automatically continued after free trial expires).

`Interface Interferences` are UI manipulation that are biased towards certain UIs elements. This includes: `Hidden Information` (options to accept conditions are small/greyed-out), `Preselection` (unfavorable options are preselected), and `Aesthetic Manipulation` (distracting manipulation of the UI). This last subclass has four subsubclasses: `Toying with emotions` (countdown to offers), `False Hierarchy` (one option is more prevalent), *Disguised Ad* (interactive games), and `Trick Questions` (double negatives).

`Forced Action` coerces users into performing certain tasks to obtain something. Three classes belong to this type: `Social Pyramid` (adding friends to obtain benefits), `Privacy Zuchering` (sharing more personal data than intended), and `Gamification` (forced grinding tasks to obtain something otherwise available with money).

In this work, we use the aforementioned taxonomy by Gray et al. [42], because it is the most updated. Although this taxonomy proved to perform well for our task, we had to extend the original meaning of `Aesthetic Manipulation` and `Forced Action` classes to include a few new DPs instances (Section 3.2.3 details how we extended the taxonomy).

THE PERVASIVENESS OF DPS IN TOP MOBILE APPS

We carry out a classification of malicious designs on 240 trending applications found on Google Play Store [10]. In the next sections, we report on how we executed the study and the results obtained.

Corpus Generation

We focus on applications with the following features: (1) available on the Android platform, (2) free of charge to download, and (3) trending in the US market.

We pick Android as it is the most popular platform among smartphone and tablet users [84]. Similarly, we only focus on free-to-use apps, due to their higher popularity [86].[2] Finally, we decide to study famous applications, to best sample apps that users may experience in their everyday life.

While it is not possible to obtain the names of most downloaded apps from the Google Play store, we can gather the list of the most trending ones. The Google algorithm that calculates apps ranking is not public, and it has changed over the years. From Google official announcements, it is nevertheless clear that app downloads, speed, and user engagement are some of the parameters used in this ranking [77]. Such ranking of apps is suited for the scope of our study since it also considers new apps that have gained high popularity in the latest period. For instance, in the months we performed the classification, the FaceApp [6] application gained much traction in a short time. Probably, considering an overall download ranking, this application would have had a hard time reaching the number of downloads of more senior apps, therefore would not have been added to our list.

The Google Play Store organizes apps in eight main categories: Photography, Family, Shopping, Social, Music and Audio, Entertainment, Personalization, and Communication. We exclude the Personalization category because apps in this category are all composed of Android launchers (e.g., set of icons, widgets, wallpapers), which are extremely different from the rest of the studied applications. Although news apps are popular among users [85], we note that they were scarce in the remaining seven categories. Therefore, we include a News and Magazines category on the list.

For each category, we select the 30 most trending mobile apps. This selection was performed by using a crawler that collected data from the SensorTower [18] website, which allows users to see the list of most trending apps. The crawling was executed among the 12th and 13th of July. The country of selection was set as the US (Europe and global selections were not possible), which had the broadest and biggest range of users of Western countries that could be selected on SensorTower.

The crawler logged 400 most trending free apps of each category and saved additional information about each application (e.g., number of installs, user ranking, and number of reviews). From this list, we selected the top 30. However, certain apps had to be skipped for one of the following reasons: (i) the application was not available anymore, (ii) the application was not available in our country, (iii) the app already appeared in a previous category, (iv) the app is a launcher. If an app needed to be skipped, we included the next one in the ranking.

Our final list was therefore composed of 30 most trending apps for eight different categories, for a total of 240 Google Play

[2] In the case of Netflix, which was the only free app with a paywall, we subscribed to their free-month service to use the application.

application. The list also includes well-known applications such as Facebook [7], Amazon [**?**], Twitter [23], Netflix [13], and Spotify [20].

Methodology

In the studies by Moser et al. [62] and Mathur et al. [61], researchers have collected screenshots of segments of pages to recognize malicious designs in e-shopping websites. In some instances, though, one can infer the presence of DPs only interacting with the artifact. For instance, `Bait and Switch` is a design that changes the meaning of certain actions to trick the user. Clicking on a download button should mean that the user wants to download a selected item, not showing an Ad asking to upgrade to premium. The Ad per se may not contain DPs, but the interaction needed to reach that interface is malicious. For this reason, we analyze each app while in use, instead of relying on static images. Particularly, we first record example usage of each app, then we classify the resulting videos.

Recording Methodology

The recording process was split among two authors of this paper, who used half of the apps each. To record each app, both authors used a One Plus 5, with the latest Android version [14]. Two new Gmail [9] accounts were created to perform the study. Furthermore, two new sim cards were also bought to protect the privacy of the researchers while using the apps.

Every application was used for ten minutes, for a total of 2,400 minutes (i.e., 40 hours) of recorded usage. During the ten minutes, the researchers performed the following tasks (when available), similarly to an inspection walkthrough [67]: (i) creating an account and log out; (ii) closing and reopening the app; (iii) visiting the market page; (iv) going to the setting page; (v) continuing shopping until checkout; (vi) trying to select product names in e-shopping; (vii) using the app for its intended use (e.g., playing games, browse news article).

This walkthrough protocol ensured consistency in our method but has the drawback that it does not cover the cases of apps with hidden features or mechanisms that are only unlocked after an app has been used for a while. Moreover, each app is the new state: It has never been opened before the beginning of our recording.

We did not purchase any products or services in apps. In e-shopping applications, we stopped right before buying the item(s). Although specific DPs may appear only after performing a purchase, we could not afford to buy products for each considered app. We did, however, subscribe to free services if the app was not usable without registering (as for Netflix).

Classification Methodology

After the recording, we randomly selected 40 of the 240 apps (five for each of the eight categories) and classified instances of DPs following the taxonomy of Grays et al. [42]. The classification was performed in pair by the first two authors of the paper. In this phase, both researchers analyzed the videos together to mitigate the risk of DP-blindness. Disagreements on a specific DP were noted for later analysis; these cases were then discussed with a third researcher, also knowledgeable about DPs: The final decision on the DP classification was taken by majority voting. This initial set allowed us to understand the power of the considered taxonomy, as well as to decide additional rules for the classification of future DPs.

In contrast to previous work [62, 61], the two researchers continued the classification process together for the remaining 200 apps. In fact, we found that DP-blindness, especially in video recordings, also affect experts in the field. After the entire classification process (which lasted 120 hours), the two authors double checked the classification sheet to find possible mistakes.

We did not count re-occurrences of DPs, meaning that each DP was reported only the first time it appeared. We consider DP as a re-occurrence of a previous one if the same UI would appear by performing a similar interaction (e.g., clicking on a button, opening the setting page). We made this decision to reduce the effect of how the app was used during the recording. Instead, we kept track of DPs if the same interaction would give a different malicious UI design as a result.

