

Shedding light on assessing Dark Patterns: Introducing the System Darkness Scale (SDS)

Christof van Nimwegen
Information and Computing Sciences
Interaction Technology, Utrecht University
Princetonplein 5, 3584CC
Utrecht, The Netherlands
c.vannimwegen@uu.nl

Kristi Bergman
Kaartje2go
Emmawijk 11
8011 CM, Zwolle
The Netherlands
kristi@kaartje2go.nl

Almila Akdag
Information and Computing Sciences
Interaction Technology, Utrecht University
Princetonplein 5, 3584CC
Utrecht University, the Netherlands
a.a.akdag@uu.nl

Dark Patterns are elements in interfaces designed to misdirect, confuse, and lure users into unintended, involuntary actions. These are not just “sloppy” or “inelegant” designs without ill intent but are rather carefully crafted with an understanding of human psychology. Dark patterns are omnipresent as part of web and game-interfaces and highly effective. Hence, there is agreement that awareness and a better understanding is needed. The current study focuses on dark patterns from a user’s perspective in order to develop the ‘System Darkness Scale’ (SDS). The SDS is a set of questionnaire items which can be used to evaluate the darkness of a system and assign a score to it. Just as the SUS proved to be a quick and reliable tool to measure usability, the SDS aims to act as a validated tool to identify in how far a system or service has incorporated “dark mechanisms”.

Keywords: Dark Patterns, evaluation tool, SUS, user perspective, (un)ethical UI design

1. INTRODUCTION

“Dark Patterns” are defined as (parts of) interfaces that are designed with malicious intentions to deceive users into performing actions they did not intend to do (Brignull, 2015). Such design instances are often crafted to confuse users, make it difficult to express their actual preferences, or manipulate users into certain actions (Luguri & Strahilevitz, 2019). An example we all know is when in a web store, extra items are put into the shopping basket.



Fig. 1. Mobile dark pattern: illusion of a hair on the screen

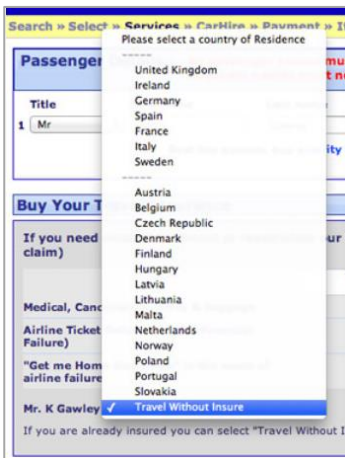


Fig. 2. Air travel web site making it purposely hard to NOT to buy travel insurance. This is hidden in the country list under “T.”

An extreme example is visible in figure 2 in which while booking a journey, the website attempts to lure the user into buying Travel Insurance. In case one does not want this insurance, one has to find the option “Don’t insure me” in the country pulldown list under “T” (Travel without insurance). An even more extreme example is a strand of hair being Photo-shopped in an image on a website, see figure 1. When seen on a mobile device this leads the user to think it is real. When the user then tries to swipe the hair off the surface, the user is actually clicking on a link and enters a web store involuntarily. As these designs entail a solid understanding of human psychology, they are very effective and widely used (Brignull, 2015) (Mathur et. Al, 2019).

However, dark patterns remain an understudied topic in HCI literature. Especially the understanding how these deceptive design strategies are perceived by the users is an open question. In this study, we address this gap by focusing on the user perspective and designing a dark patterns assessment system, i.e., a System Darkness Scale (SDS). The approach is inspired by the widely used System Usability Scale “SUS” (Brooke, 1996) to evaluate the usability of a system. By introducing SDS, we offer a tool that will be useful in identifying how ‘dark’ a system is perceived to be by the users. Furthermore, SDS as a tool will provide a reliable measurement system to compare the darkness of different systems. Dark patterns are deployed with