Taxonomy Adaptation and DPs Interpretation

Although we found the taxonomy by Grays et al. [42] to be descriptive enough after the testing phase, we had to extend and adapt it to our scenario.

First, we could not include `Forced Continuity` and `Gamification` classes. For the former class, apps continue users subscription also after the end of the plan; therefore, this class requires one to subscribe to certain programs, which we did not do during the recording phase. The `Gamification` class forces the user to repeat certain actions (often dull) to continue in the game. Unfortunately, this instance of DP is hard to perceive in the first ten minutes usage of an application. Especially at the beginning of a game, it seems that app authors try to increase user engagement and propose more interesting features.

We found instances of DPs that were not explicitly included in the considered taxonomy. For instance, watching an Ad to unlock certain features was not described. However, we found that this DP may easily fit in the `Forced Action` category (e.g., force users to perform actions to obtain something in return) [42, 33].[3]

Understanding designers' intentions and ethical decisions is hard and may lead to imprecision; thus, we limited our research exclusively to the final UI product. Therefore, in every occasion in which an interface seemed to benefit the app rather than the user, we classified the design as a DP. For instance, if an app asks for location permissions and the UI seems to prefer the 'accept' option, we consider it as a malicious design (`False Hierarchy` in this case), even though the designers may have intended this feature to speed up the interaction process.

Furthermore, to improve consistency and reduce subjectivity in the classification, we limited the number of cases we consider

[3]Few additional adaptations and classification rules are discussed in our dataset [4].

Table 1. Dark Patterns and their associated subclasses, according to the considered taxonomy. The global label indicates whether the DP can only appear once in an app (S) or multiple times (M). DA = Disguised Ad; AM = Aesthetic Manipulation; NG = Nagging; SP = Social Pyramid; FA = Forced Action; FH = False Hierarchy; RM = Roach Motel; HI = Hidden Information; PZ = Privacy Zuchering; PCP = Price Comparison Prevention; PRE = Preselection; BAS = Bait And Switch; SIB = Sneak into Basket; TWE = Toying With Emotions; TQ = Trick Questions; IntCur = Intermediate Currency. The subclasses from the original taxonomy that were not found during the classificaiton are omitted.

DP Cases	Classes	M/S
Ad with interactive game	DA	M
Moving ads button	AM	M
Small close button on ad	AM	M
A popup appears and interrupts the user in their task	NG	M
Invite friends to get something in return	SP	S
Ad appears as normal content	DA	M
A sponsored content not clearly different from rest of the content	DA	M
Icons\buttons are ads, but it's not clear	DA	M
Countdown on ads	FA	M
Daily\weekly rewards or features	FA	S
Login to obtain some rewards\bonus	FA	S
Countdown on rewards	FA	S
Watching ad to unlock feature	FA	S
There are two or more options, but the one that is more beneficial for them is more prominent	FH	M
Terms of service is small and\or greyed out	HI	M
Countdown offer	TWE	M
Multiple currencies	IntCur	S
Shame user for not doing something	TWE	M
Popup to rate	NG	M
Unable to select product names (while shopping)	PCP	S
The notifications (and\or emails and sms) are preselected	PRE	S
The option is preselected	PRE	M
App already follows pages by default	PRE	M
Send usage data preselected	PRE, PZ	M
Private settings related dps	PZ	M
Not possible to delete account	RM	S
Not possible to logout	RM	S
Sneak into basket unwanted items	SNE	S
Double negatives in selections	TQ	M
It looks like you have to login, but you can actually use the feature (app) for free	AM	M
User clicks a feature (which does not look like a premium) and get a PRO ad or open google play	BAS, DA	M

as DPs. For instance, for the `Bait and Switch` class, we only include the case in which the user clicks on a feature that looks available to be used for free, and instead finds out that it is accessible for premium users only or by downloading another app. During the classification of the first 40 apps, we found the `Bait and Switch` category to be too generic, thus too subjective; the aforementioned more conservative approach mitigated this issue. Overall, we considered 33 DP cases, for the 16 subcategories [42] (see Table 1). Given the nature of some of the studied DPs (e.g., "It was not possible to delete account"), twelve cases could be counted only once per app (S in Table 1). Each DP case may include more than one DP class. Furthermore, DPs are not mutually exclusive, as one case may appear in conjunction with one or more other DP cases. In the results, we refer to DPs as the number of occurrences of all subcategories.

Results

Among the 240 studied apps, 95% included one or more DPs in their interfaces. Overall, 1,787 DPs were found among all apps, with an average of 7.4 malicious designs per application (std. dev.: 5). Almost 10% of the apps included 0, 1 or 2 DPs (N=33), 37% of the apps contained between 3 to 6 DPs (N=89), while the remaining 49% included 7 or more (N=118).

DP Classes in Mobile Apps

Among the five DP macro-categories, apps contains an average of 2.7 classes each (std. dev.: 1.1), with 37% of the apps including 3 (N=89), 25% with 4 or 5 (N=62), 23% having 2 (N=55), and 14% including 1 or none (N=34).

Considering the 16 subcategories (Table 1), apps contained 4.3 classes on average (std.: 2.6). Most apps (63% N=152) contain at least four different subcategories.

In Figure 1 we show the total number of DPs by each subcategory and the percentage of apps with at least one occurrence of that subcategory. The most frequent DP subcategory was `Nagging`, followed by `False Hierarchy` and `Preselection`. Most apps (55%) interrupted the users in some way, to ask permissions, rate their product, or to show Ads. Often these popups gave one or more option to the user, and many times the alternative that benefited the app was aesthetically favored (see Figure 2). This contributed to the FH class being present in 61% of the apps.

60% of apps also include `Preselection` DPs. The most frequent DP among this subclass is notification preselection (push, email, and SMS) (N=121). Among these applications, 81 contain more than two notifications already preselected.

The app with most DPs in our corpus was Call Free - Free call[3], with a total of 23 DPs, belonging to 10 different subcategories. Wish [24] followed, with 20 different malicious designs in 8 subcategories. While only twelve apps contained no DPs (5% out of 240 applications). Among these: Snapseed [19], Lego Juniors [12], and Barcode Scanner [2].