malicious intent, but how strongly different dark patterns affect users may vary. While one pattern may be very severe, leading to great harm, another dark pattern may be perceived as 'acceptable'. Different combinations of dark patterns may lead to different 'severity' perceptions. As such, this study creates measure of the 'severity' or 'darkness' of a service as a whole which will enable researchers as well as designers to quickly and easily determine the 'darkness' of a system, service, or product. The aim is to build the SDS in a comparable manner to the System Usability Scale (SUS). However, whereas the SUS provides an insight into how usable a certain service is, the SDS will provide an insight into how dark a certain service is. Research on dark patterns generated an awareness on malicious design strategies and led to a discourse among HCI researchers addressing the (un)ethical nature of designing for deception. An important concern is the effects of researching and publishing results on the techniques used for deceptive design, as these could provide guidance for malicious stakeholders. However, researchers within the field generally believe that this approach is helpful and necessary. First, analysis and documentation allow for better understanding of the underlying concepts and mechanisms. Second, research on dark patterns fosters awareness and makes it easier to identify malicious patterns in the wild. Lastly, research can act as a starting point for countermeasures and regulations/legislation (Bösch, Erb, Kargl, Kopp & Pfattheicher, 2016). Regarding the latter, in 2021 the Norwegian Consumer Council filed a legal complaint against Amazon who used the Roach Motel Pattern to manipulate users. Since then, 16 other consumer organizations in Europe and the US are acting against Amazon. As such, the community is working towards more 'Ethical UI Design' (Di Geronimo, Braz, Fregnan, Palomba & Bacchelli) by creating a better understanding of dark patterns, by increasing awareness of the deceiving nature of dark patterns, and by proposing alternative design solutions. We believe our study contributes to these efforts, as a better understanding of users' perception about dark patterns will provide us with in-depth answers about their effects. In this work, we first review related literature in dark patterns. We then present our method and explain how we build a System Darkness Scale. In findings we elaborate on the elimination techniques we applied for the finalization of the SDS and describe how the SDS is applied. We conclude with a discussion and directions for future research.

2. RELATED WORK

We start this section by elaborating on the differences between persuasive technologies and deceptive design strategies. Then we introduce the taxonomies of dark patterns and report the recent

findings on the wide usage and effects of dark patterns. Finally, we discuss the importance of the recently coined term "Dark Pattern Blindness".

2.1 From Persuasion to Deception

User Interface Design refers to the design of the UI - it is about programming the look of things (Berezhnoi, 2019). Good user interface design helps users accomplish their desired goals easily and effortlessly. It aids users in reaching their goals by being self-explanatory, presenting information in an understandable fashion, and allowing for easy navigation (Sommerer & Mignonneau, 2008).

Usability is a key issue in designing effective user interfaces (Schaffer, 2009). However, applying usability techniques to enable users to perform actions efficiently, does not mean that this action will actually take place. In other words: just because users *can* do something does not guarantee that they *will*. In order to achieve the latter, users must be motivated and persuaded (Gubaidulin, 2016). As such, design is inherently a persuasive act (Nodder, 2013). Persuasive Technology aims to influence the behaviour of people by the use of design. Fogg introduced the term persuasive technology in 2003, defining it as "a computing system, device, or application intentionally designed to change a person's attitude or behaviour in a predetermined way" (Fogg, 2002). The shaping of behaviour is accomplished with the help of behavioural insights from psychology. By applying psychological insights to interface design designers can communicate information to users more precisely, aid users in decisions, nudge users toward goal completion, assist them in developing skills, and even help end or begin new habits (UX Booth, 2018).

One of the important concepts one often encounters in influencing behaviour is "nudging". Digital nudging refers to "the use of user interface design elements to guide people's behaviour in a digital environment, without restricting the individual's freedom of choice" (Meske & Potthoff, 2017). The latter part is important, as the main intention of nudging is to increase people's long-run welfare, by helping them make better choices, without forcing outcomes (Thaler & Sunstein, 2009).

While persuasive technology and nudging are often praised, there are substantial ethical considerations (Gray et al., 2018) and there is a fine line between persuasion and deception (Lång & Pudane, 2019). To distinguish the two, we can look at the intentions of the designers. In persuasive and nudging practices designers aim to encourage users to freely explore content and take actions. These practices are generally designed to guide user's behaviour with the goal to make users better off. To reach this goal, persuasive practices use techniques in which users are put at the centre of attention (Lång & Pudane, 2019) (Sunstein, 2019). In deceptive

practices on the other hand, designers aim to either trick users into taking actions, or prevent them from performing them. The persuasive techniques are no longer deployed with the user in mind, but are rather business centric (Gray, Kou, Battles, Hoggatt & Toombs, 2018).

Deception or a negative user experience can occur unintentionally due to a lack of technical skills, inexperience or little knowledge of user needs of the designer (Greenberg, Boring, Vermeulen & Dostal, 2014). Such a design solution is often called an anti-pattern. When an anti-pattern is discovered, it is often documented as 'known bad practice', so use of the design solution can be prevented in future UI design.

2.2 Dark Pattern Taxonomies

When deception occurs on purpose instead of unintentionally, we deal with so called dark patterns. The term "dark patterns" was coined by Harry Brignull in 2010. The first taxonomy of the patterns did also come from him. In this section, we provide both Brignull's and Gray's taxonomies, as these are widely used in the literature. For this study, we also used them as guidelines.

2.2.1 Brignull's Taxonomy (2015) has collected a set of artefacts containing dark patterns from blogs, websites, and social media and bundled into an overview, called the "Hall of Shame", on his website www.darkpatterns.org (Gunnarsson, 2020). The set of artifacts guided Brignull's dark patterns taxonomy, which consisted of 12 types of dark patterns. The patterns identified were: "Bait and Switch", "Disguised Ad", "Forced Continuity", "Friend Spam", "Hidden Costs", "Misdirection", "Price Comparison Prevention", "Privacy Zuckering", "Roach Motel", "Sneak into Basket", and "Trick Questions".