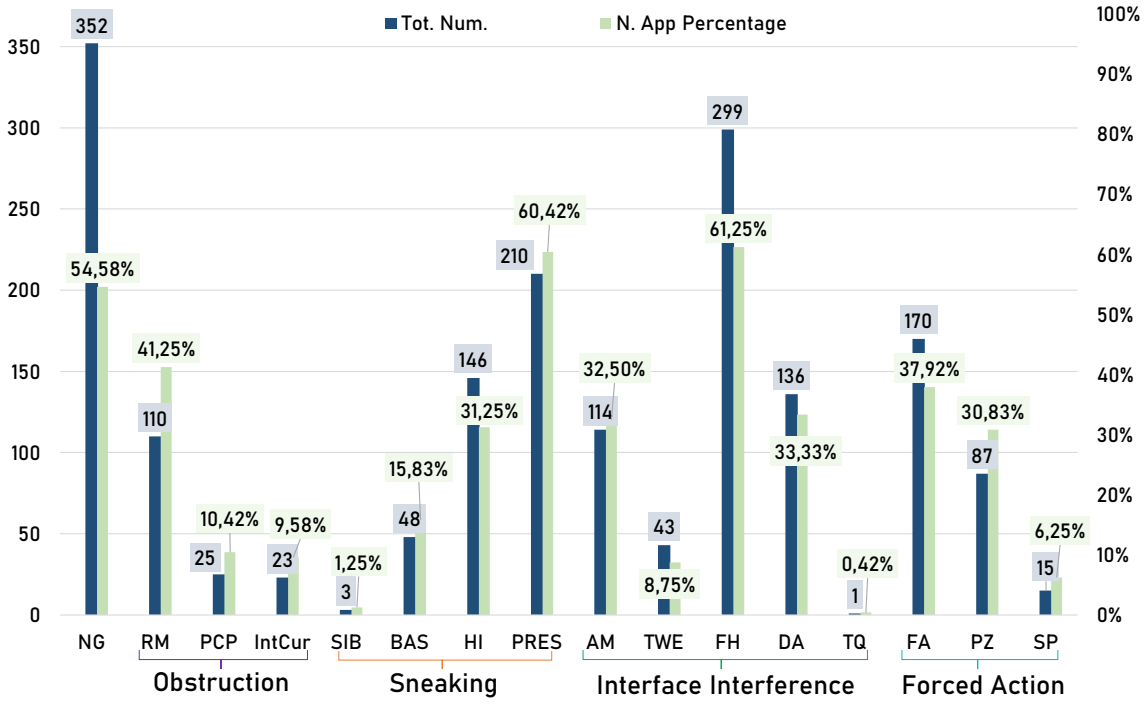

Figure 1. Total number of DPs by subcategories and percentage of apps containing that subcategory.

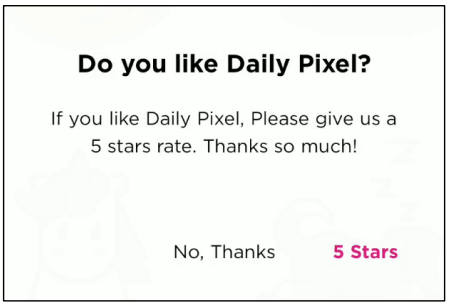

Figure 2. Popup in the Daily Pixel App. It presents `Nagging` and `False Hierarchy`, since it interrupts the user and favors the 5 Stars.

DPs and App Categories

We have also measured correlations between the number of DPs and app categories (we ran Welch ANOVA since our data failed parametric assumptions), finding that the News and Magazine category had fewer DPs when compared to other types: Music and Audio, Entertainment, Shopping, Social, and Communication ($F[7,232]= 3.390$, $p<0.05$). Besides this correlation, we did not find any other significant result.

Discussion

Through our manual classification, we found that the vast majority of trending applications use some form of malicious design to obtain certain responses from users.

Although the majority of the found DPs "simply" manipulates user interfaces, there are cases where more sensitive actions are involved. For instance, 31% of the apps contain `Privacy Zuckering`. The most common cause of PZ is privacy conditions accepted upon clicking some buttons or continuing with the registration process. We considered only cases where these labels were particularly small and hard to find. Often, this information was greyed or hidden by some other UI elements. In some other occasions, the app would activate by default the "send usage data" in the setting page. Also, particularly famous applications, such as Firefox [8] and Reddit [15], included this DP instance.

Regarding `Roach Motel` DPs, we only considered the following two cases: 'It is not possible to logout' and 'It is not possible to delete account.' This subcategory appeared in 41% of the apps; however, the majority of apps did not require an account to be used. Among the apps that allowed us to login, the vast majority did not include a 'delete account' feature within the app. Although we connected through our Gmail account whenever available, we believe that app developers should include at least a link to the Gmail account management page from their apps, since some users may be unaware of its existence or find it hard to reach it. Among the apps that do not include this feature, we found Spotify, Wish, Instagram [11], Amazon Photos [1], and many more.

Finally, some subcategories did not appear often. For instance, the `Price Comparison Prevention` was found 23 times in 240 apps. However, this DP is detectable only in Shopping applications (N=30), where it appeared in 77% of the cases.

ONLINE EXPERIMENT

Luguri and Strahilevitz [58] found that mild and aggressive DPs can have a significant impact on user behavior. While

users perceived aggressive DPs as particularly annoying, mild DPs had lower impacts on users' experience. We hypothesize that users may have developed a sort of DP-blindness to malicious design. To study this in detail, we carry out an online experiment in the form of an online survey which included videos of the apps usage.

The questionnaire received 589 answers from users with over 40 distinct nationalities and different background experiences. In the following, we report on the design of the study, its participants, and the final results, as well as discuss our analysis.

Design and Structure

The experiment, in the form of an online survey, started with a small introduction, where we stated that participants would be asked to watch videos to evaluate the overall user experience of apps. For each user that participated in the study, we donated two dollars to a charity of users choice (e.g., Wikimedia Foundation, Free Software Foundation).

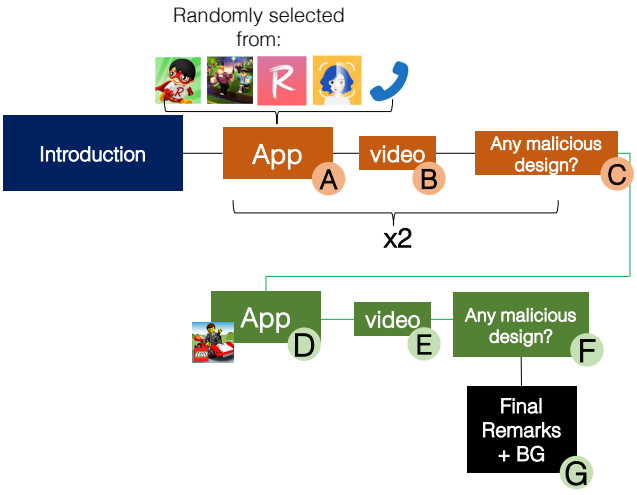

Figure 3. Structure of the online survey.

After this introduction, we followed the structure as represented in Figure 3. Each user evaluated three apps in this order: two containing malicious designs (randomly selected from: Tag with Ryan [21], Roblox [16], Romwe [17], Talkatone [22], Face Reading [5]) and one that did not (Lego Juniors).

For each application, users were first asked if they have ever used the app, only heard about it, or never come across it. If users used the app, we asked how often they used it in the last year (less than once a month, once a month, weekly, or daily), how they would rate it (one to five stars), and to briefly explain the reasons behind their rate (step *A* and *D* in Figure 3).