2.2.2 Gray's Taxonomy. Gray et al. (2018) built upon the existing taxonomy by Brignull which mixed context, strategy, and outcome. Gray's taxonomy is based on strategic motivators behind dark patterns. This categorisation is sharper, more general and more suitable for comparison among patterns (Di Geronimo et al., 2020) and as such, Gray's Taxonomy has become a standard of reference. Gray (2018) identified five primary dark pattern categories that serve as strategic motivators for designers:

(1) nagging, (2) obstruction, (3) sneaking, (4) interface interference, and (5) forced action.

Each of the five dark pattern categories includes multiple dark pattern strategies, among which the original dark pattern types as found by Brignull (2015), and some extra ones. Table 1 provides a summary of all dark pattern categories, strategies, and some of their instances.

2.3 Effectiveness and Effects of Dark Patterns

Despite their 'darkness' and questionable ethical nature, dark patterns are widespread. Mathur et al. (2019) used a web crawler to identify dark patterns on the 11K most popular shopping websites worldwide. They discovered 1818 instances of dark patterns, which were present on 1254 of them (11.1%). The more popular shopping websites were more likely to feature the dark patterns. Di Geronimo et al. (2020) analysed the prevalence of dark patterns in 240 applications and found that 95% included dark patterns in their interfaces. Overall, 1787 dark patterns were found among all apps.

Table 1. Dark pattern categories by Gray et al. (2018), underlying strategies and description.

Category	Strategy	Description
Nagging		repeated interruption of the users' desired task.
Obstruction	roach motel	situation is easy to get into, but difficult to get out of.
	price comparison prevention	makes comparing prices between products difficult.
	intermediate currency	disconnects users from the real value of money
Sneaking	forced continuity	automatically charge user after service expires
	hidden costs	shows certain price, but adds additional costs
	sneak into basket	adds additional items to cart without consent.
	bait & switch	certain action has different outcome than expected.
Interface Interference	hidden information	relevant options/actions not immediately accessible
	preselection	option selected by default prior to user interaction
	toying with emotion	uses language / style to evoke a particular emotion
	false hierarchy	gives visual precedence to one option over others.
	disguised ads	ads that blend with the medium of information.
	trick questions	question that appears to be one thing, but is another
Forced Action	social pyramid	
	privacy zuckering	
	gamification	service only earned through repeated use of service

Almost 10% of the apps included 0-2 dark patterns, 35% of the apps contained 3-6 dark patterns, while the remaining 49% included >7 dark patterns. The high prevalence of dark patterns is not surprising as numerous studies have found that they are very effective. Dark patterns perform well in multivariate tests and A/B testing (Brignull, 2011). As such, applying them will likely result in more sales, higher revenues, and the obtainment of more (personal) user data in comparison to an interface design that

does not intentionally trick the human mind. This makes dark patterns a valuable and highly effective asset in trying to reach business-oriented goals (Mayer, 2019). Luguri and Strahilevitz (2019) offered striking empirical support for the proposition that dark patterns are effective in bending consumers' will. The study assessed the effectiveness of dark patterns for accepting a certain security program by three levels of dark pattern manipulation: (1) no dark patterns, (2) mild dark patterns, and (3) aggressive dark patterns. In the no 'dark patterns' condition, only 11.3% of the participants accepted the program. When mild dark pattern tactics were deployed more than double accepted: 25.8% of participants became victim. When exposed to aggressive dark patterns, the effectiveness went up further, with 41.9% of the sample accepting the security program. In a study by Utz et al. (2019) different consent pop-up designs were created to investigate the effect of different design properties on acceptance rates. Among them were two dark patterns: Preselection and False Hierarchy. When given the options 'accept' and 'decline', users are more likely to share personal information when the 'accept' button is given visual precedence over the 'decline' button (as opposed to when the buttons have equal precedence). Nouwens et al. (2020) evaluated effects of different designs of consent banners on users' consent choices. The probability of a user accepting to a privacy notice increases with 22 percentage points when the 'reject all' button is removed from the first page of a consent banner (and is hidden on a second page), while the 'accept all' button remains present. Ever since the effectiveness of dark patterns are better understood, the question arose whether users actually recognize the use of dark patterns, or whether they are unaware of these malicious strategies. Di Geronimo et al. (2020) carried out an online experiment in which participants were asked to indicate whether they spotted malicious design in the user interfaces (UI) of several applications. Results of the experiment showed that most of the participants did not spot malicious designs in the apps containing dark patterns (55%). They explained their results by a concept known as "Dark Pattern Blindness": dark patterns are so widely spread and common among modern applications, that they become part of the normal interaction flow. As users are frequently being exposed to dark patterns, their attention for such designs is somewhat fading.

The works outlined above encompass a variety of results on the effectiveness, wide use, and effects of dark patterns. From the number of dark patterns spotted and categorized, to the pervasiveness of them, the dark pattern research paints a dark picture. The dark pattern blindness might be the most worry some effect of these deceptive strategies, as the researchers observe how users learn to live with dark patterns. Our work also aims

to understand how users perceive these patterns as well as to shed light to which strategies are still "visible" to the users and which patterns have already formed habitual behaviour in the users.

3. METHOD

In this section, we elaborate upon the experiment that we performed for generating the System Darkness Scale (SDS). The aim of the SDS is to capture the 'darkness' of a certain system as perceived by the user. In some way, the SDS can be compared to the System Usability Scale (SUS), which is commonly used to capture the 'usability' of a certain system as perceived by the user. Hence, we adopt the methodology design of the SUS. In order to assure the SDS renders the darkness of a system in a reliable manner, questionnaire items must be carefully selected. In what follows, we describe the process of generating a set of potential questionnaire items and evaluating their suitability for inclusion in the final SDS by using them in a user experiment.

3.1 Experimental Material: A List of Potential Questionnaire Items

To be able to select a set of suitable questionnaire items for the SDS, a bigger pool of potential questionnaire items needed to be generated first. It is desirable that the items within this pool cover a variety of aspects related to 'darkness', for example 'trickery', 'deception', 'evil intentions' and 'a business-centric approach'. After deriving an extensive list of such aspects, 25 potential questionnaire items were formulated, see table 2. About half of the statements is formulated in a "dark" manner, the other half is formulated in a "bright" manner. This was done on purpose, as it would prevent from response biases in the evaluation part of the study.

3.2 Experimental Design: The Extreme Ends of the Spectrum

In order to form an idea of which potential questionnaire items are best suited for the SDS, we need to test the items on their ability to capture the level of 'darkness' of a system. An item is deemed suitable if it can show the attitudes of people towards (malicious) interface design on the whole spectrum; whenever it is able to distinguish the "really bright" systems (no dark patterns) from the "really dark" ones (many severe dark patterns). To identify which items met these criteria, we needed to capture the attitudes of people for both these opposite systems. It was not possible to find two similar systems that are each other's exact opposite.

Table 2. List of Potential Questionnaire Items.

Item	Statement
01	The system tricked me into performing certain actions that I did not intend to do
02	I was repeatedly interrupted by the system without being able to stop the interruption
03	I could respond to the system as I desired at all times
04	The system obstructed me in performing certain actions
05	I could perform every action that I wanted to perform
06	The system performed certain actions I was not aware of.
07	Critical / relevant information for me as a user was readily available at all times.
08	The system gave specific actions or choices (visual) precedence over others
09	The system required me to perform specific (unwanted) actions to proceed to the next step in a process.
10	The system guided my behaviour in a way that benefited the designer of the system (e.g., online company)
11	I think that within this system, the user is put at the centre of attention
12	I think that this system brings harm to its users.
13	Using the system, I felt that I had control over my own actions and choices.
14	The system performed actions without my consent.
15	The system pushed me into spending more money than I originally anticipated.
16	I felt I had control over sharing my personal information
17	I felt the system used my emotions to trick me into performing certain actions.
18	The system caused me to spend unnecessary time, energy and attention to perform an action or choice.
19	The actions I performed using the system always resulted in the expected outcomes.
20	The wording used in the system was explicit and clear
21	The system took good care of my individual welfare.
22	I felt deceived/misled by the system
23	I felt the designer of the system had the user's best interests in mind
24	I felt like the system helped me in making better choices, that would benefit me as a user in the end.
25	The possible range of actions I could perform within the system was clear to me at all times.

Hence, we designed these two systems ourselves, in the format of an e-commerce web store. This also gave us the advantage of having more control over what a participant saw when interacting with the system. Mathur et al. (2019) identified the following Dark Pattern strategies to be often used in web stores: 'Price Comparison Prevention', 'Hidden Costs', 'Sneak into Basket', 'Hidden information', 'Preselection', 'Toying with Emotions', 'Trick Questions', 'False Hierarchy', and 'Disguised Ads'. These strategies were implemented in the "really dark" version of our web store, see figure 4. The "really bright" website (figure 3) had no dark patterns at all.