Once users completed this part, in the next page, they could watch a 30-second video and answer usability questions on the app (ease of understanding, ease of use) with a 5-point Likert scale (from 'Totally disagree' to 'Totally agree'). We also asked to rate the app again and briefly motivate the rating (step *B* and *E* in Figure 3).

Subsequently (step *C* and *F* of Figure 3), in a new page of the survey (with no possibility of going back) participants were asked if they could spot any malicious designs in the previous video. In the question, we defined a malicious design as: *"e.g., user interfaces crafted to trick the users in doing things they do not want to do, or try to manipulate the user in some way."*. To this questions, users could answer 'yes,' 'no,' or 'not sure.' If the answer was 'yes' or 'not sure,' we also asked to briefly explain the malicious design. Overall, we did not prime users on DPs and its definition, instead, we always used the more generic "malicious design" term. We made this decision to mitigate possible biases and to capture DP-blindness. Moreover, we asked this question after each video, and not after all apps, so that people would more easily remember its content and interactions.

Finally (step *G* in Figure 3), we showed users screenshots of the malicious designs for each app they evaluated. If they previously reported having identified some maliciousness, we asked if it was the same as just described. If instead, they did not spot it, we asked why ('I did not see it,' 'I did not found it to be malicious,' 'Other reasons'). Moreover, participants were asked to evaluate how annoying each malicious design was (from 'Very annoying' to 'Not annoying at all'). We concluded the experiment by asking background information.

Selected DPs and Apps

To study if users may spot DPs in user interfaces, we used five apps (from our dataset) with DPs and one without any DPs. Each user evaluated three apps, two containing a DP and one free from malicious designs. The first two apps would rotate among the aforementioned five. Instead, the last one (without DPs) was always the same.

With this design, we aimed to not only capture DP-blindness but also study potential learning effects. In fact, we hypothesized that after the first app evaluation, users would be more attentive on possible DPs.

We portrayed five instances of DPs for five different macro-classes [42] to study blindness depending on the DP category. We picked five subclasses: `Nagging`, `Intermediate Currency`, `False Hierarchy`, `Forced Action`, and `Sneak into Basket` (see Figure 4). We chose only a subset of the classes to limit the length of the survey. In addition, other classes of DPs were not suited for our study. For instance, it is hard to portray the impossibility of selecting product names in the `Price Comparison Prevention` class in a video. An additional challenge we faced was that most malicious designs do not comprise one class only. For instance, in our dataset, we could not find a `Sneak into Basket` without a `Preselection` UI (e.g., insurances preselected by default while buying products). However, it was not possible to find another suited occurrence of the `Sneaking` class. For this reason, we decided to keep the `Sneak into Basket` DP even if in conjunction with the `Preselection` one (see Figure 4, Romwe app). We a keep note of this factor in the analysis and discussion of the results.

Participants

The survey had 589 completed responses. Overall, 58% of the participants are women, 39% are men, 2.5% preferred not to disclose, and 0.5% chose to self-describe. The reported

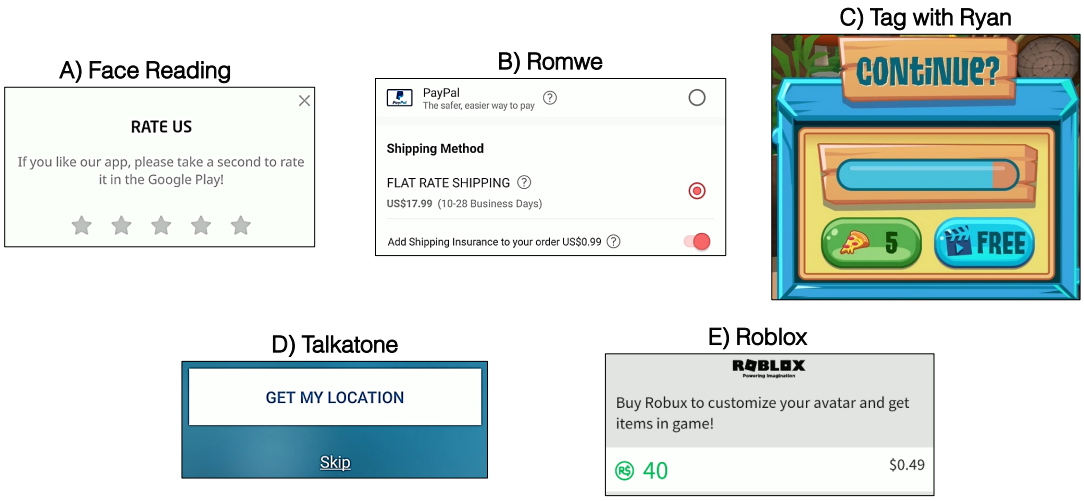

Figure 4. Screenshot of DPs of the five apps used in the survey. A) The Face Reading app contains a popup rating that interrupts the user (`Nagging`). B) The Romwe e-shopping app, adds an insurance by default when checking out (`Sneak into Basket`, and `Preselection`). C) The Tag with Ryan app, asks the user to watch an Ad to continue playing (`Forced Action`); D) The Talkatone app highlights the "Get my location" option either than the skip one (`False Hierarchy`). E) The Roblox app has many currencies (`Intermediate Currency`).

age ranged from 19 to 77 years old (avg.: 30.3, std.: 10.75). We had participants from 46 different countries (e.g., United States, UK, German, Brazilian, Italian, Swiss).

Most respondents reported to have attended secondary school (45%); in addition, several participants have a bachelor (31%), master (42%) or doctoral degree (23%). Among the college degrees, 28% were in Computer Science, Web Designer or Information Technology.

To understand how skilled participants are with modern technology, we asked how often they use mobile devices. The vast majority of our participants uses their smartphones every day (98%), while use tablets less frequently (47%).

Finally, most participants have no previous familiarity with any of the apps used in the experiment (88%); only a few have not used the application but have heard of it (9.7%).

Results

DP-Blindness

As we randomly assign two apps to each participant, a different amount of participants answered questions about different apps. For instance, 239 participants answered questions about Face Reading, while 248 answered questions about Romwe. As the third app, we assigned Lego (control) to all participants. Table 2 presents the total amount of participants per app, as well as how many users spotted a malicious design.

Regarding whether the participants could identify malicious designs, we gave the participants the following options: *yes*, *not sure*, *no*. We computed 1,767 answers (1,178 without the control) to this question as each participant answered this question to three videos. Overall, the majority of our users did not spot malicious designs in the app containing DPs (55%), some were unsure (20%), and the remaining found a malicious design in the app (25%). In the control task, 86% of users were able to recognize that the app had no DPs.

After showing the video, we asked participants to rate the apps and comment on the rating. In a new page of the survey, we asked participants if they identified any malicious designs in the videos they just watched. Therefore, in the rating page, participants were not yet primed about malicious designs. We analyzed their comments to check whether they could identify or suspect of a DP. Overall, out of 366 participants that answered the open question, only 7% somehow mentioned a DP in their answers.

At the end of the study, we showed to participants that answered 'yes' or 'not sure' to the previous question, the DPs of the apps. We asked these participants whether what they identified as the malicious design was the same as the DP shown. We did not ask this question to participants that identified malicious designs on Lego as it has no DP. Among the ones that have spotted a DP, only 24% of participants considered their answer correct, while the remaining 56% were unsure or considered their found malicious design different from the one we showed. Although most participants did not correctly identify the expected DPs, their approach towards the task showed an implicit distrusts in the application. For this reason, in the following, we will not exclude participants that did perceive a malicious design but not the expected DP.

DP-Blindness on Apps and Order

We performed a *Chi-square* test to verify the impact that different apps have on the number of detected and undetected DPs. We found that exists a correlation between apps and the number of DPs reported by the participants ($\chi(10) = 221.167$ and $p < 0.001$).

Among all apps, The Romwe application was the one that performed the worst (we ran pair-wise *Chi-square* and found significance against all other DP-apps, $p < 0.05$). The e-shopping app has the lowest percentage of DPs found (14%) when compared to all other applications containing DPs. In contrast, the remaining DP-apps performed similarly among each other.

Table 2. Participants that answered questions regarding each app. Partic. = Amount of participants; Malicious Design = Whether they identified a malicious design on the first app; Same as DP = Whether the malicious designs identified by the participants ("yes" or "not sure") are the same DPs we identified.

App	Partic.	Malicious Design			Same as DP		
		No	Not Sure	Yes	No	Some what	Yes
Face Reading	239	129	39	71	34	18	58
ROMWE	248	159	55	34	47	22	20
Roblox	227	103	66	58	36	41	47
Talkatone	246	135	44	67	34	24	53
Tag with Ryan	218	125	34	59	18	24	51
Lego	589	510	50	29	-	-	-
Total	**1,767**	**1,161**	**288**	**318**	**169**	**129**	**229**

As expected, the Lego task was the one in which respondents performed the best against all other applications ($p < 0.05$).

While analyzing the first app, the participant was not told to pay attention to the presence of a malicious design. However, while looking at the second app, they were more conscious of this objective. For this reason, we investigated how the order in which the apps were shown influences the finding of DPs. We performed a *Chi-square* test that confirmed that users are more attentive in searching DPs after the first app ($\chi(4) = 58.201$ and $p < 0.001$).

DP-Blindness and Demographic

The ability to find a malicious design might be influenced by previous knowledge they have about DPs. We checked this hypothesis performing a *Chi-square* test on the correlation between the participants' experience and their answers when asked if they noticed a DP in the app. The test confirmed that the association is statistically significant ($\chi(6) = 81.699$ and $p < 0.001$). We did not obtain any statistical differences in correlations among users age, employment status, or level of education.

DPs and Learnability

Finally, we asked participants questions regarding our survey. First, we asked them whether they learned something thanks to the experiment. In total, 49% of users answered yes, while 30% and 12% answered somewhat and no, respectively.

Subsequently, we asked whether participants would change their behavior after learning about DPs through the study. In total, 32% of our users answered they would extremely or moderately change how they download apps, 44% only slightly, and the remaining 24% answered that they would not change this aspect at all. Table 3 presents the participants' answers.

We used the *Chi-square* test to analyze whether there is a correlation between change of behavior (extremely, moderately, not at all, slightly or somewhat) and whether the participant identified a malicious design in the first app. We considered the following behaviors related to mobile applications: *download*; *use*; *rating*; and, *suggest to friends*. As output of our analysis, we found that there is no relation between a participant finding a malicious design and *downloading* the apps ($\chi(8) = 5.147$ and $p = 0.742$). On the other hand, the participants' *using*, *rating* and *suggesting to friends* are correlated to them finding malicious designs ($\chi(8) = 15.680$, $p < 0.05$; $\chi(8) = 19.073$, $p < 0.05$ and $\chi(8) = 21.038$, $p < 0.05$, respectively).

Table 3. Answers to what extent the participants think learning about Dark Patterns would change their behavior.

Affect	Mobile Apps			
	Download	Use	Rate	Suggest to Friends
Not at all	139	88	0	134
Slightly	139	88	63	73
Somewhat	118	134	117	128
Moderately	113	0	135	0
Extremely	73	100	111	112

Discussion

Noting the presence of DPs may lead to change how a user behaves with regards to a specific app. Confirming this claim, many participants declared their intentions of changing their behavior based on this increased awareness of DPs. Many survey participants reported that being aware of DPs might change how they will use and rate an app. The participants' remarks also confirm this factor, reporting that the presence of DPs led them to stop using an app and negatively rate it. However, some participants also stated that DPs are so widely spread and common among modern applications that they become part of the normal interaction flow when using apps. One this matter, when asked to give general feedback about our experiment, one of our participants stated:

> "As a remark on the watching an ad malicious design [i.e.: `Forced Action` DP in Tag with Ryan]: that may be so common already that we just do not consider it any more, and it allows us and really highlights the option to choose. Thus, I think it is good to highlight this issue, our attention for such designs are somewhat fading due to the exposure."

Some users also commented about the importance of such experiments, since they can sensitivities the population on the issue, as well as alert parents on the use their kids have on mobile applications. About this manner, one of our participants wrote:

> "... That kids are being targeted to advertising every X minutes or even seconds cannot be good for their brains and behaviour! This is a topic that must be investigated and discussed."

As previously mentioned, children are less aware of the difference between Ads and real content and are more easily manipulable than adults [34, 72, 92]. In this context, DPs can have a significant impact on the issue. For instance, the `Nagging` malicious design, the most common DP in our classification, often interrupt users to display Ads or features accessible only through payments. Given these factors, we believe that it is particularly relevant to continue the discussion on malicious designs and inform the users on the possibility of DP-blindness.

IMPLICATIONS OF THE STUDIES

The results of our studies reveal several aspects that can have practical implications for researchers, mobile app users, and

designers. In this section, we overview the outcomes as well as the impact of our findings for different stakeholders.

More empirical research is needed. One of the most surprising results of our study relates to the high pervasiveness of DPs in mobile apps; As a matter of fact, 95% of the considered apps include one or more forms of DPs. This opens questions concerning the causes behind their introduction as well as the motivations leading designers to add malicious designs in their mobile applications. We argue that more research on the harmfulness of each specific DP category should be conducted with the aim of informing users about potential privacy and security threats. Furthermore, it is still unknown whether there exist specific instances of a certain malicious designs category that are more problematic than others: in this sense, a characterization of DPs could be a useful means to address the lack of knowledge on their relevance and impact. While previous works [38, 95] have started working toward this direction, we argue that this is still an open research debate.