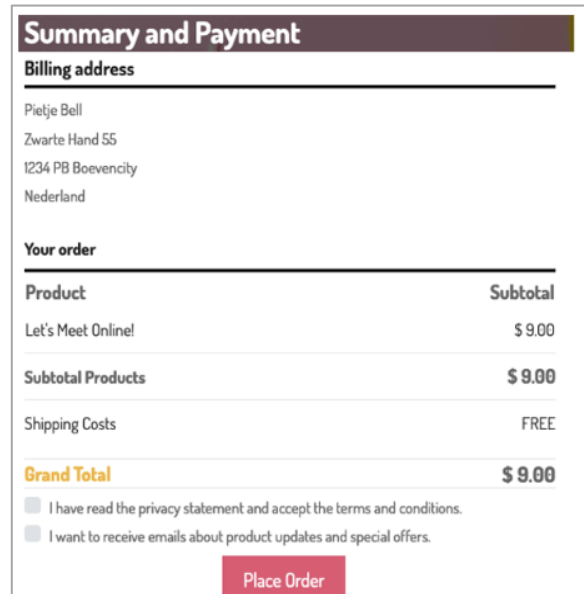


Figure 3: Really bright system: no dark patterns

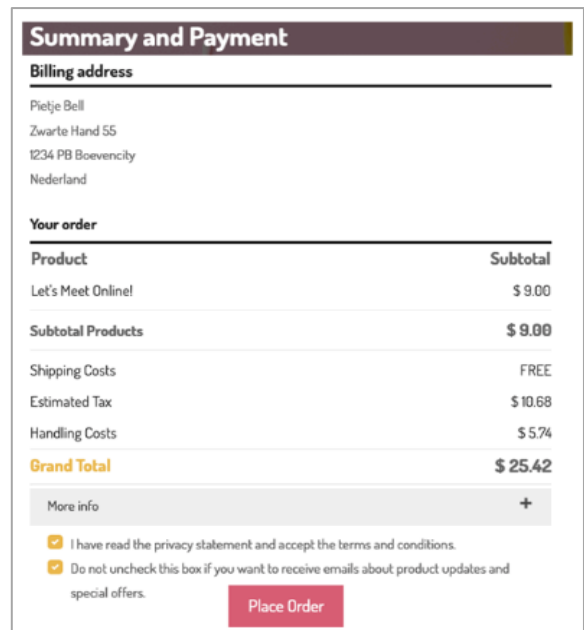


Figure 4: Really dark system. It includes the dark patterns: Hidden Costs, Hidden Information, Preselection & Trick Questions.

3.3 Experimental Task: Rating the Items

The attitudes of people towards the (malicious) interface design of the e-commerce systems was observed by performing an experiment. Within this experiment, participants (n=92) interacted with either the "really dark" or "really bright" web store (section 3.2). The assignment to either one of the versions of the web store was done randomly.

To ensure participants would see roughly the same set of pages while interacting with the web store,

they were given a shopping assignment: "You want to send four friends a postcard to let them know you still think of them in times of COVID-19. Find the best deal on COVID postcards in the shop, add them to your shopping basket, pay for them, and make sure they're coming your way! You don't need stamps or envelopes, as there's still lots of them in the drawer of your closet." After interacting with the web store, participants performed an evaluation task. The task consisted of filling out a questionnaire containing all the potential SDS questionnaire statements. The questionnaire was exactly the same for both versions of the web store. Participants needed to score each of the statements on a 5-point Likert Scale ranging from 1 (totally agree) to 5 (totally disagree).

Given the large pool of potential questionnaire items, there might be items provoking extreme agreement or disagreement among the respondents. For example, the statement "The system tricked me into sharing information I did not intend to share." could lead to extreme agreement for the "really dark system", whereas it could lead to extreme disagreement for the "really bright system". Statements that lead to extreme opposite responses are the ones that should be included in the final questionnaire. Items where there is ambiguity are not good discriminators of attitudes, and therefore should not be included in the final list of items. In the next section we detail the elimination process of such ambiguous responses.

4. FINDINGS

There were 92 participants in the between-subjects experiment, 46 in the "really dark" version and 46 in the "really bright" version. The final SDS, a subset of the 25 potential questionnaire items is presented at the end of this section.

4.1 Selecting Items for the SDS

To make judgments about which items should be selected for the final System Darkness Scale, the current study followed the unidimensional scaling approach of Trochim (2021). Several analyses were performed, each of them leading to the elimination of a set of potential questionnaire items. First a reliability analysis was performed to test the internal consistency of the initial SDS, consisting of all 25 items, as we want the final SDS to produce reliable, single 'darkness' scores. The internal consistency of the initial SDS was quite low (Cronbach's alpha = 0.5). A way to improve the internal consistency is to eliminate items that prove to be inconsistent. Item-total correlation can find inconsistent items by calculating the Pearson correlation coefficient for pairs of scores (one item of each pair is an item score, the other is the summed score of all items). The greater the value of the coefficient, the stronger

the correlation and the better the particular item contributes to the construct. Items with strong correlations should be retained and items with weak correlations should be eliminated. Eliminating these items leads to a higher in Cronbach's alpha and internal consistency. Table 3 shows the correlations with a correlation with the summed score of less than 0.4 were eliminated. The top 10 items remained candidates for inclusion in the SDS. After the reliability analysis with these items the internal consistency of the SDS improved drastically: Cronbach's alpha went up to 0.9.