On the need of automated tools. A second critical aspect concerns the definition of instruments that can detect DPs and alert users of their presence. Indeed, the research community has focused on classifying DPs and understanding their relevance [62, 61, 33, 42], while only a little effort has been spent on devising automated methodologies that can identify their presence. Moreover, we notice a considerable lack of tools that can estimate the effect of DPs and alert users of the potential threats caused by using an app. At the same time, the definition of mechanisms to recommend designers how to remove malicious designs could further reduce their diffuseness as well as help meet the ethical expectations of mobile app users. With the contributions of this paper, we hope that future research could be done to recognize commonalities among DPs and, therefore, help the automatic recognition of malicious designs.

On the perspective of users. While our work tries to bring contributions that mainly target the academic community, we hope that it also provides information to users, making them aware of the high likelihood to be involved in some form of DPs while using mobile apps. Indeed, according to our findings, users are generally not aware of and cannot correctly recognize malicious designs. On the one hand, this reinforces the idea that more automated solutions would be required. On the other hand, our findings highlight that users should be more careful when using mobile apps. As such, we recommend users to pre-screen the downloaded apps in order to look for possible anomalous UI interactions/settings and make informed decisions on the personal data shared. At the same time, as pointed out by some of our study participants, the reported findings can be useful to decide on whether to prevent kids to use certain apps.

LIMITATIONS

Although we followed previous research in the field of DPs [42] and the considered taxonomy was particularly powerful for our tasks, certain adaptations were necessary. Some malicious designs were not directly stated by authors, and for this reason, we needed to interpret definitions of classes to associate DPs. Similarly, some borderline cases may be seen as malicious design or not depending on the viewer. Different UX experts may see issues that others might not find problematic. In our classification, we tackle this problem by first conducting the task in pairs and, secondly discussing our opinions with a third researcher during testing.

To further maintain coherency among different apps and limit the length of the classification, we restrict the number of cases to be considered as malicious designs. For this reason, many DPs have not been considered in our classification. For instance, we did not include all the features discussed by Moser et al. [62] in e-shopping applications. Despite the necessity to restrict the scope of the study, the quantity of found DPs remains particularly high.

We studied free apps of the Android platform and classified instances of DPs that appear in the first ten minutes usage of the application. Different DPs may be found outside of this scenario. Similarly, certain DPs may not appear in certain apps because they require the presence of functionalities that are not implemented. Moreover, paid apps may have fewer DPs and specific DPs might appear only later in the use of an application.

While we followed a structured list of tasks among all apps during the recording [67], differences among usage might exist. Although we recognize this factor as a possible limitation of our classification, this process better represents the normal behavior that users might have on the apps. This behavior strictly depends on the features that application offers and, for this reason, it can be only partially generalized.

As for the second study, we analyzed DP-blindness through an online experiment. By design, we had to ask questions on the `Sneak into Basket` DP while it co-occurred with a `Preselection UI` since there was no other individual instance of the former. This may have introduced some form of bias due to the mixed effects of the two dark patterns. Nevertheless, in the survey we explicitly asked participants to comment on the specific malicious UI detected (if any): from the analysis of the comments, participants who perceived the presence of a problem focused on Sneak into Basket, thus suggesting that they may not have been biased by the co-occurrence of the two patterns.

Watching a video and actively using an app are two different experiences. For this reason, it might have been more difficult for respondents to spot DPs than in real-life situations. Our choice was guided by the goal of studying the effect of DP-blindness on a wide number of participants. However, in the future, an in-lab user study might be conducted to compare our results with the active use of apps containing DPs. Similarly, some users stated that it was hard for them to capture the context and goals of the apps in thirty seconds. Deciding the right length of the videos was one of the challenges of this experiment. Too long videos would have invalidated DP-blindness results since users would need to remember too many UIs and interactions. Thirty seconds was the best compromise we could find for our experiment.

CONCLUSION AND FUTURE WORK

In this paper, we presented two studies we conducted to assess the prevalence of dark patterns in mobile applications and the user's perception of the problem. We first analyzed 240 apps belonging to 8 different categories on the Google Play Store and manually identified and classified dark patterns they included, finding that 95% of the analyzed apps contain one or more Dark Patterns. Afterwards, we conducted an online experiment involving 584 respondents who were asked to rate the UI of a subset of apps considered in the first study. The outcome highlighted that most of the times users could not perceive the presence of malicious designs. These results lead to several implications and challenges, e.g., how to increase the user's awareness of dark patterns: these represent the main item of our future research agenda, which targets the definition of methods to identify and characterize dark patterns.

REFERENCES

[1] 2019. Amazon Photos. (2019). https://www.amazon.com/Amazon-Photos/b?ie=UTF8&node=13234696011

[2] 2019. Barcode Scanner Google Play page. (2019). https://play.google.com/store/apps/details?id=com.lego.bricksmore

[3] 2019. Call Free - Free Call, Google Play page. (2019). https://play.google.com/store/apps/details?id=call.free.international.phone.call

[4] 2019. Experiment Dataset. (2019). https://figshare.com/s/048c984854a59429d0f0

[5] 2019a. Face Reading Google Play page. (2019). https://play.google.com/store/apps/details?id=app.facereading.signs

[6] 2019b. FaceApp Google Play page. (2019). https://play.google.com/store/apps/details?id=io.faceapp&hl=en

[7] 2019. Facebook. (2019). https://facebook.com

[8] 2019. Firefox. (2019). https://www.mozilla.org/en-US/firefox/new/

[9] 2019. Gmail. (2019). https://gmail.com

[10] 2019. Google Play Store. (2019). https://play.google.com/store

[11] 2019. Instagram. (2019). https://www.instagram.com

[12] 2019. Lego Juniors Google Play page. (2019). https://play.google.com/store/apps/details?id=com.lego.bricksmore

[13] 2019. Netflix. (2019). https://netflix.com

[14] 2019. One Plus 5. (2019). https://oneplus.com/5

[15] 2019. Reddit. (2019). https://www.reddit.com/

[16] 2019. Roblox Google Play page. (2019). https://play.google.com/store/apps/details?id=com.roblox.client

[17] 2019. Romwe. (2019). https://play.google.com/store/apps/details?id=com.romwe

[18] 2019. SensorTower. (2019). https://sensortower.com

[19] 2019. Snapseed. (2019). https://snapseed.online

[20] 2019. Spotify. (2019). https://spotify.com

[21] 2019. Tag With Ryan Google Play Page. (2019). https://www.amazon.com/Amazon-Photos/b?ie=UTF8&node=13234696011

[22] 2019. Talkatone Google Play page. (2019). https://play.google.com/store/apps/details?id=com.talkatone.android

[23] 2019. Twitter. (2019). https://twitter.com

[24] 2019. Wish. (2019). https://www.wish.com

[25] Monica Anderson and Jingjing Jiang. 2018. *Teens, Social Media & Technology Overview 2018*. Technical Report. Pew Research Center, https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018//.