	Item	Correlation
1	10	0.630
2	14	0.608
3	06	0.587
4	01	0.558
5	15	0.543
6	09	0.542
7	17	0.468
8	22	0.468
9	08	0.447
10	12	0.405
11	02	0.387
12	04	0.379
13	18	0.367
14	20	0.085
15	03	0.084
16	24	0.078
17	05	0.065
18	11	0.059
19	21	0.052
20	19	-0.003
21	07	-0.022
22	25	-0.048
23	16	-0.100
24	23	-0.144
25	13	-0.167

Table 3. Item-total correlation of SDS items

	Item	T-value
1	15	-
2	14	65.1
3	06	43.0
4	22	39.9
5	01	33.7
6	18	33.7
7	17	33.0
8	19	32.6
9	09	26.4
10	07	22.4
11	02	22.2
12	03	20.9
13	25	20.9
14	21	20.7
15	08	20.5
16	04	20.2
17	20	19.9
18	23	19.7
19	24	19.3
20	11	18.8
21	12	17.4
22	05	17.3
23	13	17.2
24	16	16.2
25	10	14.8

Table 4. T-values of t-tests of 25 potential SDS items

After this an item discrimination analysis was performed. Item discrimination analysis checks whether a difference exists between two sample sets - here 'participants that interacted with the "dark" web shop' and 'participants that interacted with the "bright" web shop'.

As we want to select the items that lead to extreme opposite responses for the "dark" and "bright" web shop, we want the item discrimination to be as high as possible. For each of the items, the data of the

quarters of participants that gave the highest and lowest ratings were selected. Average ratings were calculated for both these groups. Thereafter, t-tests were performed on the difference between the average value for each item's top and bottom quarter participants.

Table 4 shows the t-values resulting from the t-tests. Higher t-values mean that there is a greater difference between the highest and lowest participants. In other words, items with higher t-values are better discriminators. No t-value is shown for IT15, as performing a t-test is not possible on 'perfect' data.

As we want our final SDS items to have high item-total correlation and high discrimination, we took the 10 items that were selected as possible candidates for the SDS in the previous analysis and looked closely at their t-test values within the current analysis. There were several items that were represented in the top 10 of both the Correlation (table 3) and T-value table (table 4): Item 01, 06, 09, 14, 15, 17 and 22 (in green). These items were selected to proceed to the last analysis. Item 10, 08 and 12 (in red) were eliminated. Although these items received high correlation scores, they scored (relatively) low in terms of discrimination.

Table 5. Inter-correlations between all preselected items. For inter-correlations to reach an adequate level, items 09 and 17 were eliminated from the list of potential candidates for the SDS (red coloured cells).

	01	06	09	14	15	17	22
01	x	0.5	0.5	0.6	0.6	0.3	0.6
06	0.5	x	0.5	0.7	0.6	0.2	0.5
09	0.5	0.5	x	0.5	0.4	0.3	0.4
14	0.6	0.7	0.5	x	0.7	0.3	0.7
15	0.6	0.6	0.4	0.7	x	0.4	0.7
17	0.3	0.2	0.3	0.3	0.4	x	0.4
22	0.6	0.5	0.4	0.7	0.7	0.4	x

Another two items were eliminated from the list: Item 09 and 17. By doing this, the inter-correlations between all selected items moved from (± 0.2 to ± 0.7) to (± 0.5 to ± 0.7). To summarize: we wanted our final SDS items to have high item-total correlation, high discrimination, and close inter-correlations. There were five items that met all these requirements. As such, these items were included in the final System Darkness Scale (figure 5).

4.2 Scoring the SDS

The final System Darkness Scale consists of five Likert items. In order to retrieve a single 'overall darkness score', the answers of a respondent on

System Darkness Scale

Strongly disagree
Strongly agree

1. The system tricked me into performing certain actions I did not intend to do.

			X	
1	2	3	4	5

3
2. The system performed certain actions I was not aware of.

				X
1	2	3	4	5

4
3. The system pushed me into spending more money than I originally anticipated.

		X		
1	2	3	4	5

2
4. The system performed certain actions without my consent.

	X			
1	2	3	4	5

1
5. I felt deceived / misled by the system.

				X
1	2	3	4	5

4

Total Score = 14

SDS Score = 14 * 5 = 70

Fig. 5. Final System Darkness Scale (SDS)

Lastly, an item inter-correlation analysis was performed. In Likert Scale type questionnaires, close inter-correlation is preferred. As such, the goal of this analysis was to eliminate items that provoked loose inter-correlations with other items.

Table 5 shows the inter-correlations between the former seven selected SDS items. We decided ± 0.5 to be an adequate bottom line for item inter-correlations.

each of the Likert items should be combined.

The SDS has five response options for each of the five items. Following the scoring method of the SUS, we decided to code 'Strongly Disagree' as a 0, and 'Strongly Agree' as a 4. In order to retrieve the 'overall darkness score', a respondent fills out all questionnaire items, and then sums up the scores related to the chosen response options. The total sum of scores ranges between 0 (all 0s) and 20 (all

4s). By multiplying the sum of scores by 5, the summed score is converted into the 'overall darkness score', which intuitively ranges between 0 and 100. Within this range, 0 represents 'bright' (e.g., the system is not severe at all, or even harmless), whereas 100 represents 'dark' (e.g., the system is very severe). Figure 5 shows a scored SDS scale.

5. CONCLUSION

By way of concluding, we would like to stress important points on how SDS as a tool could be used in the HCI community, as well as to present limitations. We will also list future research directions our study triggers. Lastly, we will conclude by summarizing the contributions of our study.

5.1 SDS as a Tool

The widespread use of dark patterns and recent research on the dark pattern blindness indicate a systematic habit formation in internet users. To understand which patterns are perceived to be more severe is an important step.

SDS as a tool does not give precise insights into this perception process, i.e., it is not designed in a way that points out how the user assesses each dark pattern of a system, but rather shows the overall assessment of the system.

As such, it offers an unpolished picture of user perceptions. However, if combined with experts' assessments of the same system, the differences between expert scores and user scores will indicate an easy way of spotting dark pattern blindness.

An essential point in how dark patterns are defined is the stress on the intentions of the designers. If the stakeholders and designers act with malicious intentions, the resulting design strategies probably would fall under a dark pattern. However, if a designer simply copies a template with dark patterns, without being aware of its harmful effects on the users, the resulting design is called an anti-pattern. There is no qualitative or quantitative way of assessing a difference between an anti and a dark pattern, and as such SDS will also fail to spot these differences.

The SDS scores will only showcase if a system is perceived to contain deceitful UI design or not. However, we believe that SDS will discriminate between persuasive design strategies and dark patterns, as the measurement system is based on user assessment. The SDS can be used by researchers or professionals in companies or governments to screen their services as part of a formative or summative evaluation.

5.2 Limitations

One of the biggest limitations of this study has to do with the specific system being used to develop the SDS: the web store. Results of the final SDS (e.g., the five selected questionnaire items) are based solely on participants' interaction with this web store. As such, if you would, for example, ask participants to fill out the potential pool of 25 SDS statements after interacting with a 'bright' and 'dark' version of a gaming application, you might end up with different results and therefore with a different final SDS questionnaire. A second limitation of the current study is related to assembling the pool of potential SDS questionnaire items. Due to the limited time available for performing the current study, we came up with the list of potential items ourselves. A more elegant way of assembling a list of statements would have been to perform some brainstorm sessions with experts in the field. A less satisfactory feature - rather than a limitation of the SDS is that it (accidentally) became unbalanced: The items on the final scale are all worded in the same direction (e.g., the 'dark' direction). As such, the SDS becomes prone to response bias. This bias can be caused by the fact that users of the scale do not have to think about each statement.

5.3 Future Research

Future research should validate and build upon the final System Darkness Scale (SDS) that was created within the current study. By using various different systems, various different combinations of Dark Pattern types, the SDS has the potential to develop into an even better tool, or a set of tools to evaluate the darkness of systems. In other words, further research is desired to turn the SDS into a tool that becomes widely accepted within the HCI community. Hence, we cordially invite researchers and practitioners to use the SDS, elaborate on it, and to try it out on different services. The ecological validity of SDS will be better assessed with the adaptation of the tool by the HCI community.

More concretely, we see several potential research directions. One of them is to test SDS on applications and games, platforms that are known to apply dark patterns that are different than found in e-commerce web stores. Such experiments will render a reliable assessment if an SDS built with a focus on e-commerce could be generalized to other platforms or not. Another research focus is the observation of how SDS is used by different users such as researchers, designers, as well as actual users. An evaluation study of SDS that incorporates feedback from such a variety of user base will generate a feedback loop for a better version of SDS. A last research line is the development of a dark pattern evaluation tool or method with a focus on specific patterns and how users react to them on

individual cases, rather than offering an overall assessment of a system.

5.4 Contributions

We have presented the System Darkness Scale (SDS). The final SDS consists of 5 questionnaire items, all related to some aspect of darkness. The five items that were selected out of the pool of 25 potential items were found to all be measuring the same construct ('darkness'), providing the SDS with a good level of internal consistency. The items were also all capable of discriminating between the responses provided for a 'bright' and 'dark' version of a system, thereby providing the SDS with the competence of creating an accurate representation of the 'darkness' (SDS scores towards 100) or 'brightness' (SDS scores towards 0) of a system. We believe that our study results have the potential to become a widely used tool within the HCI community.