[26] Carmelo Ardito, Paolo Buono, Danilo Caivano, Maria Francesca Costabile, and Rosa Lanzilotti. 2014. Investigating and promoting UX practice in industry: An experimental study. *International Journal of Human-Computer Studies* 72, 6 (2014), 542–551.

[27] Dan Ariely and Gregory S Berns. 2010. Neuromarketing: the hope and hype of neuroimaging in business. *Nature reviews neuroscience* 11, 4 (2010), 284.

[28] Jeffrey Bardzell and Shaowen Bardzell. 2013. What is "Critical" About Critical Design?. In *Proceedings of the Conference on Human Factors in Computing Systems*. 3297–3306.

[29] Javier A Bargas-Avila and Kasper Hornbæk. 2011. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the Conference on Human Factors in Computing Systems*. 2689–2698.

[30] Brian Bowman, Niklas Elmqvist, and TJ Jankun-Kelly. 2012. Toward visualization for games: Theory, design space, and patterns. *Transactions on Visualization and Computer Graphics* 18, 11 (2012), 1956–1968.

[31] Harry Brignull. 2019. *Dark Patterns*. Technical Report. Harry Brignull Dark Patterns website, https://www.darkpatterns.org/.

[32] Will Browne and Mike Swarbrick Jones. 2017. What works in e-commerce-a meta-analysis of 6700 online experiments. *Qubit Digital Ltd* 21 (2017).

[33] Gregory Conti and Edward Sobiesk. 2010. Malicious Interface Design: Exploiting the User. In *Proceedings of the International Conference on World Wide Web*. 271–280.

[34] Eveline A Crone and Elly A Konijn. 2018. Media use and brain development during adolescence. *Nature communications* 9, 1 (2018), 588.

[35] Nicholas S. Dalton, Emily Collins, and Paul Marshall. 2015. Display Blindness?: Looking Again at the Visibility of Situated Displays Using Eye-tracking. In *Proceedings of the Conference on Human Factors in Computing Systems*.

[36] Benedetto De Martino, Dharshan Kumaran, Ben Seymour, and Raymond J Dolan. 2006. Frames, biases, and rational decision-making in the human brain. *Science* 313, 5787 (2006), 684–687.

[37] Ward Edwards. 1954. The theory of decision making. *Psychological bulletin* 51, 4 (1954), 380.

[38] Madison Fansher, Shruthi Sai Chivukula, and Colin M Gray. 2018. # darkpatterns: UX Practitioner Conversations About Ethical Design. In *Extended Abstracts of the Conference on Human Factors in Computing Systems*. LBW082.

[39] Dan Fitton and Janet C Read. 2019. Creating a Framework to Support the Critical Consideration of Dark Design Aspects in Free-to-Play Apps. In *Proceedings of the International Conference on Interaction Design and Children*. 407–418.

[40] Batya Friedman, Peter H Kahn, and Alan Borning. 2008. Value sensitive design and information systems. *The handbook of information and computer ethics* (2008), 69–101.

[41] Joseph H. Goldberg and Anna M. Wichansky. 2003. Chapter 23 - Eye Tracking in Usability Evaluation: A Practitioner's Guide. In *The Mind's Eye*, J. Hyönä, R. Radach, and H. Deubel (Eds.). North-Holland, 493 – 516.

[42] 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 Conference on Human Factors in Computing Systems*.

[43] Saul Greenberg, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. 2014. Dark Patterns in Proxemic Interactions: A Critical Perspective. In *Proceedings of the Conference on Designing Interactive Systems*. 523–532.

[44] Barbara Grimpe, Mark Hartswood, and Marina Jirotka. 2014. Towards a closer dialogue between policy and practice: responsible design in HCI. In *Proceedings of the Conference on Human Factors in Computing Systems*. 2965–2974.

[45] Shipra Gupta. 2013. The psychological effects of perceived scarcity on consumers buying behavior. (2013).

[46] Rachel Harrison, Derek Flood, and David Duce. 2013. Usability of mobile applications: literature review and rationale for a new usability model. *Journal of Interaction Science* 1, 1 (2013), 1.

[47] David J Heeger and David Ress. 2002. What does fMRI tell us about neuronal activity? *Nature Reviews Neuroscience* 3, 2 (2002), 142.

[48] Yu-Chen Hsieh and Kuo-Hsiang Chen. 2011. How different information types affect viewers attention on internet advertising. *Computers in Human Behavior* 27, 2 (2011), 935 – 945.

[49] Ashley Karr. 2014. *Ethical Design*. Technical Report. Interactions ACM, `https://interactions.acm.org/blog/view/ethical-design`.

[50] Shinjiro Kawato and Nobuji Tetsutani. 2004. Detection and tracking of eyes for gaze-camera control. *Image and Vision Computing* 22, 12 (2004), 1031 – 1038.

[51] Ralph Kimball and B Verplank E Harslem. 1982. Designing the Star user interface. *Byte* 7, 1982 (1982), 242–282.

[52] Chris L Kleinke. 1986. Gaze and eye contact: a research review. *Psychological bulletin* 100, 1 (1986), 78.

[53] Mike Kuniavsky. 2010. *Smart things: ubiquitous computing user experience design*. Elsevier.

[54] Cherie Lacey and Catherine Caudwell. 2019. Cuteness as a Dark Pattern in Home Robots. In *Proceedings of the International Conference on Human-Robot Interaction*. 374–381.

[55] Marta Kristín Lárusdóttir, Åsa Cajander, and Jan Gulliksen. 2012. The big picture of UX is missing in Scrum projects. In *Proceedings of the International Workshop on the Interplay between User Experience (UX) Evaluation and System Development (I-UxSED)*. 49–54.

[56] Henry Latham. 2018. *Ethical design is a dangerous term*. Technical Report. Medium, UX Collective, `https://uxdesign.cc/ethical-design-is-a-dangerous-term-b314a5e385f4`.

[57] R. Alexander Lattimore. 1967. *The Odyssey of Homer*. Harper & Row.

[58] Jamie Luguri and Lior Strahilevitz. 2019. Shining a Light on Dark Patterns. *U of Chicago, Public Law Working Paper* 719 (2019).

[59] Kristiyan Lukanov, Horia A. Maior, and Max L. Wilson. 2016. Using fNIRS in Usability Testing: Understanding the Effect of Web Form Layout on Mental Workload. In *Proceedings of the Conference on Human Factors in Computing Systems*. 4011–4016.

[60] Petri Mannonen, Maiju Aikala, Hanna Koskinen, and Paula Savioja. 2014. Uncovering the user experience with critical experience interviews. In *Proceedings of the Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*. 452–455.

[61] Arunesh Mathur, Gunes Acar, Michael Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *arXiv preprint arXiv:1907.07032* (2019).