REFERENCES

- Berezhnoi, R. (2019). *What is UI Design and why is it important?* Available from: <https://ff5-studio.com/articles/what-is-user-interface-design-and-why-is-it-important/> (29 March 2022).
- Bösch, C., Erb, B., Kargl, F., Kopp, H. and Pfattheicher, S. (2016). Tales from the dark side: Privacy dark strategies and privacy dark patterns. In *Proceedings on Privacy Enhancing Technologies*. Darmstadt, Germany, July 19-22, 2016. (4): 237–254.
- Brignull, H. (2011). *Dark Patterns: Deception vs. Honesty in UI Design*. Available from <https://alistapart.com/article/dark-patterns-deception-vs.-honesty-in-ui-design> (29 March 2022).
- Brignull, H., (2013). *Dark Patterns-User Interfaces Designed to Trick People*. Available from <https://www.theverge.com/2013/8/29/4640308/dark-patterns-inside-the-interfaces-designed-to-trick-you> (29 March 2022)
- Brooke, J. (1996). Sus: a "quick and dirty" usability scale. In: Jordan, P.W., Thomas, B. and Ian Lyall McClelland. *Usability Evaluation In Industry*. CRC Press. 189–194.
- Di Geronimo, L., Braz, L., Fregnan, E., Palomba, F. and Alberto Bacchelli (2020). UI dark patterns and where to find them: a study on mobile applications and user perception. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14.
- Fogg, B.J. (2002). Persuasive technology: using computers to change what we think and do. In: *Ubiquity 2002*, 7.
- Gray, C.M., Kou, Y., Battles, B., Hoggatt, J. and Toombs, A.L. (2018). The dark (patterns) side of UX design. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*.1–14.
- Greenberg, S., Boring, S., Vermeulen, J. and Dostal, J. (2014). Dark patterns in proxemic interactions: a critical perspective. In: *Proceedings of the 2014 conference on Designing interactive systems*. 523–532.
- Gubaidulin, I. (2016). *Designing for Persuasion, Emotion and Trust*. Available from <https://uxdesign.cc/designing-for-persuasion-emotion-and-trust-cdac44c61d53> (29 March 2022).
- Gunnarsson, E. (2020). *Log In or Sign Up*. Available from <https://www.diva-portal.org/smash/get/diva2:1442280/FULLTEXT02> (29 March 2022).
- Láng, L., and Pudane, P.D. (2019). Deceptive Interfaces: A case study on Amazon's account deletion navigation and its effects on user experience. Available at <https://www.diva-portal.org/smash/get/diva2:1315156/FULLTEXT01.pdf> (29 March 2022).
- Luguri, J. and Strahilevitz, L. (2019). Shining a light on dark patterns. *Journal of Legal Analysis* 13.1 (2021): 43-109.
- Maier, M. (2019). Dark patterns - An end user perspective. Available at <https://www.diva-portal.org/smash/get/diva2:1330920/FULLTEXT01.pdf> (29 March 2022).
- Mathur, A., Acar, G., Friedman, M.J., Lucherini, E., Mayer, J., Chetty, M. and Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. In: *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–32.
- Mathur, A., Acar, G., Friedman, M.J., Lucherini, E., Mayer, J., Chetty, M. and Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. In: *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–32
- Meske, C. and Potthoff, T. (2017). The DINU-model—a process model for the design of nudges. In: *Proceedings of the 25th European Conference on Information Systems (ECIS)* pp. 2587—2597.

- Nodder, C. (2013). *Evil by design: Interaction design to lead us into temptation*. John Wiley & Sons.
- Nouwens, M., Liccardi, I., Veale, M., Karger, D. and Kagal, L. (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.
- Schaffer, E. (2009). Beyond Usability: *Designing Web Sites for Persuasion, Emotion, and Trust*. <https://www.uxmatters.com/mt/archives/2009/01/beyond-usability-designing-web-sites-for-persuasion-emotion-and-trust.php> (29 March 2022).
- Sommerer, C. and Mignonneau, L. (2008). *The Art and Science of Interface and Interaction Design* (Vol. 1). Vol. 141. Springer
- Sunstein, C.R. (2018). Better off, as judged by themselves”: a comment on evaluating nudges. *International Review of Economics* 65, 1, 1–8.
- Thaler, R.H. and Sunstein, C.R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- Trochim, W.M.K. (2021). *Likert Scaling*. Available at <https://conjointly.com/kb/likert-scaling/> (29 March 2022).
- Utz, C., Degeling, M., Fahl, S., Schaub, F. and Holz, T. (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.
- UX Booth (2018). *The Power and Danger of Persuasive Design*. Available at <https://www.uxbooth.com/articles/the-power-and-danger-of-persuasive-design/> (29 March 2022).