[62] Carol Moser, Sarita Y. Schoenebeck, and Paul Resnick. 2019. Impulse Buying: Design Practices and Consumer Needs. In *Proceedings of the Conference on Human Factors in Computing Systems*. 242:1–242:15.

[63] Emily R. Murphy, Judy Illes, and Peter B. Reiner. 2008. Neuroethics of neuromarketing. *Journal of Consumer Behaviour* 7, 4?5 (2008), 293–302.

[64] M. Nebeling, M. Speicher, and M. Norrie. 2013. W3Touch: Metrics-based Web Page Adaptation for Touch. In *Proceedings of the Conference on Human Factors in Computing Systems*.

[65] Robert Nickerson, Mark Austreich, and Jamie Eng. 2014. Mobile technology and smartphone apps: A Diffusion of innovations analysis. (2014).

[66] Jakob Nielsen. 1994a. Enhancing the Explanatory Power of Usability Heuristics. In *Proceedings of the Conference on Human Factors in Computing Systems*. 152–158.

[67] Jakob Nielsen. 1994b. Usability Inspection Methods. In *Proceedings of the Conference Companion on Human Factors in Computing Systems*. 413–414.

[68] Jakob Nielsen. 1999. *Designing Web Usability: The Practice of Simplicity*. New Riders Publishing.

[69] Jakob Nielsen and Rolf Molich. 1990. Heuristic Evaluation of User Interfaces. In *Proceedings of the Conference on Human Factors in Computing Systems*. 249–256.

[70] C. Nodder. 2013. *Evil by design: Interaction design to lead us into temptation*. John Wiley & Sons.

[71] Harri Oinas-Kukkonen and Marja Harjumaa. 2009. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems* 24, 1 (2009), 28.

[72] Gwenn Schurgin O'Keeffe, Kathleen Clarke-Pearson, and others. 2011. The impact of social media on children, adolescents, and families. *Pediatrics* 127, 4 (2011), 800–804.

[73] Justin W. Owens, Barbara S. Chaparro, and Evan M. Palmer. 2011. Text Advertising Blindness: The New Banner Blindness? *J. Usability Studies* 6, 3 (2011), 12:172–12:197.

[74] Justin W. Owens, Evan M. Palmer, and Barbara S. Chaparro. 2014. The Pervasiveness of Text Advertising Blindness. *J. Usability Studies* 9, 2 (2014), 51–69.

[75] Jay Patel, Gil Gershoni, Sanjay Krishnan, Matti Nelimarkka, Brandie Nonnecke, and Ken Goldberg. 2015. A Case Study in Mobile-Optimized vs. Responsive Web Application Design. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. 576–581.

[76] Christopher A Paul. 2016. Values at play in digital games. *New Media & Society* (2016).

[77] Sarah Perez. 2017. *Google Play now considers user engagement, not just downloads, in ranking games*. Technical Report. Techcrunch, `https://techcrunch.com/2017/02/28/google-play-now-considers-user-engagement-not-just-downloads-in-ranking-games/`.

[78] Jenny Preece, Yvonne Rogers, Helen Sharp, David Benyon, Simon Holland, and Tom Carey. 1994. *Human-Computer Interaction*. Addison-Wesley Longman Ltd.

[79] S Rajeshkumar, Ridha Omar, and Murni Mahmud. 2013. Taxonomies of user experience (UX) evaluation methods. In *Proceedings of the International Conference on Research and Innovation in Information Systems*. 533–538.

[80] J. Redström. 2006. Persuasive Design: Fringes and Foundations. In *Persuasive Technology*.

[81] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph 'Jofish' Kaye. 2005. Reflective Design. In *CC*.

[82] K. Shilton, J. A. Koepfler, and K. R. Fleischmann. 2014. How to See Values in Social Computing: Methods for Studying Values Dimensions. In *Proceedings of the Conference on Computer-Supported Cooperative Work and Social Computing*.

[83] Steven J. Stanton, Walter Sinnott-Armstrong, and Scott A. Huettel. 2017. Neuromarketing: Ethical Implications of its Use and Potential Misuse. *Journal of Business Ethics* 144, 4 (2017), 799–811.

[84] Statcounter. 2019. *Mobile Operating System Market Share Worldwide*. Technical Report. Statcounter, `https://gs.statcounter.com/os-market-share/mobile/worldwide`.

[85] Statista. 2018. *Market reach of the most popular Android app categories*. Technical Report. Statista, `https://www.statista.com/statistics/200855/favourite-smartphone-app-categories-by-share-of-smartphone-users/`.

[86] Space Technologies. 2017. *How Do Free Apps Make Money?* Technical Report. Space Technologies, `https://www.spaceotechnologies.com/how-do-free-apps-make-money/`.

[87] Sabrina M Tom, Craig R Fox, Christopher Trepel, and Russell A Poldrack. 2007. The neural basis of loss aversion in decision-making under risk. *Science* 315, 5811 (2007), 515–518.

[88] Janet Patton Tracy and Michael J Albers. 2006. Measuring cognitive load to test the usability of web sites. In *Annual Conference-society for technical communication*, Vol. 53. 256.

[89] Nynke Tromp, Paul Hekkert, and Peter-Paul Verbeek. 2011. Design for socially responsible behavior: a classification of influence based on intended user experience. *Design issues* 27, 3 (2011), 3–19.

[90] Yesim Isil Ulman, Tuna Cakar, and Gokcen Yildiz. 2015. Ethical Issues in Neuromarketing: "I Consume, Therefore I am!". *Science and Engineering Ethics* 21, 5 (2015), 1271–1284.

[91] F. Ungureanu, R. G. Lupu, A. Cadar, and A. Prodan. 2017. Neuromarketing and visual attention study using eye tracking techniques. In *2017 21st International Conference on System Theory, Control and Computing*. 553–557.

[92] Patti M Valkenburg and Jessica Taylor Piotrowski. 2017. *Plugged in: How media attract and affect youth*. Yale University Press.

[93] Giovanni Vecchiato, Laura Astolfi, Fabrizio De Vico Fallani, Jlenia Toppi, Fabio Aloise, Francesco Bez, Daming Wei, Wanzeng Kong, Jounging Dai, Febo Cincotti, Donatella Mattia, and Fabio Babiloni. 2011. On the Use of EEG or MEG Brain Imaging Tools in Neuromarketing Research. *Journal Computational Intelligence and Neuroscience*, Article 3 (2011), 3:1–3:12 pages.

[94] R. Mark Wilson, Jeannie Gaines, and Ronald Paul Hill. 2008. Neuromarketing and Consumer Free Will. *Journal of Consumer Affairs* 42, 3 (2008), 389–410.

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

[96] Dongsong Zhang and Boonlit Adipat. 2005. Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications. *International Journal of Human Computer Interaction* 18, 3 (2005), 293–308.